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

JOSÉ JAIRO DE SANTANA E SILVA

ESSAYS ON MISSPECIFICATION DETECTION IN DOUBLE BOUNDED RANDOM VARIABLES MODELING

Recife
2023

JOSÉ JAIRO DE SANTANA E SILVA

ESSAYS ON MISSPECIFICATION DETECTION IN DOUBLE BOUNDED RANDOM VARIABLES MODELING

Doctoral dissertation submitted to the Graduate Program in Statistics, Department of Statistics, Universidade Federal de Pernambuco as a requirement to obtain a Doctoral degree in Statistics. Concentration area: Mathematical Statistics

Advisor: Francisco Cribari-Neto
Co-advisor: Klaus Leite Pinto Vasconcellos

Recife

2023

Catalogação na fonte
Bibliotecária Nataly Soares Leite Moro, CRB4-1722

S586e Silva, José Jairo de Santana e
Essays on misspecification detection in double bounded random variables modeling / José Jairo de Santana e Silva. – 2023.
118 f.: il., fig., tab.

Orientador: Francisco Cribari-Neto.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2023.
Inclui referências e apêndices.

1. Estatística matemática. 2. Bootstrap. 3. Distribuição beta. I. Cribari-Neto, Francisco (orientador). II. Título.

519.5 CDD (23. ed.) UFPE- CCEN 2023 - 125

JOSÉ JAIRO DE SANTANA E SILVA

ESSAYS ON MISSPECIFICATION DETECTION IN DOUBLE BOUNDED RANDOM VARIABLES MODELING

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

Aprovado em: 27 de julho de 2023.

BANCA EXAMINADORA

Prof. Dr. Francisco Cribari Neto
Orientador, UFPE

Prof. Dr. Aldo William Medina Garay
Examinador Interno, UFPE

Prof. Dr. Getúlio José Amorim do Amaral
Examinador Interno,UFPE

Profa. Dra. Silvia Lopes de Paula Ferrari
Examinadora Externa à Instituição, USP

Prof. Dr. Gilberto Alvarenga Paula
Examinador Externo à Instituição, USP

AGRADECIMENTOS

Aos meus pais, Diva e Adeilson, por sempre me incentivarem a seguir em frente, mesmo em meio a adversidades. Vocês nunca mediram esforços para que eu pudesse atingir o objetivo de concluir o doutorado. Vocês foram o pilar necessário para manter o ânimo e a motivação.

Aos meus irmãos, Jaqueline, Júnior e Patrícia, à minha tia Fátima e aos meus sobrinhos Pedro, Maria Júlia e Heloísa. Dar orgulho a vocês sempre me motivou!

A Tati, por todo o companheirismo e cumplicidade ao longo dessa jornada.

A César, por toda a ajuda ao longo da pós-graduação. Sou muito grato a você!

A todos os amigos e colegas que fiz na pós-graduação: Luis, Lucas, Ranah, Larissa, Cristine, Adenice, Anny, Jordan, Anabeth, Eduardo, Alice, Cris, Jodavid, João, Hugo. Por todos os momentos de dificuldades e superações, bem como pelos bons momentos, meu muito obrigado.

Aos meus amigos e professores da graduação, por entenderem um pouco da minha ausência e por toda a amizade.

A todos os meus professores que muito contribuíram para minha formação, incluindo Raydonal, Gauss, Patrícia, Audrey e Maria do Carmo.

Um agradecimento especial aos meus orientadores, Francisco Cribari e Klaus Vasconcellos, por toda a paciência e vontade de me ajudar nessa caminhada. Esse tempo trabalhando com vocês me fez crescer bastante e seus ensinamentos sempre me fizeram ir mais longe.

Aos membros da banca, pela contribuição para a melhoria do trabalho.

À Capes, pelo apoio financeiro.

ABSTRACT

The beta distribution is routinely used to model variables that assume values in the standard unit interval. Several alternative laws have, nonetheless, been proposed in the literature, such as the Kumaraswamy and simplex distributions. A natural and empirically motivated question is: does the beta law provide an adequate representation for a given dataset? We test the null hypothesis that the beta model is correctly specified against the alternative hypothesis that it does not provide an adequate data fit. Our tests are based on the information matrix equality, which only holds when the model is correctly specified. They are thus sensitive to model misspecification. Simulation evidence shows that the tests perform well, especially when coupled with bootstrap resampling. We model state and county Covid-19 mortality rates in the United States. The misspecification tests indicate that the beta law successfully represents Covid-19 death rates when they are computed using either data from prior to the start of the vaccination campaign or data collected when such a campaign was under way. In the latter case, the beta law is only accepted when the negative impact of vaccination reach on death rates is moderate. The beta model is rejected under data heterogeneity, i.e., when mortality rates are computed using information gathered during both time periods.

The beta regression model is tailored for responses that assume values in the standard unit interval. In its more general formulation, it comprises two submodels, one for the mean response and another for the precision parameter. We develop tests of correct specification for such a model. The tests are based on the information matrix equality, which fails to hold when the model is incorrectly specified. We establish the validity of the tests in the class of varying precision beta regressions, provide closed-form expressions for the quantities used in the test statistics, and present simulation evidence on the tests' null and non-null behavior. We show it is possible to achieve very good control of the type I error probability when data resampling is employed and that the tests are able to reliably detect incorrect model specification, especially when the sample size is not small. Two empirical applications are presented and discussed.

Diagnostic analyses in regression modeling are usually based on residuals or local influence measures. They are used for detecting atypical observations. We develop a new approach for detecting such observations when the parameters of the model are estimated by maximum likelihood. It is based on the information matrix equality, which holds when the model is correctly specified. We consider different measures of the distance between two symmetric matrices and use them with the sample counterparts of the matrices in the information matrix equality in such a way that zero distance corresponds to correct model specification. The distance measures we use thus quantify the degree of model adequacy. We use such measures to identify observations that are atypical because they disproportionately alter the degree of model adequacy. We also introduce a modified generalized Cook distance and a new criterion that uses the two generalized Cook's distances (modified and unmodified). Empirical applications involving Gaussian and beta models are presented and discussed.

Keywords: Beta distribution; beta regression; bootstrap; information matrix test; model misspecification; Monte Carlo simulation.

RESUMO

A distribuição beta é usada rotineiramente para modelar variáveis que assumem valores no intervalo unitário padrão. Várias leis alternativas foram, contudo, propostas na literatura, tais como as distribuições Kumaraswamy e simplex. Uma questão natural e empiricamente motivada é: a lei beta fornece uma representação adequada para os dados sob análise? Nós testamos a hipótese nula de que o modelo beta está corretamente especificado contra a hipótese alternativa de que ele não fornece um ajuste adequado aos dados. Nossos testes são baseados na igualdade da matriz de informação, que somente é válida quando o modelo se encontra corretamente especificado. Os testes são, portanto, sensíveis a qualquer forma de especificação incorreta do modelo. Resultados de simulação mostram que os testes têm bom desempenho, especialmente quando utilizados com reamostragem bootstrap. Nós modelamos as taxas de mortalidade estaduais e municipais de Covid-19 nos Estados Unidos. Nossos testes de má especificação indicam que a lei beta representa adequadamente as taxas de mortalidade do Covid-19 quando estas são computadas com base em dados anteriores ao início da campanha de vacinação de Covid-19 ou com base em dados coletados quando tal campanha já se encontrava em andamento. No último caso, a lei beta só é aceita quando o impacto da vacinação sobre as taxas de mortalidade é moderado. O modelo beta é rejeitado sob heterogeneidade de dados, ou seja, quando as taxas de mortalidade são computadas usando informações coletadas durante ambos os períodos de tempo. Os testes de má especificação são estendidos para cobrir o modelo beta de regressão de precisão variável.

O modelo de regressão beta é usado com variáveis dependentes que assumem valores no intervalo unitário padrão, $(0, 1)$. Em sua formulação mais geral, contém dois submodelos, um para a média e outro para o parâmetro de precisão. Apresentamos expressões em forma fechada para estatísticas de teste da matriz de informação nessa classe de modelos. Reamostragem bootstrap é usada para alcançar melhor controle sobre a frequência de erro tipo I. São apresentados resultados de simulação de Monte Carlo sobre o comportamento dos testes, tanto sob a hipótese nula como sob a hipótese alternativa. Os resultados indicam que os testes são tipicamente capazes de detectar especificação incorreta do modelo, em especial quando o tamanho da amostra não é pequeno.

A análise de diagnóstico na modelagem de regressão é geralmente realizada com base na análise de resíduos ou influência local. Desenvolvemos uma nova abordagem para detectar pontos de dados atípicos em modelos para os quais a estimativa de parâmetros é realizada por máxima verossimilhança. A nova abordagem utiliza a igualdade da matriz de informação que é válida quando o modelo está corretamente especificado. Consideramos diferentes medidas da distância entre duas matrizes simétricas e as utilizamos com as contrapartidas amostrais das matrizes na igualdade da matriz de informação de tal forma que a distância zero corresponde à especificação correta do modelo. As medidas de distância que usamos quantificam, assim, o grau de adequação do modelo. Mostramos que elas podem ser usadas para identificar observações que contribuem desproporcionalmente para alterar o grau de adequação do modelo. Também introduzimos uma distância Cook generalizada modificada e um novo critério que utiliza as duas distâncias Cook generalizadas (modificadas e não modificadas). Aplicações empíricas envolvendo modelos de regressão gaussiano e beta são apresentadas e discutidas.

Palavras-chave: Bootstrap; distribuição beta; especificação incorreta; regressão beta; simulação de Monte Carlo; teste da matriz de informação.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 BETA DISTRIBUTION MISSPECIFICATION TESTS WITH APPLICATION TO COVID-19 MORTALITY RATES IN THE UNITED STATES

1.1 INTRODUCTION

Several variables of interest assume values in the standard unit interval, $(0, 1)$. This is the case, e.g., of rates, proportions and concentration indices. The beta distribution is commonly used to model such variables. For instance, Wiley, Herschkorn and Padian (1989) use the beta law to model the probability of HIV transmission in male-to-female sexual encounters and Bury (1999) lists applications of the beta law to engineering. Other applications of the beta distribution can be seen in Oguamanam, Martin and Huissoon (1995) (gear damage analysis), Sulaiman *et al.* (1999) (relative sunshine duration in Malaysia) and Elmer, Jones and Nagin (2018) (group-based trajectory modeling of neurological activity of comatose cardiac arrest patients). Additionally, Johnson, Kotz and Balakrishnan (1995) note that "[t]he beta distributions are among the most frequently employed to model theoretical distributions". It is also noted that the beta law arises naturally in 'normal theory' since $Z_1/(Z_1 + Z_2)$ is beta distributed if Z_1 and Z_2 are independent chi-squared random variables. The beta distribution can also be obtained as the limiting distribution of eigenvalues ratio in a sequence of random matrices.

Alternative distributions with support in the standard unit interval have been proposed in the literature and have been increasingly used in empirical analyses, such as, e.g., the Kumaraswamy (JONES, 2009) and simplex distributions (JØRGENSEN, 1997) and more recently, the unit-Weibull (MAZUCHELI; MENEZES; GHITANY, 2018) and reflected unit Burr XII (RIBEIRO *et al.*, 2021) distributions. It would then be useful to provide practitioners with a test that can be used to determine whether the beta law — which is still the most used model with fractional data — yields an adequate data fit. If not, an alternative model should be considered. This is our chief goal in this chapter. In particular, we present tests of the null hypothesis that the beta model is correctly specified against the alternative hypothesis that it is misspecified. Alternative models should be considered for the application at hand whenever the null hypothesis of correct beta model specification is rejected. In particular, we consider a general test of correct model specification that was introduced by White (1982), known as 'the information matrix test', and also some variants of it. The name of the test stems from the fact that the information matrix equality is known to only hold when the model is correctly specified. Information matrix test statistics are based on the sample counterparts of the model matrices that comprise such an equality. They were derived for several statistical models and distributions, e.g., the Gaussian linear regression model (HALL, 1987), binary data models (ORME, 1988), linear regressions with autoregressive and moving average errors (FURNO, 1996), logistic regressions (ZHANG, 2001), beta-binomial models (CAPANU; PRESNELL, 2008), and the negative binomial law (CHUA; ONG, 2013).

We obtain three information matrix test statistics for testing the null hypothesis that the beta model is correctly specified. They differ in the estimator used for the covariance matrix of a given random vector. The first two test statistics employ different estimators of the random vector's asymptotic covariance matrix whereas the third and final test statistic employs a resampling-based estimator of its exact covariance matrix. Since our numerical results show that the first two tests are considerably size-distorted in small to moderately large sample sizes, we also perform them using bootstrap critical values. It is noteworthy that the tests we develop are based on the information equality, which only holds when the model specification is not in error. As a consequence, they have power against any form of model misspecification, not only of distributional nature.

The Monte Carlo simulation evidence we report shows that the tests perform well, especially when coupled with bootstrap resampling. As noted above, three variants of the information matrix test are considered. For two of them, bootstrap resampling is used to obtain critical values that do not rely on asymptotic approximations whereas, in the remaining test, bootstrap resampling is used to estimate a covariance matrix that is used in the test statistic. Overall, the use of bootstrap resampling yields good control of the type I error frequency. Simulations in which the data were generated under the alternative hypothesis show that the tests are typically able to detect incorrect model specification, especially when the sample size is not small. Consider, e.g., the Kumaraswamy distribution, which is commonly used as an alternative law for fractional data. The numerical results we report show that when such a law is the true data-generating mechanism, the information matrix tests reject the beta model with probabilities around 0.9 for samples that contain 250 data points at the 10% significance level. Our Monte Carlo evidence also shows that the tests can successfully reject the univariate beta model when the sample size is not very small and the underlying law is beta but with non-constant means.

We model state and county Covid-19 mortality rates in the United States (US) using the univariate beta model. Three sample periods are considered: the first only includes observations from prior to the start of the nationwide vaccination campaign, the second encompasses data obtained before and after such a date, and the third and final period only includes data collected when the vaccination drive was under way. The testing inferences suggest that the beta law yields an adequate data representation for Covid-19 death rates in the first and third time periods. By contrast, the beta law is rejected when the data are heterogeneous, i.e., when the mortality rates are computed using information gathered prior to and during the nationwide vaccination drive. Interestingly, the univariate beta model is found to adequately describe the data in the third time period, in which mortality rates are negatively impacted by the reach of the vaccination drive. This happens because (i) in the initial part of the sample period vaccination was incipient and had little impact on the overall mortality figures and (ii) the negative relationship between the two variables is weakened by a few states, namely: Alaska, Arizona, Florida, Massachusetts, North Dakota, and Rhode Island. When all counties in such states are removed from the data, the inverse relationship between vaccination reach and death rates become considerably more intense, and the information matrix tests reject the adequacy of the univariate beta model, thus indicating that a more elaborate model should be used. The information matrix tests' inferences thus indicate that as long as the negative impact of vaccination reach on death rates is moderate, the beta law can be used to represent Covid-19 mortality rates. When such a negative impact becomes more pronounced, the univariate beta model should no longer be used.

The remainder of the chapter is organized as follows. The beta distribution and the corresponding maximum likelihood parameter estimation are briefly presented in Section 1.2. In Section 1.3, information matrix misspecification tests for the beta model are obtained. In particular, we introduce five tests, three of which based on bootstrap resampling. Monte Carlo simulation results are presented in Section 1.4. We evaluate the tests' null (size) and non-null (power) behaviors. An empirical analysis of Covid-19 mortality rates in the US is presented and discussed in Section 1.5. Finally, concluding remarks are offered in Section 1.6 together with directions for future research.

1.2 THE BETA DISTRIBUTION

Let Y be a beta-distributed random variable. Its density function, following the parametrization introduced by Ferrari and Cribari-Neto (2004), can be expressed as

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, 0 < \mu < 1, \phi > 0, \quad (1.1)$$

where $\mathbb{E}(Y) = \mu$ and ϕ is a precision parameter since, for fixed μ, $\mathrm{Var}(Y) = \mu(1-\mu)/(1+\phi)$ decreases as ϕ increases. We write $Y \sim \mathcal{B}(\mu,\phi)$. Unlike the standard beta parametrization, the parameters in (1.1) can be directly interpreted in terms of the distribution mean and precision. As we will see in the fifth section, it is useful to compare estimated precisions obtained from different model fits. The beta density in (1.1) is symmetric if $\mu = 0.5$ and asymmetric otherwise, and it reduces to the uniform density if $\mu = 0.5$ and $\phi = 2$. The beta density can be asymmetric to the left or to the right, and it can also be J-shaped, inverted J-shaped, and U-shaped. It is thus clear, as noted by Johnson, Kotz and Balakrishnan (1995), that "[b]eta distributions are very versatile and a variety of uncertainties can be usefully modelled by them." It is also noted that "[t]his flexibility encourages its empirical use in a wide range of applications."

Let $Y_1, \ldots, Y_n$ be independent and identically distributed (i.i.d.) beta-distributed random variables and let $y_1, \ldots, y_n$ be their observed, realized values. In what follows, $\boldsymbol{Y}$ and $\boldsymbol{y}$ denote the n-vectors of such random variables and realizations, respectively. Also, $\boldsymbol{\theta} = (\mu,\phi)^\top$ is the vector of beta parameters. Whenever required, we refer to μ and ϕ as θ_1 and θ_2, respectively. The log-likelihood function for $\boldsymbol{Y}$ evaluated at $\boldsymbol{y}$ is

$$\ell(\mu,\phi;\boldsymbol{y}) \equiv \ell(\boldsymbol{\theta};\boldsymbol{y}) = \sum_{t=1}^{n} \ell(\boldsymbol{\theta}; y_t),$$

where $\ell(\boldsymbol{\theta}; y_t) = \log(f(y_t;\mu,\phi))$ is the tth individual log-likelihood, which is given by

$$\ell(\boldsymbol{\theta}; y_t) = \log(\Gamma(\phi)) - \log(\Gamma(\mu\phi)) - \log(\Gamma((1-\mu)\phi)) + (\mu\phi - 1)y_t^* + (\phi - 2)y_t^\dagger,$$

with $y_t^* = \log(y_t/(1-y_t))$ and $y_t^\dagger = \log(1-y_t)$. Let $Y_t^* = \log(Y_t/(1-Y_t))$, $\mu^* = \mathbb{E}(Y_t^*)$, $Y_t^\dagger = \log(1-Y_t)$ and $\mu^\dagger = \mathbb{E}(Y_t^\dagger)$. It follows that $\mu^* = \psi(\mu\phi) - \psi((1-\mu)\phi)$ and $\mu^\dagger = \psi((1-\mu)\phi) - \psi(\phi)$, where ψ is the digamma function, i.e., the first derivative of the logarithm of the gamma function.

The score vector is $\nabla\ell(\boldsymbol{\theta},\boldsymbol{y}) = \partial\ell(\boldsymbol{\theta};\boldsymbol{y})/\partial\boldsymbol{\theta} = (\partial\ell(\boldsymbol{\theta};\boldsymbol{y})/\partial\mu, \partial\ell(\boldsymbol{\theta};\boldsymbol{y})/\partial\phi)^\top$, where

$$\frac{\partial\ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial\mu} = \sum_{t=1}^{n} \phi(y_t^* - \mu^*) \quad \text{and} \quad \frac{\partial\ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial\phi} = \sum_{t=1}^{n} [\mu(y_t^* - \mu^*) + (y_t^\dagger - \mu^\dagger)].$$

Fisher's information matrix for a single observation, $B(\boldsymbol{\theta})$, is defined as the expected value of the individual log-likelihood derivative outer product: $B(\boldsymbol{\theta}) = \mathbb{E}\left[\partial\ell(\boldsymbol{\theta};Y_t)/\partial\boldsymbol{\theta} \times \partial\ell(\boldsymbol{\theta};Y_t)/\partial\boldsymbol{\theta}^\top\right]$. For the beta model,

$$B(\boldsymbol{\theta}) = \begin{bmatrix} B_{\mu\mu} & B_{\mu\phi} \\ B_{\phi\mu} & B_{\phi\phi} \end{bmatrix},$$

where $B_{\mu\mu} = \phi^2 w$, $B_{\mu\phi} = B_{\mu\phi} = c$ and $B_{\phi\phi} = (\mu c)/\phi + (1-\mu)\psi'((1-\mu)\phi) - \psi'(\phi)$, ψ' being the trigamma function. The expressions for the quantities w and c can be found in Appendix A. The total information matrix, i.e., the information matrix for the complete sample, is $nB(\boldsymbol{\theta})$.

The maximum likelihood estimator of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, cannot be expressed in closed-form. Parameter estimates are typically obtained by numerically maximizing the log-likelihood function using a Newton or quasi-Newton nonlinear optimization algorithm. In what follows, we will use Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with analytical first derivatives for maximum likelihood estimation; for details, see Nocedal and Wright (2006).

1.3 BETA MISSPECIFICATION TESTS

Our goal in what follows is to obtain tests of correct model specification for the beta distribution. Our focus is on the information matrix test introduced in full generality by White (1982). Let $\boldsymbol{\theta}_0 = (\mu_0, \phi_0)^\top$ be the true parameter value. The beta model is taken to be correctly specified if Y_t follows the beta law with parameter vector $\boldsymbol{\theta}_0$ $\forall t$.

Let $A(\boldsymbol{\theta}) = \mathbb{E}\left[\partial^2\ell(\boldsymbol{\theta};Y_t)/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top)\right]$ be the expected Hessian of $\ell(\boldsymbol{\theta};Y_t)$. When the model is correctly specified and under the assumptions listed in Sections 2 and 3 of White (1982), the information matrix equality holds: $B(\boldsymbol{\theta}_0) = -A(\boldsymbol{\theta}_0)$; alternatively, $A(\boldsymbol{\theta}_\mathbf{0}) + B(\boldsymbol{\theta}_0) = O_{k\times k}$, where $O_{k\times k}$ denotes a k-dimensional square matrix of zeros, with $k = 2$. Evidence that such an equality fails to hold is thus taken as evidence of incorrect model specification. Our interest lies in testing the null hypothesis $\mathcal{H}_0$: $A(\boldsymbol{\theta}_\mathbf{0}) + B(\boldsymbol{\theta}_0) = O_{k\times k}$ (correct beta model specification) against the alternative hypothesis $\mathcal{H}_1 : A(\boldsymbol{\theta}_\mathbf{0}) + B(\boldsymbol{\theta}_0) \neq O_{k\times k}$ (beta model misspecification).

In what follows, we will present three information matrix test statistics that can be used to test the correct beta model specification. At the outset, we derive several quantities that are used in such test statistics. We obtain, for the beta model,

$$A_n(\boldsymbol{\theta};\boldsymbol{Y}) = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^2\ell(\boldsymbol{\theta};Y_t)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right] = \frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix}A_{n_{\mu\mu}} & A_{n_{\phi\mu}}\\ A_{n_{\mu\phi}} & A_{n_{\phi\phi}}\end{bmatrix},$$

where $A_{n_{\mu\mu}} = -\phi^2 w$, $A_{n_{\phi\mu}} = A_{n_{\mu\phi}} = (Y_t^* - \mu^*) - c$ and $A_{n_{\phi\phi}} = -(\mu c)/\phi - (1-\mu)\psi'((1-\mu)\phi) + \psi'(\phi)$. Expressions for c and w can be found, as noted earlier, in Appendix A. Additionally,

$$B_n(\boldsymbol{\theta};\boldsymbol{Y}) = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial\ell(\boldsymbol{\theta};Y_t)}{\partial\boldsymbol{\theta}}\times\frac{\partial\ell(\boldsymbol{\theta};Y_t)}{\partial\boldsymbol{\theta}^\top}\right] = \frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix}B_{n_{\mu\mu}} & B_{n_{\phi\mu}}\\ B_{n_{\mu\phi}} & B_{n_{\phi\phi}}\end{bmatrix},$$

where $B_{n_{\mu\mu}} = \phi^2(Y_t^* - \mu^*)^2$, $B_{n_{\phi\mu}} = B_{n_{\mu\phi}} = \phi\,(Y_t^* - \mu^*)\left[\mu(Y_t^* - \mu^*) + (Y_t^\dagger - \mu^\dagger)\right]$ and $B_{n_{\phi\phi}} = \left[\mu(Y_t^* - \mu^*) + (Y_t^\dagger - \mu^\dagger)\right]^2$. Notice that $A_n(\boldsymbol{\theta};\boldsymbol{Y})$ and $B_n(\boldsymbol{\theta};\boldsymbol{Y})$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ are consistent estimators of $A(\boldsymbol{\theta}_0)$ and $B(\boldsymbol{\theta}_0)$, respectively.

We also need to obtain

$$\boldsymbol{D}_n(\boldsymbol{\theta}) \equiv \boldsymbol{D}_n(\boldsymbol{\theta};\boldsymbol{Y}) = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{d}(\boldsymbol{\theta};Y_t),$$

where

$$\boldsymbol{d}(\boldsymbol{\theta};Y_t) = \operatorname{vech}\left(\frac{\partial^2\ell(\boldsymbol{\theta};Y_t)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} + \frac{\partial\ell(\boldsymbol{\theta};Y_t)}{\partial\boldsymbol{\theta}}\times\frac{\partial\ell(\boldsymbol{\theta};Y_t)}{\partial\boldsymbol{\theta}^\top}\right)$$

is a 3×1 vector with lth component given by

$$d_l(\boldsymbol{\theta};Y_t) = \frac{\partial^2\ell(\boldsymbol{\theta};Y_t)}{\partial\theta_i\partial\theta_j} + \frac{\partial\ell(\boldsymbol{\theta};Y_t)}{\partial\theta_i}\times\frac{\partial\ell(\boldsymbol{\theta};Y_t)}{\partial\theta_j},$$

where $i=j=1$ for $l=1$; $i=1$ and $j=2$ for $l=2$; $i=j=2$ for $l=3$. For the beta distribution, we obtain

$$\begin{aligned}
d_1(\boldsymbol{\theta};Y_t) &= \phi^2[(Y_t^* - \mu^*)^2 - w],\\
d_2(\boldsymbol{\theta};Y_t) &= (Y_t^* - \mu^*) - c + \phi(Y_t^* - \mu^*)\left[\mu(Y_t^* - \mu^*) + (Y_t^\dagger - \mu^\dagger)\right],\\
d_3(\boldsymbol{\theta};Y_t) &= -\mu\frac{c}{\phi} - (1-\mu)\psi'((1-\mu)\phi) + \psi'(\phi) + \left[\mu(Y_t^* - \mu^*) + (Y_t^\dagger - \mu^\dagger)\right]^2.
\end{aligned}$$

Note that $\boldsymbol{D}_n(\boldsymbol{\theta};\boldsymbol{Y}) = \operatorname{vech}(A_n(\boldsymbol{\theta};\boldsymbol{Y}) + B_n(\boldsymbol{\theta};\boldsymbol{Y}))$ is a vector that contains three elements. The information matrix test statistics we consider are functions of such a restrictions vector evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

Let

$$\begin{aligned}
V(\boldsymbol{\theta}) = \mathrm{E}\Big\{&\left[\boldsymbol{d}(\boldsymbol{\theta};Y_t) - \nabla\boldsymbol{D}(\boldsymbol{\theta})A(\boldsymbol{\theta})^{-1}\nabla\ell(\boldsymbol{\theta};Y_t)\right]\\
&\times\left[\boldsymbol{d}(\boldsymbol{\theta};Y_t) - \nabla\boldsymbol{D}(\boldsymbol{\theta})A(\boldsymbol{\theta})^{-1}\nabla\ell(\boldsymbol{\theta};Y_t)\right]^\top\Big\},
\end{aligned}$$

where $\boldsymbol{D}(\boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{d}(\boldsymbol{\theta}; Y_t)]$ and $\nabla \boldsymbol{D}(\boldsymbol{\theta}) = \partial \boldsymbol{D}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^\top$. White (1982) showed that, under correct model specification, $\sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}; \boldsymbol{Y})$ is asymptotically normally distributed with zero mean and covariance matrix $V(\boldsymbol{\theta}_0)$ and noticed that a natural consistent estimator for $V(\boldsymbol{\theta}_0)$ is

$$V_{n1}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\Big\{\Big[\boldsymbol{d}(\boldsymbol{\theta}; Y_t) - \nabla \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y}) A_n(\boldsymbol{\theta}; \boldsymbol{Y})^{-1} \nabla \ell(\boldsymbol{\theta}; Y_t)\Big] \\ \times \Big[\boldsymbol{d}(\boldsymbol{\theta}; Y_t) - \nabla \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y}) A_n(\boldsymbol{\theta}; \boldsymbol{Y})^{-1} \nabla \ell(\boldsymbol{\theta}; Y_t)\Big]^\top\Big\}$$

evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, where $\nabla \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y}) = \partial \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y})/\partial \boldsymbol{\theta}^\top$. Closed-form expressions for the elements of $\nabla \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y})$ in the beta model are given in Appendix A.

The first information matrix test statistic is

$$\zeta_1 = n\boldsymbol{D}_n(\hat{\boldsymbol{\theta}})^\top [V_{n1}(\hat{\boldsymbol{\theta}})]^{-1} \boldsymbol{D}_n(\hat{\boldsymbol{\theta}}),$$

where q is the number of components of $\boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y})$ considered ($q \leq 3$). Under $\mathcal{H}_0$, ζ_1 is asymptotically distributed as χ^2_q. The test is then carried out using critical values from such a distribution, i.e., $\mathcal{H}_0$ is rejected at significance level $\alpha \in (0,1)$ if $\zeta_1 > \chi^2_{q,1-\alpha}$, where $\chi^2_{q,1-\alpha}$ is the $1-\alpha$ χ^2_q quantile.

Alternative information matrix test statistics can be obtained by considering different consistent estimators for $V(\boldsymbol{\theta}_0)$. Chesher (1983) and Lancaster (1984) showed that it is possible to use a covariance matrix estimator that does not require third order log-likelihood derivatives. They use the fact that, under $\mathcal{H}_0$, $\nabla \boldsymbol{D}(\boldsymbol{\theta}_0) = -\mathbb{E}[\boldsymbol{d}(\boldsymbol{\theta}_0; Y_t) \times \nabla \ell(\boldsymbol{\theta}_0; Y_t)^\top]$ (LANCASTER, 1984). Let

$$L_n(\boldsymbol{\theta}; \boldsymbol{Y}) = -\frac{1}{n}\sum_{t=1}^{n}\Big[\boldsymbol{d}(\boldsymbol{\theta}; Y_t) \times \nabla \ell(\boldsymbol{\theta}; Y_t)^\top\Big].$$

The Chesher-Lancaster estimator of $V(\boldsymbol{\theta}_0)$ is

$$V_{n2}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\Big\{\Big[\boldsymbol{d}(\boldsymbol{\theta}; Y_t) + L_n(\boldsymbol{\theta}; \boldsymbol{Y}) B_n(\boldsymbol{\theta}; \boldsymbol{Y})^{-1} \nabla \ell(\boldsymbol{\theta}; Y_t)\Big] \\ \times \Big[\boldsymbol{d}(\boldsymbol{\theta}; Y_t) + L_n(\boldsymbol{\theta}; \boldsymbol{Y}) B_n(\boldsymbol{\theta}; \boldsymbol{Y})^{-1} \nabla \ell(\boldsymbol{\theta}; Y_t)\Big]^\top\Big\}$$

evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The corresponding information matrix test statistic is

$$\zeta_2 = n\boldsymbol{D}_n(\hat{\boldsymbol{\theta}})^\top [V_{n2}(\hat{\boldsymbol{\theta}})]^{-1} \boldsymbol{D}_n(\hat{\boldsymbol{\theta}}).$$

Under $\mathcal{H}_0$, ζ_2 is asymptotically distributed as χ^2_q and, as before, the test is carried out using asymptotic critical values.

It is noteworthy that $V_{n1}(\hat{\boldsymbol{\theta}})$ and $V_{n2}(\hat{\boldsymbol{\theta}})$ are consistent estimators of $V(\boldsymbol{\theta}_0)$, the latter being the asymptotic covariance matrix of $\sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}};\boldsymbol{Y})$. A consistent estimator of the exact covariance matrix of such a vector, say $V_{S_n}(\boldsymbol{\theta}_0)$, can be obtained by using parametric bootstrap resampling, as shown by Dhaene and Hoorelbeke (2004). The bootstrap estimator of $V_{S_n}(\boldsymbol{\theta}_0)$ based on B bootstrap samples, say $\hat{V}^*_B$, can be computed as follows:

1. Using the original sample $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$, compute $\hat{\boldsymbol{\theta}}$.
2. Obtain a random sample of size n, say $\boldsymbol{Y}^*_b = (Y^*_1, \ldots, Y^*_n)^\top$, from the beta law with $\boldsymbol{\theta}$ replaced with $\hat{\boldsymbol{\theta}}$, i.e., perform the pseudo-data generation from $f(\cdot;\hat{\boldsymbol{\theta}})$.
3. Using $\boldsymbol{Y}^*_b$, compute $\hat{\boldsymbol{\theta}}^*_b$ and $\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}^*_b;\boldsymbol{Y}^*_b)$.
4. Execute steps (2) and (3) B times, where B is a large positive integer.
5. Using the bootstrap replicates $\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}^*_1;\boldsymbol{Y}^*_1), \ldots, \boldsymbol{D}_n(\hat{\boldsymbol{\theta}}^*_B;\boldsymbol{Y}^*_B)$, compute the bootstrap estimator of $V_{S_n}(\boldsymbol{\theta}_0)$ as

$$\hat{V}^*_{n3,B} = \frac{n}{B-1}\sum_{b=1}^{B}(\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}^*_b;\boldsymbol{Y}^*_b) - \bar{\boldsymbol{D}})(\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}^*_b;\boldsymbol{Y}^*_b) - \bar{\boldsymbol{D}})^\top,$$

where $\bar{\boldsymbol{D}} = B^{-1}\sum_{b=1}^{B}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}^*_b;\boldsymbol{Y}^*_b)$.

For fixed n and as $B \to \infty$, it follows that $\hat{V}^*_{n3,B} \xrightarrow{p} V_{S_n}(\hat{\boldsymbol{\theta}})$ (DHAENE; HOORELBEKE, 2004). We thus arrive at a third information matrix test statistic for testing the correct beta model specification. It is given by

$$\zeta_3 = n\boldsymbol{D}_n(\hat{\boldsymbol{\theta}})^\top(V^*_{n3,B})^{-1}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}).$$

Under $\mathcal{H}_0$, for fixed B and $n \to \infty$, ζ_3 is asymptotically distributed as $T^2_{q,B-1}$, i.e., as Hotelling's T-squared distribution with q and $B-1$ degrees of freedom (DHAENE; HOORELBEKE, 2004). As before, the test is performed using asymptotic critical values.

The information matrix test statistics ζ_1, ζ_2 and ζ_3 measure the sample evidence against the correct beta model specification. When they assume large values and $\mathcal{H}_0$ is rejected at the usual significance levels, an alternative model should be used. A word of caution, however, is in order. The test based on ζ_3 is expected to perform well in small to moderately large samples since the test statistic uses a bootstrap estimator of the exact covariance matrix of $\sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}};\boldsymbol{Y})$. The tests based on ζ_1 and ζ_2, by contrast, may be considerably size-distorted when n is not large since the test statistics use estimators of the asymptotic covariance matrix of $\sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}};\boldsymbol{Y})$ and such an asymptotic covariance matrix may be a poor approximation for its exact counterpart when n is not large. To remedy that, we recommend that ζ_1 and ζ_2 testing inferences be based on critical values obtained from bootstrap resampling instead of on $\chi^2_{q,1-\alpha}$ (asymptotic critical values). To that end, for $i = 1,2$:

1. Using the original sample $\boldsymbol{Y} = (Y_1,\dots,Y_n)^\top$, compute $\hat{\boldsymbol{\theta}}$ and ζ_i.
2. Obtain a random sample of size n, say $\boldsymbol{Y}^*_b = (Y^*_1,\dots,Y^*_n)^\top$, from the beta law with $\boldsymbol{\theta}$ replaced with $\hat{\boldsymbol{\theta}}$.
3. Using $\boldsymbol{Y}^*_b$, compute $\hat{\boldsymbol{\theta}}^*_b$ and $\zeta^*_{i,b}$.
4. Execute steps (2) and (3) B times.
5. Reject $\mathcal{H}_0$ at significance level α if ζ_i exceeds the $1-\alpha$ quantile of $\zeta^*_{i,1},\dots,\zeta^*_{i,B}$.

The use of bootstrap resampling when performing testing inferences based on the information matrix test statistics ζ_1 and ζ_2 may considerably reduce size distortions since the critical values used in such tests are now obtained from estimates of the test statistics' exact null distributions.

As noted earlier, it is possible to test $q \leq 3$ restrictions. In what follows, we will test two restrictions since numerical evaluations not shown here for brevity revealed that the third element of $\boldsymbol{D}_n(\hat{\boldsymbol{\theta}};\boldsymbol{Y})$ always assumes very small values and has very small variance, especially when dispersion is low, which renders near singular estimates of $V(\boldsymbol{\theta}_0)$. As noted by White (1982), when an indicator is identically null it should be ignored; see the example on page 10 of his article. Unlike what happens in his example, the maximum likelihood estimators in our case cannot be expressed in closed form, and that is why we had to resort to numerical evaluations to determine whether there is a non-relevant restriction. We thus test $q = 2$ restrictions by using $\boldsymbol{d}(\boldsymbol{\theta};Y_t) = (d_1(\boldsymbol{\theta};Y_t), d_2(\boldsymbol{\theta};Y_t)))^\top$. Correspondingly, we drop the last row of $\nabla \boldsymbol{D}_n(\boldsymbol{\theta};\boldsymbol{Y})$. The asymptotic null distribution of ζ_1 and ζ_2 is χ^2_2, and that of ζ_3 is $T^2_{2,B-1}$, where B is the number of bootstrap replications used in the estimation of $V_{S_n}(\boldsymbol{\theta}_0)$.

According to White (1982), it is expected that the tests will be consistent (i.e., have unit power asymptotically) against any alternative which renders the usual maximum likelihood inference techniques invalid. In our case, maximum likelihood inference involves the estimation of the beta distribution mean and precision parameters. When Y follows other laws or when the values of the beta parameters are not the same for all observations, the test statistics are expected to diverge in probability so that unit power is achieved asymptotically. We performed Monte Carlo simulations using a number of alternative models as the true data generating mechanism, which include alternative laws, data inflation (i.e., data that contain zero and/or one values), and neglected regression structure. The results from these simulations are presented in the next section. They show evidence of asymptotic unit power under all sources of model misspecification we considered.

1.4 NUMERICAL EVIDENCE

We will now numerically evaluate the performance of the information matrix tests when used to determine whether the beta distribution yields a satisfactory data fit, i.e., when used to determine whether the beta model is correctly specified. Data generation is carried out under the null and alternative hypotheses (correct and incorrect model specification, respectively). Beta random number generation is performed using the acceptance-rejection method based on uniform random draws obtained using the Mersenne Twister method. Parameter estimates are obtained by numerically maximizing the beta log-likelihood function using the BFGS quasi-Newton algorithm with analytical first derivatives. The starting values used in the estimation of μ and ϕ are, respectively, $\bar{y}$ and $\bar{y}(1-\bar{y})/\widehat{\mathrm{Var}}(Y)-1$, where $\bar{y}=n^{-1}\sum_{t=1}^{n} y_t$ and $\widehat{\mathrm{Var}}(Y)=(n-1)^{-1}\sum_{t=1}^{n}(y_t-\bar{y})^2$. The number of Monte Carlo and bootstrap replications are, respectively, 5000 and 500. The null hypothesis is $\mathcal{H}_0$: "the beta model is correctly specified" and the alternative hypothesis is $\mathcal{H}_1$: "the beta model is misspecified".

The following tests are performed: ζ_1, ζ_{1B}, ζ_2, ζ_{2B}, and ζ_3. The ζ_{1B} and ζ_{2B} tests employ bootstrap critical values, and the ζ_3 test statistic uses a bootstrap covariance matrix estimate. The simulations were performed using the R statistical computing environment; see R Core Team (2023).

At the outset, data generation is carried out under $\mathcal{H}_0$, i.e., the observations are obtained as random draws from the beta distribution with mean μ and precision ϕ. The significance levels and sample sizes are, respectively, $\alpha = 10\%, 5\%$ and 1% and $n = 50, 100, 250, 500, 1000, 5000$. In the tables below, we omit rows corresponding to powers of 100% after the first row with such powers.

In what follows, we will report the tests' null and non-null rejection rates obtained from size (data generated under $\mathcal{H}_0$) and power (data generated under $\mathcal{H}_1$) simulations, respectively. Additionally, we will present p-value plots and size-power plots for the ζ_1, ζ_2 and ζ_3 tests, i.e., for the tests that do not employ bootstrap critical values. Based on the size simulations (the data-generating process is beta), we plot the tests' empirical sizes (vertical axis) against nominal sizes, i.e., against values of $\alpha \in (0,1)$ (horizontal axis). The 45° line indicates perfect agreement between actual and nominal sizes. Curves that lie above (below) such a diagonal line for a given range of values of α are indicative of liberal (conservative) behavior at those significance levels. It should be noted that, in this graphical analysis, α is not fixed at three values (0.10, 0.05 and 0.01) but varies from close to zero up to close to one. We thus obtain a comprehensive view of the tests' null behaviors. We also present plots that relate the tests' empirical powers (vertical axis) to the corresponding sizes (horizontal axis), computed for values of α ranging from close to zero up to close to one. The non-null rejection rates are computed using a data-generating process that differs from the beta law. It should be noted that since the non-null rejection rates are plotted using the empirical critical value for each nominal size (and not using asymptotic critical values) it is possible to compare the tests' non-null behaviors by properly accounting for any existing size distortions. The higher the curve, the more powerful the test. For more details on these plots, see Davidson and MacKinnon (1998).

In the first size simulation, the data are generated from the beta law with $\mu = 0.2$ and $\phi = 20, 40, 80, 120$. The null rejection rates of the ζ_1, ζ_{1B}, ζ_2, ζ_{2B} and ζ_3 tests are shown in Table 1. All entries are percentages. The reported results lead to interesting conclusions. First, the ζ_1 and ζ_2 tests, which use asymptotic critical values, are quite liberal when the sample size is not very large; even with $n = 1000$, considerable size distortions take place. Second, such tests have effective sizes that are close to the nominal sizes when bootstrap (rather than asymptotic) critical values are used. For example, when $\phi = 40$ and $n = 100$, the sizes of ζ_1 and ζ_2, at $\alpha = 10\%$, are 17.6% and 37.4%; when bootstrap critical values are used, these rates drop to 10.7% and 9.8%, respectively. The use of bootstrap resampling thus considerably reduces size distortions. Third, the size distortions of ζ_1 decrease when the value of ϕ increases. For example, the test's null rejection rates for $n = 100$ and $\alpha = 10\%$ are 20.1% and 12.3% when $\phi = 20$ and $\phi = 120$, respectively. It is worth noticing that the variance of Y decreases when the value of ϕ increases, and that translates into more accurate testing inferences. Fourth, the ζ_3 test tends to be conservative when $n \leq 1000$, and displays null rejection rates close to the nominal levels with $n = 5000$.

In the second set of size simulations, data generation was performed from the beta distribution with $\mu = 0.5$ and the same precision values as before. The tests' null rejection rates are presented in Table 2. All entries are percentages. In general, the new results are similar to those in the previous scenario. The ζ_1 and ζ_2 tests remain liberal, with ζ_1 exhibiting considerably higher null rejection rates relative to previous results. For example, when $\phi = 40$, $\alpha = 10\%$ and $n = 100$, the null rejection rate of ζ_1 is 28.4% whereas in the previous scenario it was 17.6%. The testing inferences are less accurate here because there exists more uncertainty since the variance of the beta distribution is maximal when $\mu = 0.5$; recall that such a variance is $\mu(1-\mu)/(1+\phi)$. The figures in Table 2 further show that the ζ_{1B} and ζ_{2B} tests display the smallest size distortions, being accurate even when n is small. For example, when $\phi = 20$ and $n = 50$, the sizes of ζ_{1B} and ζ_{2B}, at $\alpha = 10\%$, are 10.0% and 9.6%, respectively. It is thus clear that bootstrap resampling works remarkably well. Additionally, the ζ_3 test remains conservative when $\mu = 0.5$, but only for $\alpha = 10\%$ and 5%. The test exhibits small size distortions when $n \geq 250$. For instance, when $\phi = 20$ and $n = 250$, the test's null rejection rate, at $\alpha = 10\%$, is 9.3%.

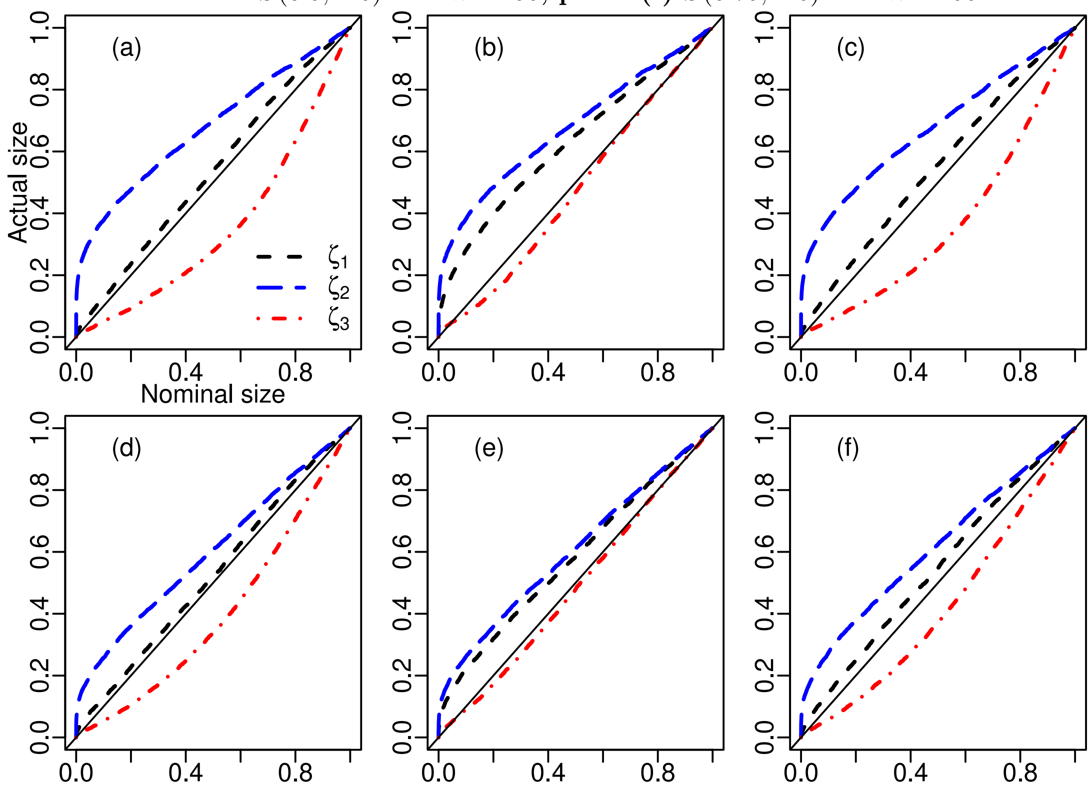
The third and final set size simulations was performed using $\mu = 0.75$ with the same precision values as before. We used $\mu = 0.75$ (and not $\mu = 0.8$) to avoid symmetry relative to the first scenario. The null rejection rates, expressed as percentages, are presented in Table 3. Overall, the results in this scenario are similar to those in Table 1 ($\mu = 0.2$). The ζ_1 and ζ_2 tests are liberal when $n \leq 1000$ and only become accurate with $n = 5000$. The ζ_{1B} and ζ_{2B} tests have the smallest size distortions. Such tests deliver accurate inferences even when n is small. For example, when $n = 50$, $\phi = 40$ and $\alpha = 10\%$, their null rejection rates are 10.1% and 9.7%, respectively. It should also be noted that the ζ_3 test exhibits conservative behavior when $n \leq 500$. For example, with $n = 500$, $\phi = 20$ and $\alpha = 10\%$, its null rejection rate is 8.7%.

The results presented above show that, in general, the ζ_1 test exhibits less liberal behavior when the mean of the distribution is not in the middle of the standard unit interval. For instance, when $\phi = 120$, $n = 250$ and $\alpha = 10\%$, the test's null rejection rates for $\mu = 0.2, 0.5, 0.75$ are 12.7%, 22.1% and 14.5%, respectively. Recall that the beta density is symmetric if $\mu = 0.5$ and asymmetric otherwise. It seems that the ζ_1 test incorrectly finds increasing evidence against the beta model as the distribution becomes more symmetric.

The results also show that the ζ_2 test is quite liberal in all scenarios, especially when the sample size is small. Finally, the ζ_3 test becomes more conservative as the distribution mean moves away from 0.5. For example, when $\phi = 80$, $n = 500$ and $\alpha = 10\%$, the test's null rejection rates for $\mu = 0.2, 0.5, 0.75$ are 7.0%, 9.5% and 7.2%, respectively.

Figure 1 contains p-value plots for the ζ_1, ζ_2 and ζ_3 tests corresponding to different values of μ. The sample sizes are $n = 100, 250$ and $\phi = 120$. The three curves move closer to the diagonal line when the sample increases from $n = 100$ to $n = 250$, thus indicating that the tests' size distortions for all nominal sizes decrease as n increases. It is also clear that ζ_1 and ζ_2 are liberal and ζ_3 is conservative regardless of the value of α, ζ_1 being less size-distorted than ζ_2, especially when the underlying beta law is asymmetric ($\mu \neq 0.5$). Interestingly, for all values of α, under distributional asymmetry (symmetry), ζ_1 (ζ_3) is the most accurate test.

Figure 1 – P-value plots; panel (a): $\mathcal{B}(0.2, 120)$ and $n = 100$, panel (b) $\mathcal{B}(0.5, 120)$ and $n = 100$, panel (c) $\mathcal{B}(0.75, 120)$ and $n = 100$, panel (d) $\mathcal{B}(0.2, 120)$ and $n = 250$, panel (e) $\mathcal{B}(0.5, 120)$ and $n = 250$, panel (f) $\mathcal{B}(0.75, 120)$ and $n = 250$.



Source: Author

We will now shift the focus to the tests' powers, i.e., to their ability of correctly identifying that the null hypothesis is false. In these simulations, the true data-generating process is not the standard beta law, i.e., it is not the beta distribution with constant

Table 1 – Null rejection rates (%), $\mu = 0.2$.

n	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3
	$\phi = 20$					$\phi = 40$				
	$\alpha = 10\%$									
50	17.5	9.9	48.6	9.5	5.5	15.4	10.0	50.1	10.0	4.2
100	20.1	11.0	39.1	10.7	6.1	17.6	10.7	37.4	9.8	5.2
250	19.0	11.0	27.9	10.8	7.1	17.0	9.9	27.1	10.2	6.6
500	17.1	10.6	21.3	10.8	8.2	15.8	10.4	20.9	10.5	7.4
1000	14.8	10.1	17.3	9.9	9.1	14.0	10.2	17.1	10.4	8.6
5000	11.9	10.2	12.3	10.3	9.9	11.6	10.4	12.2	10.2	9.7
	$\alpha = 5\%$									
50	10.1	5.3	41.8	4.8	3.3	9.2	5.3	43.7	5.2	2.5
100	12.4	5.7	31.9	5.5	3.5	10.6	5.3	30.7	5.4	3.0
250	12.6	5.3	21.3	5.4	4.0	10.4	4.9	20.7	5.3	3.7
500	11.1	5.6	15.7	5.5	4.5	9.9	5.5	15.2	5.6	4.1
1000	9.1	5.3	11.7	5.4	4.8	8.5	5.3	11.3	5.4	4.7
5000	6.8	5.5	7.2	5.6	4.9	6.5	5.3	7.1	5.4	4.9
	$\alpha = 1\%$									
50	3.8	1.1	30.2	1.2	1.2	3.5	1.0	31.6	1.3	0.9
100	4.7	1.4	22.4	1.4	1.4	3.9	1.1	20.8	1.2	0.9
250	4.6	1.0	12.9	1.3	1.5	3.9	1.2	11.9	1.4	1.3
500	4.8	1.1	8.2	1.4	1.5	3.9	1.1	8.0	1.2	1.3
1000	3.7	1.6	5.2	1.6	1.5	3.4	1.3	5.2	1.5	1.6
5000	2.3	1.5	2.5	1.4	1.2	1.8	1.2	2.2	1.2	0.8
	$\phi = 80$					$\phi = 120$				
	$\alpha = 10\%$									
50	13.8	9.9	48.5	10.4	4.3	12.8	10.6	49.0	10.3	4.1
100	14.0	9.8	37.8	10.1	4.5	12.3	10.2	37.7	10.2	4.7
250	14.5	10.1	26.5	9.5	5.8	12.7	10.3	25.7	10.7	5.2
500	14.0	10.2	20.9	9.7	7.0	12.4	9.9	21.0	10.2	6.2
1000	13.4	10.6	17.6	10.6	7.9	12.6	10.8	17.0	10.7	7.6
5000	11.1	9.9	12.0	9.9	9.5	11.0	10.1	12.2	9.9	9.4
	$\alpha = 5\%$									
50	7.8	4.8	41.5	4.8	2.5	7.7	5.4	42.1	5.3	2.2
100	8.1	5.0	30.7	5.0	2.4	7.3	4.9	30.9	5.3	2.6
250	8.6	5.2	20.7	4.9	3.2	8.0	6.0	19.7	5.8	2.9
500	8.3	5.1	14.6	4.9	3.9	7.1	5.0	14.6	5.3	3.0
1000	7.8	5.3	11.6	4.9	4.1	7.1	5.5	11.5	5.7	3.8
5000	6.3	5.3	7.1	5.2	4.6	5.5	4.9	6.4	5.0	4.4
	$\alpha = 1\%$									
50	2.6	1.2	29.3	1.5	0.9	3.0	1.2	30.1	1.3	0.6
100	3.0	1.2	20.1	1.2	0.8	2.5	1.1	21.1	1.2	0.9
250	2.9	1.3	11.3	1.1	0.9	3.1	1.6	12.1	1.7	1.1
500	2.7	1.0	7.4	1.0	1.3	2.3	1.4	7.6	1.4	0.8
1000	2.6	1.2	4.7	1.1	1.2	2.6	1.5	5.5	1.3	0.9
5000	1.8	1.1	2.2	1.1	1.1	1.3	1.0	1.7	0.9	0.9

Source: Author

Table 2 – Null rejection rates (%), $\mu = 0.5$.

n	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3
	$\phi = 20$					$\phi = 40$				
	$\alpha = 10\%$									
50	30.5	10.0	49.1	9.6	6.9	31.2	9.6	49.8	10.3	7.1
100	28.2	10.5	38.2	10.8	8.0	28.4	10.9	38.7	10.7	8.3
250	23.2	11.2	27.3	10.8	9.3	21.8	9.6	26.2	9.5	8.5
500	17.6	9.4	19.8	9.3	8.3	19.3	10.9	21.5	10.8	8.9
1000	15.7	10.2	16.8	10.1	9.5	16.1	10.9	17.1	10.8	9.7
5000	12.1	10.2	12.4	10.2	10.3	12.1	10.5	12.3	10.4	9.7
	$\alpha = 5\%$									
50	22.1	4.9	42.2	4.7	4.8	22.1	5.0	42.1	5.1	4.5
100	20.0	5.4	31.9	5.4	5.0	20.3	5.5	32.2	5.0	5.1
250	16.4	6.0	21.1	6.1	5.3	14.9	4.8	19.4	4.6	4.9
500	12.0	4.8	14.4	4.8	4.8	12.7	5.6	15.2	5.8	5.1
1000	10.2	5.3	11.0	5.4	4.8	10.8	5.6	11.8	5.6	5.1
5000	6.6	5.4	6.9	5.4	5.4	6.4	4.9	6.6	4.9	5.1
	$\alpha = 1\%$									
50	9.7	1.4	30.3	1.1	2.5	9.4	1.3	29.6	1.3	2.2
100	9.6	1.3	21.4	1.3	2.5	10.1	1.2	21.2	1.3	2.3
250	8.5	1.2	12.6	1.3	2.0	6.7	1.1	11.1	0.9	1.6
500	5.0	1.3	7.0	1.3	1.7	5.9	1.3	8.2	1.4	1.6
1000	4.3	1.4	5.1	1.4	1.6	4.1	1.1	5.2	1.2	1.5
5000	2.1	1.4	2.3	1.4	1.3	1.6	1.1	1.8	1.1	1.0
	$\phi = 80$					$\phi = 120$				
	$\alpha = 10\%$									
50	31.6	10.9	48.9	10.4	8.0	31.7	9.6	50.1	9.2	7.7
100	27.7	10.3	38.9	10.3	7.9	28.1	10.4	38.1	10.3	7.5
250	22.8	10.8	27.3	10.8	8.3	22.1	10.5	26.1	10.7	8.9
500	18.7	10.1	20.8	10.0	9.5	18.5	10.2	20.6	10.3	9.4
1000	14.8	10.5	15.8	10.4	9.4	15.5	10.0	16.7	9.9	9.9
5000	11.3	9.7	11.7	9.7	9.9	12.6	10.8	12.8	10.8	10.3
	$\alpha = 5\%$									
50	22.3	5.3	41.6	5.6	5.4	22.1	4.9	42.1	4.5	5.2
100	20.5	5.2	31.5	5.1	4.8	20.3	5.3	31.0	5.4	4.8
250	16.0	5.3	20.7	5.1	4.8	15.2	5.3	20.0	5.2	5.1
500	12.4	5.1	14.8	4.9	4.8	12.1	5.0	14.4	5.1	5.2
1000	10.0	5.2	10.9	5.2	5.4	9.6	5.3	10.7	5.4	5.6
5000	6.2	5.0	6.5	5.0	5.2	7.1	5.5	7.4	5.6	5.4
	$\alpha = 1\%$									
50	10.7	1.1	30.4	1.1	2.9	9.6	0.9	29.7	1.0	2.7
100	9.7	1.2	20.8	0.9	2.2	10.0	1.2	21.0	1.2	2.1
250	7.7	0.9	12.1	1.0	1.8	7.7	1.1	12.0	1.1	1.9
500	5.3	1.1	7.0	1.2	1.6	5.4	1.2	7.5	1.2	1.9
1000	3.9	1.2	4.8	1.2	1.6	4.1	1.4	5.1	1.3	1.4
5000	2.1	1.4	2.2	1.3	1.2	2.1	1.2	2.3	1.2	1.1

Source: Author

Table 3 – Null rejection rates (%), $\mu = 0.75$.

n	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3
	$\phi = 20$					$\phi = 40$				
	$\alpha = 10\%$									
50	21.6	10.8	52.3	10.3	5.6	18.0	10.1	49.3	9.7	5.7
100	20.7	9.6	37.1	9.8	5.9	20.0	10.4	37.7	10.1	6.0
250	18.8	10.1	26.0	9.9	7.2	18.3	10.5	27.1	10.5	7.2
500	16.5	9.4	20.3	9.4	8.7	16.1	9.9	20.6	9.5	7.9
1000	15.1	10.1	16.9	10.1	9.2	14.0	10.3	16.5	9.8	8.5
5000	11.2	9.7	11.5	9.7	10.0	12.2	10.8	12.7	11.0	9.3
	$\alpha = 5\%$									
50	12.9	5.4	44.5	5.4	3.5	10.5	5.5	42.3	4.9	3.4
100	13.2	4.9	30.4	4.7	3.6	12.1	5.2	30.8	5.4	3.3
250	12.9	5.0	20.0	5.0	4.4	12.3	5.0	20.8	5.4	4.1
500	10.5	4.7	14.2	4.9	4.6	10.4	4.9	14.6	4.5	4.1
1000	9.1	5.1	11.2	5.0	4.5	8.8	4.8	10.8	4.7	5.0
5000	6.3	4.8	6.6	4.9	4.8	6.5	5.3	7.1	5.2	5.1
	$\alpha = 1\%$									
50	4.8	1.3	31.8	1.4	1.8	4.4	1.3	29.3	1.1	1.4
100	4.5	0.9	20.0	0.8	1.4	4.2	1.3	20.2	1.1	1.2
250	5.5	1.2	11.8	1.2	1.6	4.5	0.9	12.2	1.3	1.4
500	4.2	1.0	7.1	1.0	1.7	3.9	1.1	7.1	1.2	1.5
1000	3.3	1.1	4.9	1.0	1.3	3.0	1.3	4.7	1.1	1.4
5000	1.8	1.0	2.0	0.9	1.2	1.8	1.1	2.1	1.2	1.2
	$\phi = 80$					$\phi = 120$				
	$\alpha = 10\%$									
50	15.5	10.4	48.9	10.5	4.0	14.4	10.8	49.7	10.5	4.6
100	16.4	10.3	38.0	9.9	5.2	14.5	10.3	37.9	10.2	5.3
250	15.3	9.8	26.7	10.0	5.6	14.5	10.4	27.4	10.9	5.6
500	14.6	10.1	20.0	10.4	7.2	13.3	10.5	19.8	10.5	6.7
1000	14.5	11.4	18.1	11.5	8.1	13.9	11.4	17.5	11.1	8.0
5000	12.0	10.7	12.9	10.7	10.3	11.3	10.2	12.3	10.3	9.5
	$\alpha = 5\%$									
50	9.5	5.4	42.6	5.3	2.5	8.7	5.5	42.9	5.3	2.8
100	9.4	5.3	31.5	5.4	3.0	8.3	5.3	30.4	5.1	2.9
250	9.2	5.0	19.9	5.3	3.3	8.7	5.6	20.6	5.4	3.1
500	9.1	5.3	14.8	5.5	4.0	8.3	5.4	14.7	5.9	3.4
1000	8.8	5.6	12.5	5.4	4.1	8.3	5.8	11.9	5.7	3.9
5000	6.8	5.6	7.4	5.5	5.0	5.9	5.2	6.7	5.1	4.7
	$\alpha = 1\%$									
50	3.9	1.3	29.7	1.2	0.9	3.2	1.1	31.2	1.0	0.9
100	3.3	1.0	20.6	1.4	1.2	2.8	1.1	20.0	0.9	1.0
250	3.6	1.4	11.4	1.3	1.3	3.0	1.3	12.5	1.3	0.9
500	3.3	1.2	7.5	1.1	1.2	3.1	1.3	8.0	1.2	0.8
1000	3.1	1.3	5.2	1.4	1.1	2.7	1.2	5.1	1.2	0.9
5000	1.7	1.1	2.0	1.2	1.3	1.6	1.2	2.1	1.2	1.1

Source: Author

parameters. Since the ζ_1 and ζ_2 tests are oftentimes considerably size-distorted, they are carried out using exact (not asymptotic) critical values obtained from the size simulations. The significance levels are $\alpha = 10\%, 5\%$.

At the outset, we use the Kumaraswamy law (JONES, 2009), $\mathcal{KW}(\omega, \phi)$, as the true data-generating mechanism. Here, ω is the distribution median and ϕ is a precision parameter. The parameter values are (i) $\omega = 0.2$ and $\phi = 5, 7.5$, (ii) $\omega = 0.5$ and $\phi = 10, 15$, and (iii) $\omega = 0.75$ and $\phi = 15, 25$. The tests' non-null rejection rates are presented in Table 4. All entries are percentages. The figures in this table show that the tests' powers are similar for $n \geq 100$, being close to 100% when $n \geq 250$. When $n = 50$, the ζ_3 test is generally the most powerful test. The reported results also show that the tests' powers increase with ϕ. That is, higher precision translates into more powerful tests. Also, when $\omega = 0.2$, the ζ_3 test exhibits slightly higher powers than the ζ_1 and ζ_{1B} tests, and these in turn exhibit noticeably higher powers than ζ_2 and ζ_{2B}. For illustration, with $\omega = 0.2$, $\phi = 7.5$, $n = 100$ and $\alpha = 5\%$, the non-null rejection rates of the ζ_1, ζ_{1B}, ζ_2, ζ_{2B} and ζ_3 tests are 61.3%, 65.2%, 56.7%, 56.8% and 69.7%, respectively. Here, ζ_3 is the best performer. It is also noteworthy that ζ_3 is the most powerful test when $\omega = 0.5$ for all values of ϕ and α. Additionally, it is seen that the ζ_2 and ζ_{2B} tests are more powerful than the ζ_1 and ζ_{1B} tests. Finally, when $\omega = 0.75$, for all values of α and ϕ, the ζ_1, ζ_{1B} and ζ_3 tests display similar powers, which are considerably higher than those of ζ_2 and ζ_{2B}.

Figure 2 contains size-power plots for ζ_1, ζ_2 and ζ_3. The sample size is $n = 100$ and the empirical powers were computed using $\mathcal{KW}(\omega, \phi)$ data-generating processes. The tests' powers are very similar for empirical sizes in excess of 0.4. For empirical sizes up to 0.4, ζ_3 is the clear winner, especially in the left and middle panels; in the right panel, the curves relative to ζ_1 and ζ_3 nearly coincide, both clearly lying above that of ζ_2. Also, ζ_1 is the worst performer when the distribution median lies at the center of the standard unit interval, i.e., $\omega = 0.5$; see panel (b).

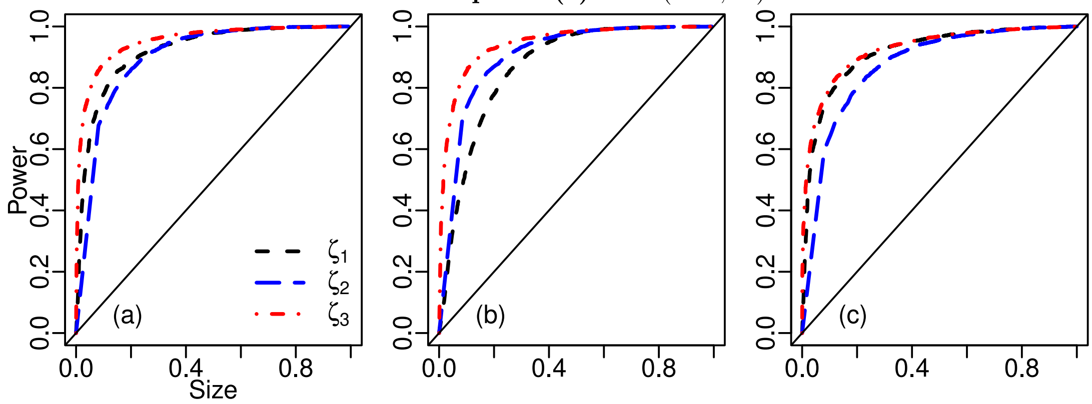
In the second scenario of power simulations, all samples are randomly generated from the unit Weibull law (MAZUCHELI; MENEZES; GHITANY, 2018), $\mathcal{UW}(\omega, \phi)$, where ω is the distribution median and ϕ is a precision parameter. For brevity, we only report results obtained using $\omega = 0.2, 0.5, 0.75$ and $\phi = 5$. The tests' non-null rejection rates are given in Table 5. All entries are percentages. It is worth noticing that all empirical powers are nearly equal to 100% when $n = 250$. When $\omega = 0.2$, ζ_1 is the best performer. The ζ_3

Table 4 – Non-null rejection rates (%), data generated from $\mathcal{KW}(\omega, \phi)$.

n	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3
	$\omega = 0.2$ and $\phi = 5$					$\omega = 0.2$ and $\phi = 7.5$				
	$\alpha = 10\%$									
50	37.6	36.6	29.8	30.2	37.6	48.9	52.7	41.5	42.3	49.0
100	59.4	57.7	56.4	56.3	65.3	76.6	80.2	72.0	72.6	79.0
250	93.1	92.7	96.2	95.9	98.0	98.4	98.9	99.3	99.2	99.4
500	99.5	99.4	100.0	100.0	100.0	99.9	99.9	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\alpha = 5\%$									
50	23.9	21.4	17.6	17.3	29.6	35.1	36.9	28.0	28.1	40.8
100	40.3	39.0	38.8	39.0	54.6	61.3	65.2	56.7	56.8	69.7
250	82.8	81.8	91.6	90.5	95.6	94.6	96.5	97.9	97.5	98.8
500	97.8	97.7	99.8	99.8	100.0	99.4	99.7	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0
5000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\omega = 0.5$ and $\phi = 10$					$\omega = 0.5$ and $\phi = 15$				
	$\alpha = 10\%$									
50	20.5	22.2	32.8	33.7	40.9	29.1	27.8	46.2	44.0	51.9
100	43.3	45.2	59.2	60.1	69.1	54.2	55.3	73.8	74.0	80.9
250	92.8	93.7	96.2	96.3	98.1	96.9	97.1	99.3	99.3	99.7
500	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\alpha = 5\%$									
50	8.9	10.2	20.0	21.3	32.6	12.5	12.4	31.4	30.3	42.9
100	24.0	25.6	43.8	44.4	59.0	33.7	34.6	58.5	59.7	71.9
250	83.0	83.4	91.5	91.4	95.7	91.3	91.2	97.6	97.6	98.9
500	99.8	99.8	99.9	99.9	99.9	99.9	99.9	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\omega = 0.75$ and $\phi = 15$					$\omega = 0.75$ and $\phi = 25$				
	$\alpha = 10\%$									
50	32.7	30.7	23.1	24.0	25.7	49.1	51.9	37.6	38.2	41.9
100	55.8	52.4	39.7	39.3	47.7	77.9	78.6	64.1	64.0	69.2
250	88.9	87.8	79.7	79.0	86.5	99.2	99.2	97.3	97.7	98.3
500	99.2	99.1	98.6	98.6	99.1	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\alpha = 5\%$									
50	18.2	16.9	13.9	14.5	19.3	32.3	34.4	25.3	25.5	32.3
100	39.1	36.4	25.4	27.1	38.6	64.3	65.3	49.1	49.1	60.3
250	81.6	78.2	64.3	65.7	79.3	97.7	98.2	93.4	93.5	96.5
500	98.0	97.7	95.6	95.6	98.1	100.0	100.0	99.9	99.9	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

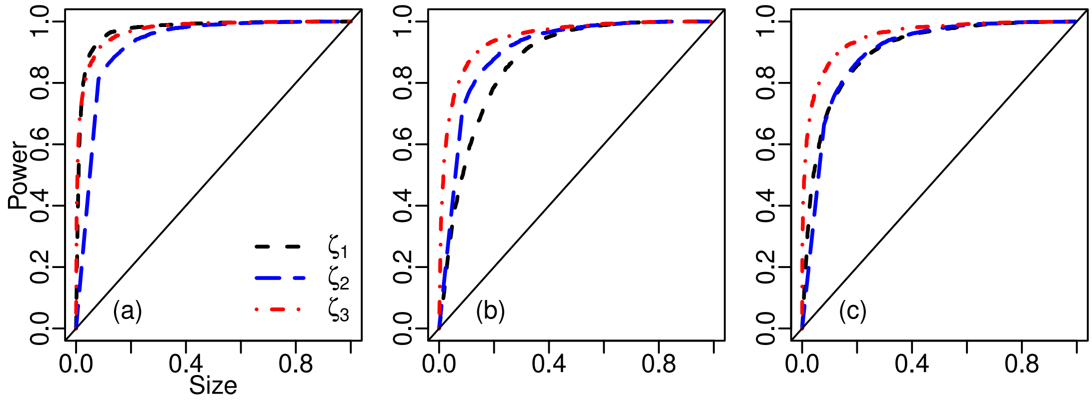
Source: Author

Figure 2 – Size-power plots, $\mathcal{KW}(\omega,\phi)$, $n=100$; panel (a): $\mathcal{KW}(0.2,7.5)$, panel (b): $\mathcal{KW}(0.5,15)$, panel (c): $\mathcal{KW}(0.75,25)$.



Source: Author

Figure 3 – Size-power plots, $\mathcal{UW}(\omega,\phi)$, $n=100$; panel (a): $\mathcal{UW}(0.2,5)$, panel (b): $\mathcal{UW}(0.5,5)$, panel (c): $\mathcal{UW}(0.75,5)$.



Source: Author

test is slightly more powerful than the other tests when $\omega = 0.5$ and 0.75.

Size-power plots are presented in Figure 3. The sample size is $n = 100$ and the tests' empirical powers were computed using unit Weibull data-generating mechanisms. In Figure 3 panel (a), the size-power curves of the ζ_1 and ζ_3 tests are clearly above that of the ζ_2 test for empirical sizes up to approximately 40%. In panel (b) of Figure 3, for empirical sizes up to about 50%, the curve of the ζ_3 test is above the curve of the ζ_2 test, which in turn is above that of the ζ_1 test. Finally, panel (c) of Figure 3 clearly favors ζ_3 for empirical sizes up to approximately 50%.

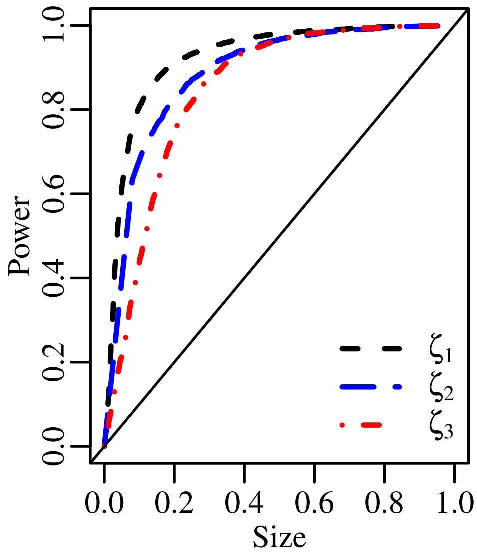
The next set of power simulation results was obtained using simplex (JØRGENSEN, 1997) data-generating mechanisms: all samples are randomly generated from $\mathcal{S}(\mu,\sigma)$, where μ is the distribution mean and σ is the dispersion parameter. For brevity, we only present results for $\mu = 0.75$ and $\sigma = 2$. The tests' non-null rejection rates, expressed as percentages, can be found in Table 6. It is noteworthy that the powers of the ζ_1, ζ_{1B}, ζ_2 and ζ_{2B} tests are quite high for $n \geq 250$. Also, the ζ_3 test is clearly less powerful than the

Table 5 – Non-null rejection rates (%), data generated from $\mathcal{UW}(\omega, \phi)$.

n	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3
	$\omega = 0.2$ and $\phi = 5$					$\omega = 0.5$ and $\phi = 5$					$\omega = 0.75$ and $\phi = 5$				
	$\alpha = 10\%$														
50	68.0	60.9	54.1	53.3	53.4	24.1	26.2	41.2	42.4	49.6	45.6	44.0	39.7	40.2	47.1
100	94.8	91.1	83.9	83.6	85.1	55.6	57.4	74.7	75.2	81.2	72.2	68.9	72.0	71.3	77.6
250	100.0	100.0	99.9	99.9	99.9	97.4	97.9	99.5	99.6	99.8	96.9	95.8	99.3	99.4	99.5
500	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.7	99.4	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\alpha = 5\%$														
50	51.7	43.3	40.7	40.4	44.0	10.2	11.8	27.5	28.5	40.3	28.2	26.0	25.8	25.9	37.9
100	89.4	82.7	72.3	72.9	77.4	34.8	36.3	60.9	61.1	72.9	54.8	49.3	56.3	55.7	69.0
250	100.0	99.9	99.5	99.5	99.7	92.8	92.7	98.4	98.4	99.1	91.0	89.1	97.9	97.8	98.8
500	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.5	98.3	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	100.0	100.0	100.0
5000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Source: Author

Figure 4 – Size-power plot, $\mathcal{S}(\mu,\sigma)$, $n=100$.



Source: Author

competing tests. For example, when $n=100$ and $\alpha=10\%$, the powers of the ζ_1, ζ_{1B}, ζ_2 and ζ_{2B} tests exceed 60% whereas that of the ζ_3 test is approximately equal to 22%.

Table 6 – Non-null rejection rates (%), data generated from $\mathcal{S}(\mu,\sigma)$.

n	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3
	$\mu=0.75$ and $\sigma=2$				
	$\alpha=10\%$				
50	39.5	27.8	42.0	40.4	9.8
100	80.6	62.9	68.0	66.2	22.2
250	99.3	98.1	97.5	97.6	79.0
500	100.0	100.0	100.0	100.0	99.6
1000	100.0	100.0	100.0	100.0	100.0
	$\alpha=5\%$				
50	20.4	13.7	28.9	26.5	5.9
100	61.6	37.0	53.7	51.3	12.6
250	97.6	94.7	94.8	94.0	58.0
500	100.0	100.0	100.0	100.0	98.8
1000	100.0	100.0	100.0	100.0	100.0

Source: Author

We present size-power plots constructed using the tests' empirical powers under simplex laws in Figure 4. The sample size is $n=100$. It can be seen that the curves relative to the ζ_1 and ζ_2 tests are similar. They both lie considerably above that of the ζ_3 test for effective sizes up to 40%.

Next, we consider the case in which the data are generated from the beta law but with a regression structure for the mean. That is, we use the beta regression model introduced by Ferrari and Cribari-Neto (2004) as the true model. Here, $\log(\mu_t/(1-\mu_t)) = \beta_1 + \beta_2 x_{t2}$. The true parameter values are $\beta_1 = -0.25$, $\beta_2 = 0.5$ and $\phi = 120$. The covariate values were generated from $LN(0, 0.5)$, i.e., as realizations from the log-normal distribution with parameters 0 and 0.5. Table 7 contains the tests' non-null rejection rates, all expressed as percentages. In general, all tests have high powers when the sample size is not very small. In particular, for $n = 250$ and $\alpha = 10\%$, the tests have powers close to or equal to 100%. When $n = 50$, ζ_1 is clearly less powerful than ζ_2 and ζ_3.

Table 7 – Non-null rejection rates (%), data generated from the beta distribution with a mean regression structure.

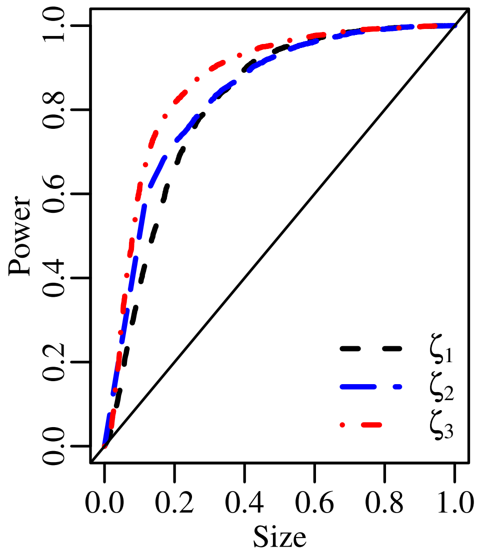
n	ζ_1	ζ_{1B}	ζ_2	ζ_{2B}	ζ_3
			$\alpha = 10\%$		
50	37.3	42.1	56.1	53.3	47.8
100	89.3	90.9	91.0	90.7	91.8
250	97.7	97.7	100.0	100.0	100.0
500	100.0	100.0	100.0	100.0	100.0
			$\alpha = 5\%$		
50	16.3	19.9	40.6	38.2	30.2
100	72.1	75.0	81.8	82.3	79.6
250	87.3	87.7	99.6	99.5	100.0
500	99.9	99.9	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0

Source: Author

In Figure 5, we present the size-power plot of ζ_1, ζ_2 and ζ_3 for $n = 50$. In general, the tests have similar powers when the effective size is smaller than 20% or larger than 60%. In the middle region of the graph, ζ_3 is the most powerful test.

We also performed simulations using the inflated beta distribution introduced by Ospina and Ferrari (2010) as the true model. It combines continuous and discrete components, and is used when Y_t assumes values in $[0, 1)$, $(0, 1]$ or $[0, 1]$ (inflation at zero, inflation at one, and double inflation, respectively). A common practice is to fit the standard beta distribution after replacing the inflated data points by $[Y_t(n-1)+0.5]/n$ (SMITHSON; VERKUILEN, 2006). We consider inflation at zero with $\Pr(Y_t = 0) = \lambda$. After the data were generated, all inflated values (zeros) were replaced by $0.5/n$, and then the standard beta law was fitted. The null hypothesis is false since the beta model is not the true data generating process. We wish to evaluate the information matrix tests'

Figure 5 – Size-power plot, $\mathcal{B}(\mu_t, \phi)$, $n = 50$.



Source: Author

ability to detect that the beta model is misspecified. Data generation was carried out using $\mu = 0.5$, $\phi = 20$ and $\lambda = 0.025$. We will not present the simulation results for brevity, but we note that the information matrix tests proved to be very powerful in this setting with non-null rejection rates close to 100% at $\alpha = 5\%$ for $n = 100$.

Overall, the results presented above favor the ζ_{1B}, ζ_{2B} and ζ_3 tests. The ζ_1 and ζ_2 tests typically display very large size distortions and their use should be avoided except when n is large. Regarding the ζ_{1B}, ζ_{2B} and ζ_3 tests, we note that the latter may be considerably conservative for some beta law parameter values. As a result, we recommend the use of the ζ_{1B} and ζ_{2B} tests in empirical analyses. Such tests showed good control of the type I error frequency and also good power in situations in which the data-generating process is not beta, in particular when $n \geq 250$.

It is also possible to test the null hypothesis that the variable of interest is beta-distributed using two alternative tests, namely: Anderson-Darling (AD) and Cramér-von Mises (CVM). They are usually carried with the modification proposed by Braun (1980), which accounts for unknown parameters in the distribution under test (in our case, beta). We performed Monte Carlo simulations to assess the finite sample behaviors of such tests using the configurations previously described. We do not present such results for brevity. We note, however, that both tests are conservative, i.e., their null rejection rates are smaller than the significance levels. For instance, when $n = 100$ ($n = 500$) the AD and CVM null rejection rates at the 10% significance level are, respectively, 7.6% and 6.3% (9.2% and 8.5%). Also, such non-parametric tests are substantially less powerful than the information matrix tests introduced in this chapter. For instance, when the true data-generating process is $\mathcal{UW}(0.5, 7.5)$ ($\mathcal{KW}(0.5, 15)$), the AD and CVM non-null rejection rates at $\alpha = 10\%$ are, respectively, 29.1% and 42.0% (28.9% and 39.3%) when $n = 5000$.

1.5 COVID-19 MORTALITY RATES IN THE US

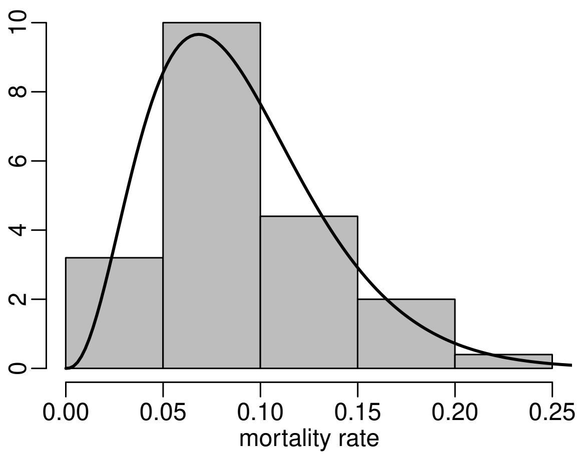
We will now present and discuss an analysis of Covid-19 mortality rates in the US. We use the three information matrix tests to determine whether the standard beta model provides an adequate representation of the data. We will also briefly comment on inferences drawn from the AD and CVM tests. Maximization of the beta log-likelihood function was performed using the BFGS method with analytical first derivatives. We used $B = 1000$ bootstrap replications for performing the ζ_{1B}, ζ_{2B} and ζ_3 information matrix tests. In what follows, we model state and county level data for three time periods. In each case, we will report the information matrix tests' p-values, the point estimates of the beta parameters and their standard errors. We report clustered standard errors computed using information on each state's region and on to each county's state.

The Covid-19 epidemic began in late 2019. It is estimated that approximately 247 million people had been infected with the new coronavirus by October 2021. The United States was the first country in the Americas to face a serious public health crisis brought on by the new coronavirus. In December 2020, on the 14th to be exact, the US government began a campaign to vaccinate healthcare workers and followed by vaccinating the general population. Covid-19 death rates started to decrease as vaccination progressed.

Our variable of interest are Covid-19 mortality rates per one hundred people. At the outset, we will work with statewide data, i.e., we use data on the 50 US states ($n = 50$). The death rates were computed using the cumulative number of deaths between January 22 and December 14 of 2020. We refer to this period as 'period 1'. The source of the data on Covid-19 deaths is the Centers for Disease Control and Prevention (<https://data.cdc.gov/>). Data on state populations in 2020 were obtained from Ribeiro *et al.* (2021). Since the sample size is small, we only consider bootstrap-based information matrix testing inferences. We wish to determine whether the univariate beta model provides an adequate representation of the data. The model has a simple structure and is based on the assumption that the observations are i.i.d. Can it provide an acceptable and useful representation of the US Covid-19 mortality rates?

The minimum, mean, median, and maximum mortality rates, and the standard deviation are 0.0164, 0.0903, 0.0894, 0.2001 and 0.0423, respectively. The maximal value corresponds to New Jersey. The maximum likelihood estimates of the beta parameters (clustered standard errors in parentheses) are $\hat{\mu} = 0.0900$ (0.0099) and $\hat{\phi} = 39.8208$ (15.5240). The p-values of the ζ_{1B}, ζ_{2B} and ζ_3 tests of correct beta specification are 0.2870, 0.6070 and 0.4472, respectively. The model is not rejected at the usual significance levels. We thus conclude that it adequately represents the US state mortality rates. In Figure 6 we present the histogram of the mortality rates together with the beta density evaluated at the maximum likelihood estimates. The estimated density clearly provides a good approximation to the data histogram.

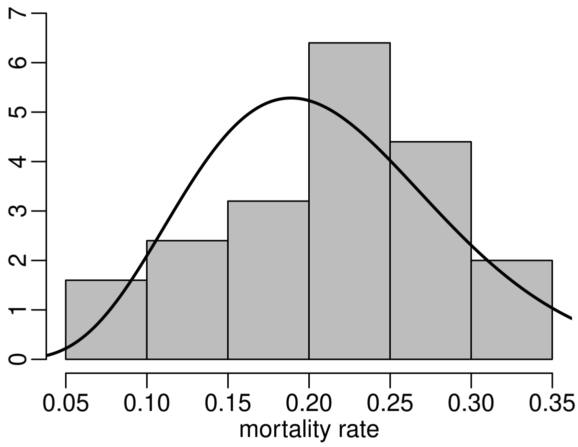
Figure 6 – Histogram and fitted beta density, period 1, state data.



Source: Author

The previous analysis was performed using mortality rates computed up to December 14, 2020. Next, we will conduct a similar analysis, but based on more recent data. We consider state mortality rates calculated using data from January 22, 2020 to October 31, 2021. We refer to this more extended time period as ‘period 2’. The minimum, mean, median, maximum and standard deviation values are 0.0550, 0.2120, 0.2227, 0.3370 and 0.0709, respectively. The maximum likelihood point estimates are $\hat{\mu} = 0.2112$ (0.0199) and $\hat{\phi} = 27.8235$ (8.9936). The estimated precision is now approximately 30% smaller than in the previous scenario. The p-values of the ζ_{1B}, ζ_{2B} and ζ_3 tests are 0.0600, 0.0570 and 0.0269, respectively. All tests reject the correct specification of the univariate beta model at the 10% significance level; ζ_3 rejects $\mathcal{H}_0$ at 5%. Figure 7 presents the data histogram and the estimated beta density. The estimated beta density does not adequately represent the data asymmetry.

Figure 7 – Histogram and fitted beta density, period 2, state data.



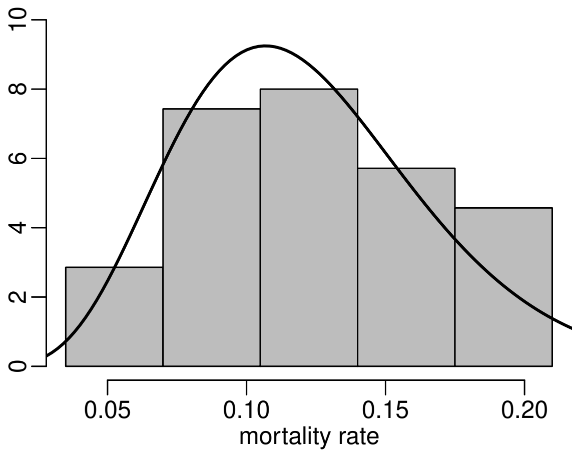
Source: Author

Unlike the previous results, all tests now reject the beta distribution at $\alpha = 10\%$. The data now cover two very different periods, namely: before and after the start of the nationwide vaccination campaign. There is thus clear data heterogeneity. The much smaller estimated precision (approx. 28 vs approx. 40) is probably due to such heterogeneity.

The mortality rates in the two periods show high positive correlation (0.8252), as expected, given the cumulative nature of the observations. The univariate beta model is not rejected by the information matrix tests when the shorter time period is used. It thus provides a good description of the statewide Covid-19 mortality rates. The second time period, however, covers the Covid-19 vaccination campaign. Since the reach and impact of such a campaign was uneven across the 50 states, for reasons that include partisan political connotations and other factors, Covid-19 mortality rates greatly differ before and after the beginning of the immunization campaign. There is thus clear heterogeneity in the two periods.

The two analyses presented so far are based on cumulative time periods, namely: (i) January 22 to December 14, 2020 (without vaccination) and (ii) January 22, 2020 to October 31, 2021 (without and with vaccination). In the following, we will only consider the most recent period (December 15, 2020 to October 31, 2021), 'period 3'. The minimal and maximal values are 0.0385 and 0.1985 whereas the mean and median values are 0.1218 and 0.1174, respectively; the standard deviation is 0.0432. The maximum likelihood estimates of μ and ϕ are 0.1216 (0.0153) and 52.8670 (11.1468), respectively. The estimated precision is even larger than that obtained by only considering the pre-vaccination time period (approx. 53 vs approx. 40). Recall that much lower precision was obtained when the longest time period was considered (approx. 28). The ζ_{1B}, ζ_{2B} and ζ_3 p-values are, respectively, 0.5900, 0.2860 and 0.5087. These large p-values indicate that there is very little evidence against the beta law. We thus conclude that despite the impact of vaccination on Covid-19 mortality, the univariate beta model still provides a good representation of the data. The data histogram and the fitted beta density are presented in Figure 8. Visual inspection of such a figure suggests that the beta law yields a reasonably good data fit. Interestingly, there is less skewness than in the previous two cases.

Figure 8 – Histogram and fitted beta density, period 3.

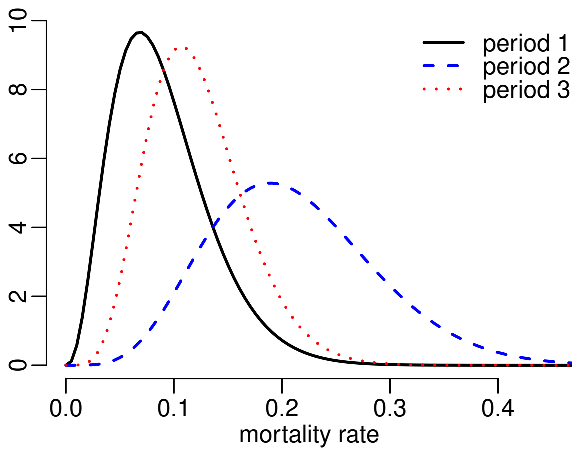


Source: Author

The three fitted beta densities are presented in Figure 9. Notice that the estimated densities for periods 1 and 3 are similarly shaped and with somewhat similar precisions. By contrast, the fitted beta density obtained using data that cover both the period in which there was no vaccination and that of the vaccination drive is much more disperse. As noted earlier, heterogeneity in the data leads to poor data fit. The information matrix tests indicated that the beta model yields an adequate data representation in periods 1 and 3, but in for period 2. It seems that the tests correctly detected that the heterogeneous nature of the data renders the beta law unable to adequately represent Covid-19 mortality rates.

We presented above an analysis of statewide Covid-19 mortality data in the US. The inferences obtained from the information matrix tests were quite informative. Such tests indicated that the beta law is able to adequately represent the data in two disjoint periods — before and after the start of the nationwide vaccination campaign —, but not when the two periods are combined.

Figure 9 – Fitted beta densities, state data.



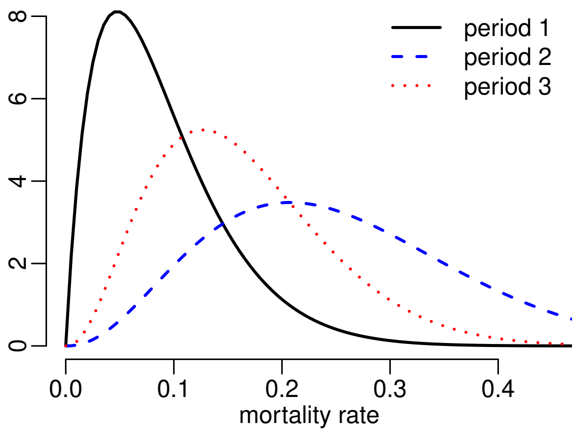
Source: Author

In what follows we will use death rates per 100 persons computed for US counties for periods 1, 2 and 3. The data on the cumulative total of deaths was obtained from the New York Times repository (<https://github.com/nytimes/covid-19-data>). In order to avoid inaccurate records, we only considered, in each time period, counties with at least one Covid-19 death and at least 15000 inhabitants. The sample sizes for periods 1, 2 and 3 are $n = 2073$, $n = 2080$ and $n = 2080$ respectively. Since the sample sizes are large, we will use all tests, i.e., ζ_1, ζ_{1B}, ζ_2, ζ_{2B} and ζ_3, and $\alpha = 5\%$. Mortality rates were calculated using the estimated populations in 2020 obtained from the Economic Research Service of the US Department of Agriculture (<https://www.ers.usda.gov>).

Initially, we will consider period 1. The minimum, mean, median, maximum and standard deviation values of the mortality rates are 0.0013, 0.0883, 0.0764, 0.4554 and 0.0596, respectively. The maximum likelihood estimates are $\hat{\mu} = 0.0884$ (0.0055) and $\hat{\phi} = 22.1618$ (1.9529). The ζ_1, ζ_{1B}, ζ_2, ζ_{2B} and ζ_3 tests' p-values are 0.0978, 0.1370, 0.0927, 0.1540 and 0.1022, respectively. No test rejects the beta law at $\alpha = 5\%$. The tests that use bootstrap resampling also do not reject such a hypothesis at $\alpha = 10\%$. The p-values of the tests that use asymptotic critical values are slightly smaller than 0.10. Overall, we conclude that Covid-19 mortality rates can be adequately represented by the beta law in period 1.

We will now consider the second period. The minimum, mean, median, maximum and standard deviation values are, respectively, 0.0122, 0.2513, 0.2418, 0.7376 and 0.1133. Also, $\hat{\mu} = 0.2512$ (0.0118) and $\hat{\phi} = 13.3931$ (1.1877). The estimate of ϕ is approximately 40% smaller than in the previous scenario. There is thus considerably more

Figure 10 – Fitted beta densities, county data.



Source: Author

uncertainty. The p-values of the ζ_1, ζ_{1B}, ζ_2, ζ_{2B} and ζ_3 tests are 0.0157, 0.0570, 0.0155, 0.0650 and 0.0087, respectively. The beta law is rejected at $\alpha = 1\%$ ($\alpha = 5\%$) by ζ_1, ζ_2 and ζ_3 (ζ_{1B} and ζ_{2B}). We conclude that the beta law does not provide an adequate data representation in period 2.

Next, we will perform inferences with data from period 3. The minimum, mean, median, maximum and standard deviation of the mortality rates are 0.0085, 0.1632, 0.1528, 0.4734 and 0.0786 respectively. The point estimates are $\hat{\mu} = 0.1631$ (0.0083) and $\hat{\phi} = 20.4761$ (1.5648). The p-values of the ζ_1, ζ_{1B}, ζ_2, ζ_{2B}, and ζ_3 tests are 0.0529, 0.0830, 0.0360, 0.0750, and 0.0903, respectively. Except for ζ_2, no test rejects the beta law at $\alpha = 5\%$. We thus conclude that it can be used to adequately represent county-level Covid-19 mortality rates in the third and final period. We will return to these results later.

Figure 10 contains the estimated densities for the three time periods obtained using county data. They are similar to those obtained using statewide data; see Figure 9. Notice that there is considerably more uncertainty when data from period 2 are used.

Interestingly, similar testing inferences were obtained with state and county data, namely: (i) the univariate beta model provides an adequate description of Covid-19 mortality rates with data either from prior to the nationwide vaccination drive or from when such a drive was under way; (ii) there is evidence against the correct specification of the beta model when Covid-19 mortality rates are computed using data that cover both periods (no vaccination and nationwide vaccination). The tests thus indicate that the beta distribution is not an adequate model for Covid-19 mortality rates under substantial data heterogeneity.

As noted earlier, we also performed the AD and CVM tests using both state and county data. The corresponding p-values for state data are: 0.4691 and 0.8734, period 1; 0.2277 and 0.4339, period 2; 0.9413 and 0.3360, period 3. With county data, we obtained the following p-values: 0.7414 and 0.5299, period 1; 0.3250 and 0.4856, period 2; 0.8765 and 0.8010, period 3. All p-values are quite large, and hence the beta model is not rejected in all scenarios, i.e., for the three time periods and when state or county data are used. In particular, unlike the information matrix tests, the two non-parametric tests are not able to reject the beta model when there is marked data heterogeneity (period 2). By contrast, our tests indicate that the univariate beta model is only appropriate when there is reasonable homogeneity in the data (periods 1 and 3).

We will now further examine (i) the data heterogeneity that caused the rejection of beta law in period 2 and (ii) the acceptance of the beta law in period 3 when the vaccination drive was under way. As noted earlier, the Covid-19 mortality rates computed for period 2 cover two quite distinct periods: January 22, 2020 through December 14, 2020 (period 1) and December 15, 2020 through October 21, 2021 (period 3). (Recall that period 2 consists of the merging of periods 1 and 3.) The correlation coefficient between statewide death rates in periods 1 and 3 is weak: 0.3735. This small correlation strength is indicative that the mortality rates in such periods obey different dynamics. This was expected because, unlike what took place in period 3, there was no nationwide vaccination drive in period 1. Additionally, the states with the lowest mortality rates in period 1 (period 3) are Vermont, Hawaii, Maine, Oregon, and Utah (Vermont, Hawaii, New York, Alaska, and Maine) whereas those with the highest death rates in period 1 (period 3) are New Jersey, Massachusetts, Mississippi, Rhode Island, and North Dakota (Arizona, Alabama, West Virginia, Florida, and Georgia). Consider, e.g., New Jersey and Massachusetts. They are the states with the highest Covid-19 mortality rates in period 1, and yet their corresponding ranks in period 3 are 28 and 32. Arizona and Alabama display the highest death rates in period 3, and yet their ranks in period 1 are 14 and 9, respectively. Again, it is clear that the death rates in periods 1 and 3 (which are combined in period 2) are considerably heterogeneous.

Next, we will examine again Covid-19 mortality rates in period 3; in particular, we will examine the finding that the univariate beta model yields an adequate representation for such rates. There was a nationwide vaccination drive under way in period 3, and its reach negatively impacted death rates. We obtained data on the total number of fully vaccinated people by October 31, 2021. The source of the data is the Our World in Data repository (<https://ourworldindata.org/us-states-vaccinations>). The correlation between death and vaccination rates in period 3 is -0.5858 (state data). A natural question is: Given that mortality rates are negatively impacted by vaccination rates, why was the univariate beta model found to be correctly specified? Why use a fixed mean model if the distribution mean appears to be impacted by an explanatory variable (vaccination rate)? At the outset, we note that some states considerably weaken the inverse relationship between the two variables in period 3, namely: Alaska, Arizona, Florida, Massachusetts, North Dakota, and Rhode Island. In particular, the Arizona, Florida, Massachusetts, and Rhode Island (Alaska and North Dakota) Covid-19 mortality rates are higher (lower) than expected based on the corresponding vaccination levels. The inverse correlation between death and vaccination rates becomes considerably stronger when computed without such states: -0.7592 (state data). We removed from the data all counties of the six states that weaken the impact of vaccination reach on death rates, and performed the tests again. The ζ_1, ζ_{1B}, ζ_2, ζ_{2B}, and ζ_3 p-values become 0.0289, 0.0600, 0.0162, 0.0530, and 0.0460, respectively. The ζ_1, ζ_2 and ζ_3 tests now reject the univariate beta model at $\alpha = 5\%$ whereas the ζ_{1B} and ζ_{2B} p-values are only marginally larger than 0.05. Hence, there is now evidence against the model. Overall, the information matrix tests' inferences suggest that, as long as the negative impact of vaccination reach on death rates is moderate (complete data), the beta law can be adequately used to represent Covid-19 mortality rates. When such a negative impact becomes more pronounced (incomplete data, counties of six states removed from the data), the univariate beta model no longer should be used. In that case, practitioners should search for a more elaborate model. By contrast, the two non-parametric tests continue to accept the univariate beta model even when the Alaska, Arizona, Florida, Massachusetts, North Dakota, and Rhode Island counties are not considered; the AD and CVM p-values are 0.3025 and 0.5788, respectively.

Finally, using the three county data samples, we compare the data fits yielded by the beta distribution to those obtained with the following alternative laws: Kumaraswamy, simplex, and unit Weibull. To that end, we computed, for each sample period and for each distribution, the values of the following information criteria: Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc), Bayesian Information

Criterion (BIC), Hannan-Quinn Information Criterion (HQIC), Weighted-Average Information Criterion (WIC) and Empirical Information Criterion (EIC). The latter employs bootstrap resampling and proved to be very effective in dynamic beta modeling; see Cribari-Neto, Scher and Bayer (2022). We used 1000 bootstrap replications, i.e., 1000 pseudo-samples were generated for computing the EIC values. We also computed the AD and CVM statistics. For all measures, smaller values indicate better data fits. The results are presented in Table 8. They show that, according to all information criteria (AIC, AICc, BIC, HQIC, WIC, and EIC), the best data fits in the three sample periods are yielded by the beta law. Considering the two non-parametric test statistics, in period 1 (period 2) [period 3], the beta model was the winner according to both of them (the runner-up according to both statistics, slightly behind the Kumaraswamy law) [the winner according to CVM and the runner-up according to AD, behind the Kumaraswamy model]. Considering the eight measures and the three sample periods, the beta law was the winner in 21 out of the 24 cases. Figure 11 contains the data histogram and the estimated beta density for period 3, as in Figure 8, together with the fitted Kumaraswamy (KW), simplex and unit Weibull (UW) densities. Visual inspection of the figure shows that the beta law best fits the data histogram. In order to further examine the two best data fits, we produced quantile-quantile (QQ) plots for the beta and Kumaraswamy laws, again using data from period 3. In both panels of Figure 12, empirical quantiles are plotted against theoretical quantiles, the 45° degree line indicating perfect agreement between both sets of quantiles. The Kumaraswamy and beta laws fit the data quite well up to approximately 0.35 and 0.45, respectively. It is then clear that the latter outperforms the former in the sense that it yields better agreement between empirical and theoretical quantiles.

1.6 CONCLUDING REMARKS

The beta distribution is commonly used to model variables that assume values in the standard unit interval. We developed information matrix tests that can be used to test whether the univariate beta model yields an adequate representation of the data. The null hypothesis of correct model specification is tested against the alternative hypothesis that the model specification is in error. The tests seek to verify whether the information matrix equality holds. As is well known, this equality only holds when the model is correctly specified. The tests' small sample behavior can be improved by using data resampling

Table 8 – Goodness-of-fit measures.

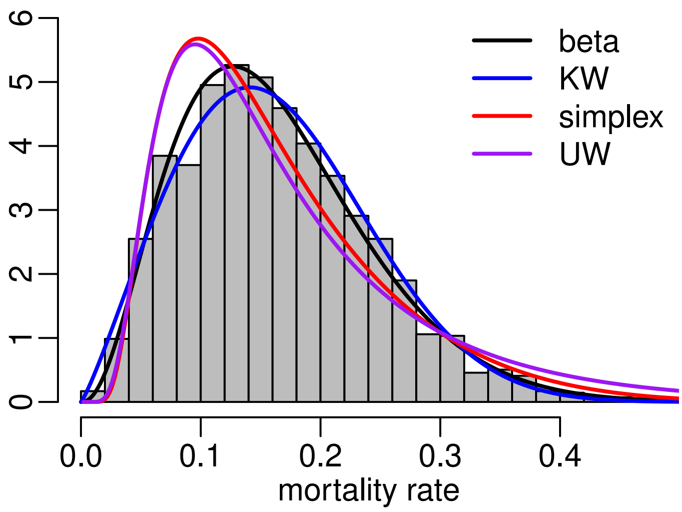
Period	Criterion	beta	KW	simplex	UW
1	AIC	−6466.687	−6441.745	−5993.029	−6144.892
	AICc	−6466.681	−6441.739	−5993.023	−6144.887
	BIC	−6455.413	−6430.471	−5981.755	−6133.619
	HQIC	−6462.555	−6437.613	−5988.897	−6140.760
	WIC	−6457.755	−6432.813	−5984.096	−6135.960
	EIC	−6477.157	−6452.015	−6013.894	−6160.247
	AD	2.958	3.170	5.544	3.788
	CVM	0.652	0.657	1.387	0.911
2	AIC	−3287.865	−3266.338	−3076.004	−3033.648
	AICc	−3287.859	−3266.332	−3075.994	−3033.643
	BIC	−3276.584	−3255.058	−3064.720	−3022.368
	HQIC	−3283.731	−3262.204	−3071.866	−3029.515
	WIC	−3278.926	−3257.399	−3067.061	−3024.710
	EIC	−3300.680	−3279.034	−3093.013	−3049.644
	AD	4.031	3.114	5.707	4.181
	CVM	0.674	0.471	0.966	2.120
3	AIC	−4873.815	−4862.928	−4666.304	−4558.565
	AICc	−4873.809	−4862.922	−4666.298	−4558.559
	BIC	−4862.534	−4851.647	−4655.024	−4547.284
	HQIC	−4869.681	−4858.794	−4662.170	−4554.431
	WIC	−4864.876	−4853.989	−4657.365	−4549.626
	EIC	−4885.201	−4875.239	−4680.756	−4571.892
	AD	2.595	2.335	4.443	4.093
	CVM	0.522	0.606	1.020	0.881

Source: Author

(bootstrap). We presented the results of extensive Monte Carlo simulations that showed that the tests have good power against different forms of model misspecification, including the case in which the univariate beta model is fitted using data that have an underlying regression structure.

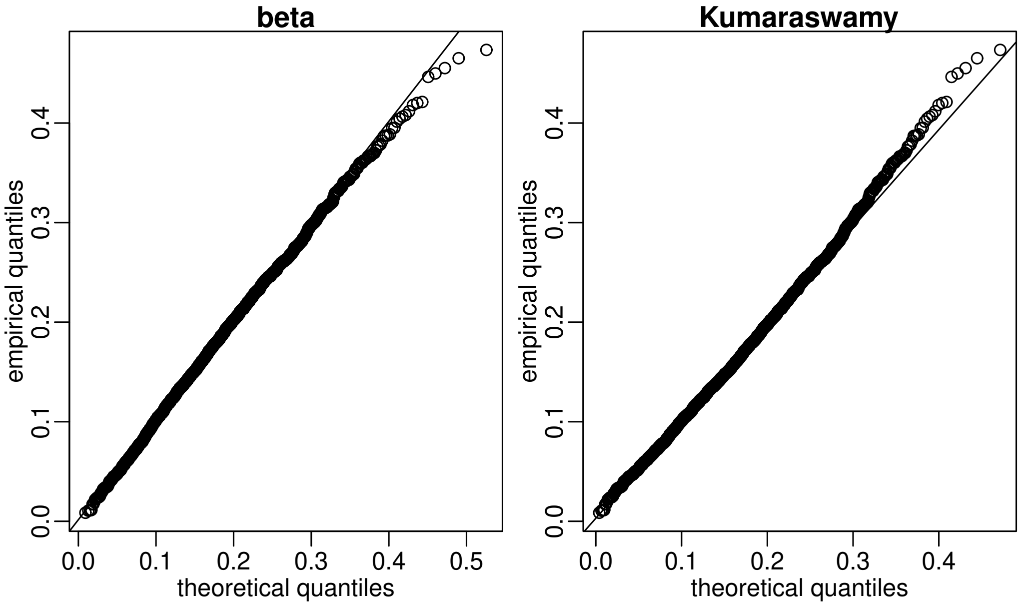
We presented an empirical analysis of Covid-19 mortality rates in the US. We considered three sample periods: (i) before, (ii) before and after, and (iii) after the beginning of the nationwide vaccination drive. The testing inferences indicated that the beta law yields a good representation of the data in the pre-vaccination period. There is also evidence in favor of such a model when mortality rates are computed using data that only cover the vaccination drive period as long as the negative impact of vaccination reach on death rates is moderate; when such an impact is strong, the univariate beta model is rejected. The beta law is also rejected by the information matrix tests when mortality rates are computed using data that cover both periods (before and after the start of the vaccination campaign). The rejection of the beta distribution in this case is due to data

Figure 11 – Histogram and fitted densities, period 3.



Source: Author

Figure 12 – Quantile-quantile plots, period 3.



Source: Author

heterogeneity. We recommend fitting alternative laws to the data and also mixtures of beta distributions.

Our results should be viewed as an initial exploration on the usefulness of information matrix tests for fractional data analysis. The tests we presented proved to be quite useful when applied to the univariate beta model. In future research, we will extend the results presented in this chapter to cover other univariate laws that are used to model fractional data (e.g., Kumaraswamy and simplex). We will also seek to extend our results to regression settings, in particular to the beta regression model introduced by Ferrari and Cribari-Neto (2004), and to dynamic beta models, such as the βARMA model introduced by Rocha and Cribari-Neto (2009), Rocha and Cribari-Neto (2017); see also Cribari-Neto, Scher and Bayer (2022) and Scher *et al.* (2020). The beta parameterization used in this chapter, which is indexed by mean and precision parameters, will be helpful for the aforementioned extensions of our results.

2 BETA REGRESSION MISSPECIFICATION TESTS

2.1 INTRODUCTION

The beta regression model introduced by Ferrari and Cribari-Neto (2004) is tailored for double bounded response variables, i.e., variables that assume values in (a, b), where a and b are known real constants $(-\infty < a < b < \infty)$. The most common situation is that in which the variable of interest assumes values in $(0, 1)$, e.g., rates and proportions. The dependent variable is assumed to follow the beta distribution indexed by its mean and a precision parameter. Both parameters of the beta distribution are impacted by explanatory variables, being connected to linear predictors that involve regressors and unknown parameters through link functions. The model thus comprises two separate submodels, one for the mean response and another for the precision.

Model selection strategies for beta regressions were developed by Bayer and Cribari-Neto (2015), Bayer and Cribari-Neto (2017), bias correction point estimation were developed by Grün, Kosmidis and Zeileis (2012) and Ospina, Cribari-Neto and Vasconcellos (2006), residuals were proposed by Espinheira, Ferrari and Cribari-Neto (2007) and Espinheira, Santos and Cribari-Neto (2017), local influence measures were derived by Ferrari, Espinheira and Cribari-Neto (2011), non-nested hypothesis testing inferences were developed by Cribari-Neto and Lucena (2015), bootstrap testing inferences were outlined by Lima and Cribari-Neto (2020), bootstrap prediction intervals were considered by Espinheira, Ferrari and Cribari-Neto (2014), and hypothesis tests that incorporate small sample corrections were developed by Bayer and Cribari-Neto (2013), Ferrari and Pinheiro (2011) and Guedes, Cribari-Neto and Espinheira (2021). Bayer, Tondolo and Müller (2018) introduced control charts based on beta regressions. The use of parametric link functions in beta regressions was studied by Canterle and Bayer (2019) and Rauber, Cribari-Neto and Bayer (2020). For details on the class of beta regression models, we refer readers to Cribari-Neto and Zeileis (2010) and Douma and Weedon (2019).

The beta regression model has been extensively used in many different areas to model the behavior of random variables with support in the standard unit interval. Using data on 124 countries in a beta regression analysis, Cribari-Neto and Souza (2013) measured the impact of average intelligence on the prevalence of atheists. Similar results for the United States based on state level data were obtained by Souza and Cribari-Neto

(2018). Souza and Cribari-Neto (2015) studied the impact of religiosity and average intelligence on homosexuality non-acceptance. Cordeiro *et al.* (2021) used beta regressions to model mortality rates in Europe during the COVID-19 first wave. A beta regression analysis of county-level excessive alcohol use, rurality, and COVID-19 case fatality rates in the United States was performed by Pro *et al.* (2021). Cribari-Neto (2023) performed a beta regression analysis of statewide COVID-19 mortality in Brazil. Swearingen *et al.* (2011) used the model to gain knowledge on ischemic stroke volume in NINDS rt-PA clinical trials.

In an empirical analysis, it is important to assess whether a fitted varying precision beta regression is correctly specified before drawing inferences and conclusions from it. The model can be misspecified in many different ways. For instance, the mean link function may be incorrectly specified, the precision link function may not be adequate, the practitioner may have failed to include an important regressor in the mean and/or precision linear predictor, there may be neglected nonlinearity, there may be neglected varying precision, and so on. The assessment of the adequacy of a fitted beta regression model is typically done using diagnostic tools, for example, by means of residual analyses based on residual half normal plots with simulated envelopes. Such analyses are useful, but involve a degree of subjectivity.

As a complement to diagnostic analyses based on residuals, our goal here is to propose hypothesis tests that can be used to assess whether the beta regression model in use is correctly specified. The null hypothesis of correct model specification is tested against the alternative hypothesis of model misspecification. The tests are based on the information matrix equality, which is known to only hold when the model is correctly specified, and are known as 'information matrix tests'. We establish the validity of information matrix tests of correct model specification for beta regressions and present closed-form expressions for information matrix test statistics in this class of models. We also present simulation evidence on the tests' finite sample performance. The numerical results obtained indicate that excellent control of the type I error frequency is achieved by using bootstrap resampling. They also show that the tests are able to detect incorrect model specification with high probability, especially when the sample size is not small. For instance, according to the numerical evidence we report, one of the tests we introduce was able to reject, at the 10% significance level, the hypothesis of correct model specification of a beta regression in which an important regressor was missing from the mean submodel with probability 0.95 based on a sample of 250 observations. The inferential tools we present allow practitioners to decide whether their fitted beta regressions are correctly specified within the framework of a hypothesis test in which it is possible to control the frequency of incorrectly concluding that the model specification is in error. Also, in the proposed tests, the probability of erroneously concluding in favor of correct model specification decays to zero as the sample size increases. Our results extend and generalize those of Silva, Cribari-Neto and Vasconcellos (2022), who presented similar tests for the beta distribution.

Since information matrix tests tend to be considerably size-distorted in finite samples, analytical corrections to the test's critical values and test statistic were obtained, respectively, by Chesher and Spady (1991) and Cribari-Neto (1997). An information matrix misspecification test statistic for the normal linear regression model was obtained by Hall (1987). It is worth noticing that misspecification is more likely to take place in beta regressions than in normal linear regressions, since the former have a more elaborate structure that includes two submodels, two link functions and two linear predictors.

We present two empirical applications of the proposed misspecification tests in beta regression analyses. They involve physiological biometrics and environmental biometrics. The former deals with modeling arms and android body fat using data collected at a public hospital in Brazil. In this first application, the tests clearly indicate for which models there is substantial evidence of incorrect model specification. Using the model with the largest tests' p-values, we construct curves that represent the impacts of age and the body mass index on the mean proportion of arms body fat. Such plots uncover important information on the different impact patterns for men and women and also for individuals with different levels of physical activity. In the second application, which relates to behavioral biometrics, the interest lies in modeling the impact of average intelligence on the mean proportion of religious disbelievers in different countries. The information matrix tests indicate that the model used in Cribari-Neto and Souza (2013) is correctly specified. We also perform the misspecification tests by only considering restrictions on mean and precision covariates directly related to average intelligence. They show that the only model for which the specification of such effects seems correct is the loglog model. This showcases an advantage of information matrix tests: one can focus on a chosen subset of restrictions, e.g., those related to the covariates that are of most relevance to the analysis.

The tests we introduce can be viewed as an alternative to that in Pereira and Cribari-Neto (2014). The authors present an adaptation of RESET test introduced by Ramsey (1969) to inflated beta regressions, which encompass the class of beta regressions. The test consists of adding powers of the fitted mean linear predictor to the mean submodel, and testing their exclusion using, e.g., the likelihood ratio test. Our tests have several advantages relative to the RESET test. First, they are obtained from an identity that is directly related to the correct specification of the model. Second, they allow practitioners to focus on a subset of restrictions related to a selected subset of covariates, as exemplified in one of our empirical applications. Third, as revealed by our simulation results, the RESET test lacks power in important settings. Finally, the null hypothesis under evaluation is well defined in the tests we develop since we test whether the information matrix equality holds. It is not clear, however, what is the null hypothesis of the RESET test in beta regressions. In the classical linear regression model, one tests whether the mean of error term conditional on the regressors is null. It is not clear how that translates to beta regression settings.

We also note that there are formulations of the beta regression model that are more general than that used in this chapter. For instance, Rauber, Cribari-Neto and Bayer (2020) consider the use a parametric link function in the mean submodel. We chose to work with the standard varying precision beta regression because it is most commonly used formulation of the model in empirical studies.

Finally, it is worth noting that information matrix tests have been derived for several statistical and econometric models, e.g., the Gaussian linear regression model (HALL, 1987), binary data models (ORME, 1988), linear regressions with autoregressive and moving average errors (FURNO, 1996), logistic regressions (ZHANG, 2001), beta-binomial models (CAPANU; PRESNELL, 2008), the negative binomial law (CHUA; ONG, 2013), and copulas (PROKHOROV; SCHEPSMEIER; ZHU, 2019). We add to the literature by extending such tests to the class of varying precision beta regressions.

The chapter unfolds as follows. The beta regression model is briefly presented in Section 2.2. In Section 2.3, information matrix misspecification tests for beta regressions are obtained. Monte Carlo simulation evidence is presented in Section 2.4. We present numerical results on the tests' null (size) and non-null (power) behavior. Empirical applications are presented and discussed in Section 2.5. Finally, concluding remarks are offered in Section 2.6.

2.2 THE BETA REGRESSION MODEL

Let Y be a random variable that follows the beta distribution with density function given by Ferrari and Cribari-Neto (2004)

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

$0<\mu<1$, $\phi>0$, where $\mathbb{E}(Y)=\mu$ and ϕ is a precision parameter since. Here, $\mathrm{Var}(Y) = \mu(1-\mu)/(1+\phi)$ which, for fixed μ, decreases as ϕ increases. We write $Y \sim \mathcal{B}(\mu,\phi)$.

Let $Y_1,\ldots,Y_n$ be independent beta-distributed random variables such that $Y_i \sim \mathcal{B}(\mu_i,\phi_i)$, $i=1,\ldots,n$. The varying precision beta regression model is given by

$$g_1(\mu_i) = \sum_{j=1}^{p} \beta_j x_{ij} = \eta_{1i} \quad \text{and} \quad g_2(\phi_i) = \sum_{j=1}^{q} \delta_j z_{ij} = \eta_{2i},$$

where $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)^\top \in \mathbb{R}^p$ and $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_q)^\top \in \mathbb{R}^q$ are unknown parameter vectors $(p+q=k<n)$, $\boldsymbol{\eta}_1 = (\eta_{11},\ldots,\eta_{1n})^\top$ and $\boldsymbol{\eta}_2 = (\eta_{21},\ldots,\eta_{2n})^\top$ are linear predictor vectors, and $x_{i1} = z_{i1} = 1\forall i$. Also, $x_{i2},\ldots,x_{ip}$ and $z_{i2},\ldots,z_{iq}$ are the covariates used in the mean and precision submodels, respectively. Here, $g_1:(0,1)\to\mathbb{R}$ and $g_2:(0,\infty)\to\mathbb{R}$ are strictly monotonic and three times differentiable link functions. Common choices for the mean (precision) link function are logit, probit, loglog, cloglog and cauchit (log). When all observations share the same precision, i.e., when $\phi_1=\cdots=\phi_n=\phi$, the above model reduces to its fixed precision formulation. In this case, $Y_i \sim \mathcal{B}(\mu_i,\phi)$, $i=1,\ldots,n$.

In what follows, $\boldsymbol{Y}$ and $\boldsymbol{y}$ denote n-vectors of beta-distributed random variables and their realizations, respectively. Let$\boldsymbol{\theta} = (\boldsymbol{\beta}^\top,\boldsymbol{\delta}^\top)^\top \in \Theta \subseteq \mathbb{R}^k$ be the vector containing all regression coefficients, where Θ denotes the parameter space. The log-likelihood function for $Y_1,\ldots,Y_n$ with observed values $y_1,\ldots,y_n$ is

$$\ell \equiv \ell(\boldsymbol{\theta};\boldsymbol{y}) \equiv \ell(\boldsymbol{\beta},\boldsymbol{\delta};\boldsymbol{y}) = \sum_{i=1}^{n} \ell(\mu_i,\phi_i;y_i),$$

where $\ell(\mu_i,\phi_i;y_i) = \log\Gamma(\phi_i) - \log\Gamma(\mu_i\phi_i) - \log\Gamma((1-\mu_i)\phi_i) + (\mu_i\phi_i-1)y_i^* + (\phi_i-2)y_i^\dagger$, with $y_i^* = \log(y_i/(1-y_i))$ and $y_i^\dagger = \log(1-y_i)$. It can be easily verified that $\mathbb{E}(Y_i^*) = \psi(\mu_i\phi_i) - \psi((1-\mu_i)\phi_i)$, $\mathbb{E}(Y_i^\dagger) = \psi((1-\mu_i)\phi_i) - \psi(\phi_i)$, where ψ is the digamma function. We will denote $\mathbb{E}(Y_i^*)$ and $\mathbb{E}(Y_i^\dagger)$ by μ_i^* and $\mu_i^\dagger$, respectively.

Let $\boldsymbol{U} \equiv \boldsymbol{U}(\boldsymbol{\theta}) = (\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta})^\top, \boldsymbol{U}_{\boldsymbol{\delta}}(\boldsymbol{\theta})^\top)^\top = (\partial\ell(\boldsymbol{\theta};\boldsymbol{y})/\partial\boldsymbol{\beta}^\top, \partial\ell(\boldsymbol{\theta};\boldsymbol{y})/\partial\boldsymbol{\delta}^\top)^\top = \nabla\ell(\boldsymbol{\theta};\boldsymbol{y})$ denote the score function. The log-likelihood derivatives with respect to the rth and Rth components of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are, respectively,

$$U_{\beta_r}(\boldsymbol{\theta}) = \frac{\partial\ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial\beta_r} = \sum_{i=1}^{n} \frac{\partial\ell_i(\mu_i,\phi_i;y_i)}{\partial\mu_i}\frac{d\mu_i}{d\eta_{1i}}\frac{\partial\eta_{1i}}{\partial\beta_r},$$
$$U_{\delta_R}(\boldsymbol{\theta}) = \frac{\partial\ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial\delta_R} = \sum_{i=1}^{n} \frac{\partial\ell_i(\mu_i,\phi_i;y_i)}{\partial\phi_i}\frac{d\phi_i}{d\eta_{2i}}\frac{\partial\eta_{2i}}{\partial\delta_R},$$

where

$$\frac{\partial\ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial\beta_r} = \sum_{i=1}^{n} \phi_i(y_i^* - \mu_i^*)\frac{1}{g_1'(\mu_i)}x_{ir},$$
$$\frac{\partial\ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial\delta_R} = \sum_{i=1}^{n} [\mu_i(y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)]\frac{1}{g_2'(\phi_i)}z_{iR},$$

$r = 1,\ldots,p$ and $R = 1,\ldots,q$. Here, $g_1'(\mu_i)$ and $g_2'(\phi_i)$ denote the derivatives of g_1 and g_2 with respect to μ_i and ϕ_i, respectively.

The maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\delta}}^\top)^\top = \arg\max_{\boldsymbol{\theta}\in\mathbb{R}^{p+q}} \ell(\boldsymbol{\theta};\boldsymbol{y})$ cannot be expressed in closed-form. Parameter estimates are usually obtained by numerically maximizing the beta regression model log-likelihood function with the aid of a nonlinear optimization algorithm. In what follows, we will use the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with analytical gradient; for details, see Nocedal and Wright (2006).

Fisher's information matrix for a single observation, say $\dot{B}_i(\boldsymbol{\theta})$, is the $k \times k$ matrix given by the expected value of the individual log-likelihood derivative outer product: $\dot{B}_i(\boldsymbol{\theta}) = \mathbb{E}(\partial\ell(\boldsymbol{\theta};Y_i)/\partial\boldsymbol{\theta} \times \partial\ell(\boldsymbol{\theta};Y_i)/\partial\boldsymbol{\theta}^\top)$. We write $\dot{B}_i(\boldsymbol{\theta})$ as

$$\dot{B}_i(\boldsymbol{\theta}) = \begin{bmatrix} \dot{B}_i^{\boldsymbol{\beta\beta}} & \dot{B}_i^{\boldsymbol{\beta\delta}} \\ & \\ \dot{B}_i^{\boldsymbol{\delta\beta}} & \dot{B}_i^{\boldsymbol{\delta\delta}} \end{bmatrix}, \tag{2.1}$$

where $\dot{B}_i^{\boldsymbol{\beta\beta}}$, $\dot{B}_i^{\boldsymbol{\beta\delta}} = (\dot{B}_i^{\boldsymbol{\delta\beta}})'$ and $\dot{B}_i^{\boldsymbol{\delta\delta}}$ are matrices with dimensions $p \times p$, $p \times q$ and $q \times q$, respectively. For $r, s = 1,\ldots,p$ and $R, S = 1,\ldots,q$, let $\dot{B}_i^{(r,s)}$, $\dot{B}_i^{(r,R)}$, $\dot{B}_i^{(R,S)}$ denote the (r,s), (r,R) and (R,S) elements of the matrices $\dot{B}_i^{\boldsymbol{\beta\beta}}$, $\dot{B}_i^{\boldsymbol{\beta\delta}}$ and $\dot{B}_i^{\boldsymbol{\delta\delta}}$, respectively. They can be expressed as $\dot{B}_i^{(r,s)} = -\phi_i^2 w_i \left(d\mu_i/d\eta_{1i}\right)^2 x_{ir}x_{is}$, $\dot{B}_i^{(r,R)} = -c_i\left(d\mu_i/d\eta_{1i}\right)\left(d\phi_i/d\eta_{2i}\right)x_{ir}z_{iR}$ and $\dot{B}_i^{(R,S)} = -p_i\left(d\phi_i/d\eta_{2i}\right)^2 z_{iR}z_{iS}$. The expressions for the quantities w_i, c_i and p_i can be found in Appendix B.

2.3 BETA REGRESSION MISSPECIFICATION TESTS

Our interest lies in testing the null hypothesis that the beta regression model is correctly specified against the alternative hypothesis that the model specification is in error. We consider the information matrix test introduced in full generality by White (1982). Let $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^\top, \boldsymbol{\delta}_0^\top)^\top$ be the true parameter value. We say that the beta regression model is correctly specified if Y_i follows the beta law with parameter vector $\boldsymbol{\theta}_0$ $\forall i$.

2.3.1 Misspecification tests for beta regressions

Assume that the limits $A(\boldsymbol{\theta})$ and $B(\boldsymbol{\theta})$ given by $A(\boldsymbol{\theta}) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \dot{A}_i(\boldsymbol{\theta})$ and $B(\boldsymbol{\theta}) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \dot{B}_i(\boldsymbol{\theta})$ exist, where $\dot{A}_i(\boldsymbol{\theta}) = \mathbb{E}(\partial^2\ell(\boldsymbol{\theta}; Y_i)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top)$. The information matrix equality $B(\boldsymbol{\theta}_0) = -A(\boldsymbol{\theta}_0)$ holds when the model is correctly specified. It is valid for $\forall\boldsymbol{\theta}_0 \in \Theta$. We then wish to test $\mathcal{H}_0 : A(\boldsymbol{\theta}_0) + B(\boldsymbol{\theta}_0) = O_{k\times k}$ against $\mathcal{H}_1 : A(\boldsymbol{\theta}_0) + B(\boldsymbol{\theta}_0) \neq O_{k\times k}$, where $O_{k\times k}$ denotes a k-dimensional square matrix of zeros. Rejection of the null hypothesis is evidence of model misspecification.

In what follows, we will present information matrix test statistics that can be used to test the correct beta regression model specification. Let $A_i(\boldsymbol{\theta}) = \partial\ell^2(\boldsymbol{\theta}; Y_i)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ and $A_n(\boldsymbol{\theta}; \boldsymbol{Y}) = n^{-1}\sum_{i=1}^n A_i(\boldsymbol{\theta})$. We write $A_n(\boldsymbol{\theta}; \boldsymbol{Y})$ as

$$A_n(\boldsymbol{\theta}; \boldsymbol{Y}) = \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} A_i^{\boldsymbol{\beta\beta}} & A_i^{\boldsymbol{\beta\delta}} \\ \\ A_i^{\boldsymbol{\delta\beta}} & A_i^{\boldsymbol{\delta\delta}} \end{bmatrix},$$

where $A_i^{\boldsymbol{\beta\beta}}$, $A_i^{\boldsymbol{\beta\delta}} = (A_i^{\boldsymbol{\delta\beta}})'$ and $A_i^{\boldsymbol{\delta\delta}}$ are matrices with dimensions $p\times p$, $p\times q$ and $q\times q$, respectively. For $r,s = 1,\ldots,p$ and $R,S = 1,\ldots,q$, let $A_i^{(r,s)}$, $A_i^{(r,R)}$ and $A_i^{(R,S)}$ denote the (r,s), (r,R) and (R,S) elements of the matrices $A_i^{\boldsymbol{\beta\beta}}$, $A_i^{\boldsymbol{\beta\delta}}$ and $A_i^{\boldsymbol{\delta\delta}}$, respectively. Such elements are expressed by

$$\begin{aligned}
A_i^{(r,s)} &= \frac{\partial^2\ell(\boldsymbol{\theta}; Y_i)}{\partial\beta_r\partial\beta_s} = \left[-\phi_i^2 w_i \frac{d\mu_i}{d\eta_{1i}} - \phi_i(y_i^* - \mu_i^*)\frac{d}{d\mu_i}\frac{d\mu_i}{d\eta_{1i}}\right]\left(\frac{d\mu_i}{d\eta_{1i}}\right) x_{ir}x_{is},\\
A_i^{(r,R)} &= \frac{\partial^2\ell(\boldsymbol{\theta}; Y_i)}{\partial\beta_r\partial\delta_R} = [-c_i + (y_i^* - \mu_i^*)]\frac{d\phi_i}{d\eta_{2i}}\frac{d\mu_i}{d\eta_{1i}} x_{ir}z_{iR},\\
A_i^{(R,S)} &= \frac{\partial^2\ell(\boldsymbol{\theta}; Y_i)}{\partial\delta_R\partial\delta_S} = \Bigg\{-p_i\left(\frac{d\phi_i}{d\eta_{2i}}\right)^2 + \left[\mu_i(y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)\right]\\
&\qquad \times \left(\frac{d}{d\phi_i}\frac{d\phi_i}{d\eta_{2i}}\right)\frac{d\phi_i}{d\eta_{2i}}\Bigg\} z_{iR}z_{iS}.
\end{aligned}$$

Expressions for the above derivatives are given in Appendix B

Also, let $B_i(\boldsymbol{\theta}) = \partial\ell(\boldsymbol{\theta};Y_i)/\partial\boldsymbol{\theta} \times \partial\ell(\boldsymbol{\theta};Y_i)/\partial\boldsymbol{\theta}^\top)$ and $B_n(\boldsymbol{\theta};\boldsymbol{Y}) = n^{-1}\sum_{i=1}^n B_i(\boldsymbol{\theta})$. We write $B_n(\boldsymbol{\theta};\boldsymbol{Y})$ as

$$B_n(\boldsymbol{\theta};\boldsymbol{Y}) = \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} B_i^{\boldsymbol{\beta\beta}} & B_i^{\boldsymbol{\beta\delta}} \\ \\ B_i^{\boldsymbol{\delta\beta}} & B_i^{\boldsymbol{\delta\delta}} \end{bmatrix},$$

where $B_i^{\boldsymbol{\beta\beta}}$, $B_i^{\boldsymbol{\beta\delta}} = (B_i^{\boldsymbol{\delta\beta}})'$ and $B_i^{\boldsymbol{\delta\delta}}$ are matrices with dimensions $p\times p$, $p\times q$ and $q\times q$, respectively. For $r,s = 1,\dots,p$ and $R,S = 1,\dots,q$, let $B_i^{(r,s)}$, $B_i^{(r,R)}$ and $B_i^{(R,S)}$ denote the (r,s), (r,R) and (R,S) elements of the matrices $B_i^{\boldsymbol{\beta\beta}}$, $B_i^{\boldsymbol{\beta\delta}}$ and $B_i^{\boldsymbol{\delta\delta}}$, respectively. They are given by

$$\begin{aligned}
B_i^{(r,s)} &= \frac{\partial\ell(\boldsymbol{\theta};Y_i)}{\partial\beta_r}\times\frac{\partial\ell(\boldsymbol{\theta};Y_i)}{\partial\beta_s} = \left[\phi_i(y_i^* - \mu_i^*)\frac{d\mu_i}{d\eta_{1i}}\right]^2 x_{ir}x_{is},\\
B_i^{(r,R)} &= \frac{\partial\ell(\boldsymbol{\theta};Y_i)}{\partial\beta_r}\times\frac{\partial\ell(\boldsymbol{\theta};Y_i)}{\partial\delta_R} = \phi_i\,(y_i^* - \mu_i^*)\Big[\mu_i(y_i^* - \mu_i^*)\\
&\qquad + (y_i^\dagger - \mu_i^\dagger)\Big]\frac{d\phi_i}{d\eta_{2i}}\frac{d\mu_i}{d\eta_{1i}}x_{ir}z_{iR},\\
B_i^{(R,S)} &= \frac{\partial\ell(\boldsymbol{\theta};Y_i)}{\partial\delta_R}\times\frac{\partial\ell(\boldsymbol{\theta};Y_i)}{\partial\delta_S} = \Big[\mu_i(y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)\Big]^2\left(\frac{d\phi_i}{d\eta_{2i}}\right)^2 z_{iR}z_{iS}.
\end{aligned}$$

Expressions for the derivatives in the above formulas can be found in Appendix B. As shown in subsection 2.3.2, $A_n(\boldsymbol{\theta};\boldsymbol{Y})$ and $B_n(\boldsymbol{\theta};\boldsymbol{Y})$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n$ are consistent estimators of $A(\boldsymbol{\theta}_0)$ and $B(\boldsymbol{\theta}_0)$, respectively.

Let $C_i(\boldsymbol{\theta}) = A_i(\boldsymbol{\theta}) + B_i(\boldsymbol{\theta})$ and let $\boldsymbol{d}_i(\boldsymbol{\theta}) = \text{vech}\,(C_i(\boldsymbol{\theta}))$ be a vector of dimension $K\times 1$, where $K = k(k+1)/2$. Here, vech is the operator that, when applied to a square matrix, returns the vector formed by its lower triangular portion (including the diagonal).

Also, let

$$\boldsymbol{D_n}(\boldsymbol{\theta}) \equiv \boldsymbol{D_n}(\boldsymbol{\theta};\boldsymbol{Y}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{d}_i(\boldsymbol{\theta}).$$

For $r,s = 1,\dots,p$; $R,S = 1,\dots,q$, let $C_i^{(r,s)}$, $C_i^{(r,R)}$ and $C_i^{(R,S)}$ denote the (r,s), (r,R) and

(R,S) elements of the matrix $C_i(\boldsymbol{\theta})$. These elements are expressed by

$$
\begin{aligned}
C_i^{(r,s)} =& \left\{ \phi_i \frac{d\mu_i}{d\eta_{1i}} \left[-\phi_i w_i \frac{d\mu_i}{d\eta_{1i}} + (y_i^* - \mu_i^*) \frac{d}{d\mu_i} \frac{d\mu_i}{d\eta_{1i}} + \phi_i (y_i^* - \mu_i^*)^2 \frac{d\mu_i}{d\eta_{1i}} \right] \right\} \\
& \times x_{ir} x_{is}, \\
C_i^{(r,R)} =& \frac{d\mu_i}{d\eta_{1i}} \frac{d\phi_i}{d\eta_{2i}} \Big\{ \phi_i (y_i^* - \mu_i^*) \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger) \right] \\
& + (y_i^* - \mu_i^*) - c_i \Big\} x_{ir} z_{iR}, \\
C_i^{(R,S)} =& \left\{ -p_i \left(\frac{d\phi_i}{d\eta_{2i}} \right)^2 + \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger) \right] \left(\frac{d}{d\phi_i} \frac{d\phi_i}{d\eta_{2i}} \right) \frac{d\phi_i}{d\eta_{2i}} \right. \\
& \left. + \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger) \right]^2 \left(\frac{d\phi_i}{d\eta_{2i}} \right)^2 \right\} z_{iR} z_{iS}.
\end{aligned}
$$

Appendix B contains expressions for the above derivatives.

The vector $\boldsymbol{D_n}(\boldsymbol{\theta};\boldsymbol{Y}) = \text{vech}(A_n(\boldsymbol{\theta};\boldsymbol{Y}) + B_n(\boldsymbol{\theta};\boldsymbol{Y}))$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n$ is known as 'the vector of restrictions'. It contains the lower triangular elements of $A_n(\boldsymbol{\theta};\boldsymbol{Y}) + B_n(\boldsymbol{\theta};\boldsymbol{Y})$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n$. Suppose the limits Suppose the limits

$$\nabla \boldsymbol{D}(\boldsymbol{\theta}) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \mathbb{E}\left(\partial \boldsymbol{d}_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^\top \right)$$

and

$$
\begin{aligned}
V(\boldsymbol{\theta}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} & \left(\boldsymbol{d}_i(\boldsymbol{\theta}) - \nabla \boldsymbol{D}(\boldsymbol{\theta}) A(\boldsymbol{\theta})^{-1} \nabla \ell(\boldsymbol{\theta}; Y_i) \right) \\
& \times \left(\boldsymbol{d}_i(\boldsymbol{\theta}) - \nabla \boldsymbol{D}(\boldsymbol{\theta}) A(\boldsymbol{\theta})^{-1} \nabla \ell(\boldsymbol{\theta}; Y_i) \right)^\top
\end{aligned}
$$

exist. When the model is correctly specified, the asymptotic distribution of the random vector $\sqrt{n}\boldsymbol{D_n}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{Y})$ is, under some regularity conditions, multivariate normal with zero mean and covariance matrix $V(\boldsymbol{\theta}_0)$; see Theorem 1 in the next subsection. Therefore, if $V_n(\hat{\boldsymbol{\theta}}_n)$ is consistent for $V(\boldsymbol{\theta}_0)$, then, under correct model specification,

$$n \boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n; \boldsymbol{Y})^\top V_n(\hat{\boldsymbol{\theta}}_n)^{-1} \boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n; \boldsymbol{Y})$$

is asymptotically chi-squared distributed with K degrees of freedom (χ^2_K).

A natural estimator for $V(\boldsymbol{\theta}_0)$ is

$$
\begin{aligned}
V_{n1}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} & \left(\boldsymbol{d}_i(\boldsymbol{\theta}) - \nabla \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y}) A_n(\boldsymbol{\theta}; \boldsymbol{Y})^{-1} \nabla \ell(\boldsymbol{\theta}; Y_i) \right) \\
& \times \left(\boldsymbol{d}_i(\boldsymbol{\theta}) - \nabla \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y}) A_n(\boldsymbol{\theta}; \boldsymbol{Y})^{-1} \nabla \ell(\boldsymbol{\theta}; Y_i) \right)^\top,
\end{aligned}
$$

evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n$, where $\nabla \boldsymbol{D}_n(\boldsymbol{\theta};\boldsymbol{Y}) = \partial \boldsymbol{D}_n(\boldsymbol{\theta};\boldsymbol{Y})/\partial \boldsymbol{\theta}^\top$. Expressions for the elements of the matrix $\nabla \boldsymbol{D}_n(\boldsymbol{\theta};\boldsymbol{Y})$ for the beta regression model are presented in Appendix B.

We will now obtain a closed-form expression for $\nabla \boldsymbol{D}_n(\boldsymbol{\theta};\boldsymbol{Y})$ in beta regressions that only involve simple matrix operations. The only caveat is that the ordering of the elements in $\boldsymbol{d}_i(\boldsymbol{\theta})$ must obey a well-defined rule which differs from the one previously stated. We will return to that later.

The number of rows of the matrix $\nabla \boldsymbol{D}_n(\boldsymbol{\theta};\boldsymbol{Y})$ equals the number of elements of the vector $\boldsymbol{d}_i(\boldsymbol{\theta})$, i.e., $K = k(k+1)/2$, where $k = p+q$. Note that

$$K = \frac{(p+q)(p+q+1)}{2} = \frac{p(p+1)}{2} + \frac{q(q+1)}{2} + pq.$$

Let $\bar{P}$, $\bar{Q}$, and $\bar{R}$ denote, respectively, the three terms in the sum above.

Let X be the $n \times p$ matrix of the mean regressors, whose p columns are $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$, where each $\boldsymbol{x}_j$ is an $n \times 1$ vector, $j = 1, \ldots, p$. Let $\boldsymbol{x}_{j,k}$, for each fixed j and $k = 1, \ldots, p$, be the $n \times 1$ vector that equals the direct (Hadamard) product of $\boldsymbol{x}_j$ by $\boldsymbol{x}_k$ (i.e., ith component of the product is the product of the ith components). We now define the matrix $X * X$, of dimension $n \times \bar{P}$, with columns given as

$$X * X = [\boldsymbol{x}_{1,1} \ldots \boldsymbol{x}_{p,1} \; \boldsymbol{x}_{2,2} \ldots \boldsymbol{x}_{p,2} \; \boldsymbol{x}_{3,3} \ldots \boldsymbol{x}_{p,p-1} \; \boldsymbol{x}_{p,p}].$$

Let Z be the $n \times q$ matrix of precision regressors, whose q columns are $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_q$, where each $\boldsymbol{z}_j$ is an $n \times 1$ vector, $j = 1, \ldots, q$. Let $\boldsymbol{z}_{j,k}$, for each fixed j and $k = 1, \ldots, q$, be the $n \times 1$ vector given by the direct (Hadamard) product between $\boldsymbol{z}_j$ and $\boldsymbol{z}_k$. We now define the matrix $Z * Z$, of dimension $n \times \bar{Q}$, with columns given as

$$Z * Z = [\boldsymbol{z}_{1,1} \ldots \boldsymbol{z}_{p,1} \; \boldsymbol{z}_{2,2} \ldots \boldsymbol{z}_{p,2} \; \boldsymbol{z}_{3,3} \ldots \boldsymbol{z}_{p,p-1} \; \boldsymbol{z}_{p,p}].$$

Consider the $\bar{R}$ vectors $\boldsymbol{v}_{j,k}$, $j = 1, \ldots, p$ and $k = 1, \ldots, q$, where each $\boldsymbol{v}_{j,k}$ is the $n \times 1$ vector representing the direct product of $\boldsymbol{z}_j$ by $\boldsymbol{x}_k$. Define the matrix $Z * X$, of dimension $n \times \bar{R}$, as

$$Z * X = [\boldsymbol{v}_{1,1} \; \boldsymbol{v}_{2,1} \; \ldots \boldsymbol{v}_{p,1} \; \boldsymbol{v}_{1,2} \ldots \boldsymbol{v}_{p,2} \ldots \boldsymbol{v}_{1,q} \ldots \boldsymbol{v}_{p,q}].$$

Now define the matrix W, of dimension $3n \times K$, as

$$W = \begin{bmatrix} X * X & O_{n \times \bar{Q}} & O_{n \times \bar{R}} \\ O_{n \times \bar{P}} & Z * Z & O_{n \times \bar{R}} \\ O_{n \times \bar{P}} & O_{n \times \bar{Q}} & Z * X \end{bmatrix},$$

where $O_{r\times c}$ denotes an $r\times c$ matrix of zeros. For each $r,s=1,\dots,p$, the derivative of $C_i^{(r,s)}$ with respect to β_t is an expression of the type $\alpha_i^{\boldsymbol{\beta},\boldsymbol{\beta},\boldsymbol{\beta}} x_{ir}\, x_{is}\, x_{it}$, where the coefficient $\alpha_i^{\boldsymbol{\beta},\boldsymbol{\beta},\boldsymbol{\beta}}$ does not involve r,s,t. Similarly, the derivative of $C_i^{(r,s)}$ with respect to δ_R is an expression of the form $\alpha_i^{\boldsymbol{\beta},\boldsymbol{\beta},\boldsymbol{\delta}} x_{ir}\, x_{is}\, z_{iR}$, where the coefficient $\alpha_i^{\boldsymbol{\beta},\boldsymbol{\beta},\boldsymbol{\delta}}$ does not involve r,s,R. The derivatives of $C_i^{(r,R)}$ and $C_i^{(R,S)}$ can be expressed in a similar fashion.

Let $\Lambda^{\boldsymbol{\beta}}_{\boldsymbol{\beta},\boldsymbol{\beta}}$, $\Lambda^{\boldsymbol{\delta}}_{\boldsymbol{\beta},\boldsymbol{\beta}}$, $\Lambda^{\boldsymbol{\beta}}_{\boldsymbol{\delta},\boldsymbol{\delta}}$, $\Lambda^{\boldsymbol{\delta}}_{\boldsymbol{\delta},\boldsymbol{\delta}}$, $\Lambda^{\boldsymbol{\beta}}_{\boldsymbol{\beta},\boldsymbol{\delta}}$, and $\Lambda^{\boldsymbol{\delta}}_{\boldsymbol{\beta},\boldsymbol{\delta}}$ be $n\times n$ diagonal matrices such that the diagonal entries of $\Lambda^{\boldsymbol{\beta}}_{\boldsymbol{\beta},\boldsymbol{\beta}}$ are the coefficients of the derivatives of $C_i^{(r,s)}$ with respect to $\boldsymbol{\beta}$, and so on. We now assemble the matrix U, of dimension $3n\times k$ and defined using six blocks, as

$$U=\begin{bmatrix}\Lambda^{\boldsymbol{\beta}}_{\boldsymbol{\beta},\boldsymbol{\beta}}X & \Lambda^{\boldsymbol{\delta}}_{\boldsymbol{\beta},\boldsymbol{\beta}}Z\\ \Lambda^{\boldsymbol{\beta}}_{\boldsymbol{\delta},\boldsymbol{\delta}}X & \Lambda^{\boldsymbol{\delta}}_{\boldsymbol{\delta},\boldsymbol{\delta}}Z\\ \Lambda^{\boldsymbol{\beta}}_{\boldsymbol{\beta},\boldsymbol{\delta}}X & \Lambda^{\boldsymbol{\delta}}_{\boldsymbol{\beta},\boldsymbol{\delta}}Z\end{bmatrix}.$$

It is now possible to express $\nabla \boldsymbol{D}_n(\boldsymbol{\theta};\boldsymbol{Y})$ as the $K\times k$ matrix given by

$$\nabla \boldsymbol{D}_n(\boldsymbol{\theta};\boldsymbol{Y}) = n^{-1}W^\top U. \tag{2.2}$$

This matrix involves unknown parameters. In $V_{n1}(\hat{\boldsymbol{\theta}}_n)$, we use $\nabla \boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n;\boldsymbol{Y}) = n^{-1}W^\top \hat{U}$, where $\hat{U}$ is obtained by replacing all entries of the aforementioned diagonal matrices by their maximum likelihood estimates.

An important caveat relates to the way the columns of W are defined since it provides a well-defined ordering to be used for the components of the vector $\boldsymbol{d}_i(\boldsymbol{\theta})$. The manner in which we sequentially use $X*X$, then $Z*Z$, then $Z*X$ indicates the vector $\boldsymbol{d}_i(\boldsymbol{\theta})$ should no longer be defined simply as $\mathrm{vech}(C_i)$, but as the vector

$$\begin{bmatrix}\mathrm{vech}(C_i^{\boldsymbol{\beta},\boldsymbol{\beta}})\\ \mathrm{vech}(C_i^{\boldsymbol{\delta},\boldsymbol{\delta}})\\ \mathrm{vec}(C_i^{\boldsymbol{\beta},\boldsymbol{\delta}})\end{bmatrix},$$

where $C_i^{\boldsymbol{\beta},\boldsymbol{\beta}} = A_i^{\boldsymbol{\beta},\boldsymbol{\beta}} + B_i^{\boldsymbol{\beta},\boldsymbol{\beta}}$, etc., vec being the operator that yields a column vector when applied to a matrix by stacking its columns one underneath the other. Note that the orders of the columns within each of the products we defined were strategically chosen so that the order of the elements of the vector $\boldsymbol{d}_i(\boldsymbol{\theta})$ has this relatively simple form.

We will show in the next subsection that $V_{n1}(\hat{\boldsymbol{\theta}}_n)$ is a consistent estimator of $V(\boldsymbol{\theta}_0)$ under correct model specification. The following test statistic can then be used to test the correct beta regression model specification:

$$\zeta_1 = n\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n)^\top V_{n1}(\hat{\boldsymbol{\theta}}_n)^{-1}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n).$$

Under $\mathcal{H}_0$, ζ_1 is asymptotically distributed as χ^2_K. The test is performed using asymptotic critical values. We reject the null hypothesis at significance level $\alpha \in (0,1)$ if $\zeta_1 > \chi^2_{K;1-\alpha}$, where $\chi^2_{K;1-\alpha}$ is the $1-\alpha$ quantile of the χ^2_K distribution.

Lancaster (1984) proposed a consistent estimator of the matrix $V(\boldsymbol{\theta})$ that does not involve third-order log-likelihood derivatives. They used the fact that, under $\mathcal{H}_0$, $\mathbb{E}(\partial \boldsymbol{d}_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top + \boldsymbol{d}_i(\boldsymbol{\theta})\nabla\ell(\boldsymbol{\theta};Y_i))^\top) = O_{K\times k}$, $i = 1,\ldots,n$ (LANCASTER, 1984). Therefore, under $\mathcal{H}_0$,

$$-\nabla\boldsymbol{D}(\boldsymbol{\theta}_0) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\mathbb{E}(\boldsymbol{d}_i(\boldsymbol{\theta}_0)\nabla\ell(\boldsymbol{\theta}_0;Y_i)^\top).$$

Let $L_n(\boldsymbol{\theta};\boldsymbol{Y}) = -n^{-1}\sum_{i=1}^n \boldsymbol{d}_i(\boldsymbol{\theta})\nabla\ell(\boldsymbol{\theta};Y_i)^\top$. Their estimator of $V(\boldsymbol{\theta}_0)$ is

$$V_{n2}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\Big(\boldsymbol{d}_i(\boldsymbol{\theta}) + L_n(\boldsymbol{\theta};\boldsymbol{Y})B_n(\boldsymbol{\theta};\boldsymbol{Y})^{-1}\nabla\ell(\boldsymbol{\theta};Y_i)\Big) \times \Big(\boldsymbol{d}_i(\boldsymbol{\theta}) + L_n(\boldsymbol{\theta};\boldsymbol{Y})B_n(\boldsymbol{\theta};\boldsymbol{Y})^{-1}\nabla\ell(\boldsymbol{\theta};Y_i)\Big)^\top$$

evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n$.

We will show in the next subsection that this is a consistent estimator of $V(\boldsymbol{\theta}_0)$ under correct model specification in beta regressions. The corresponding information matrix test statistic is

$$\zeta_2 = n\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n)^\top V_{n2}(\hat{\boldsymbol{\theta}}_n)^{-1}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n).$$

Under $\mathcal{H}_0$, ζ_2 is distributed as χ^2_K asymptotically, and we reject the null hypothesis of correct model specification if $\zeta_2 > \chi^2_{K;1-\alpha}$. We note that the computation of $V_{n2}(\hat{\boldsymbol{\theta}}_n)$ does not require $\nabla\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n;\boldsymbol{Y})$.

The beta regression misspecification tests we presented may be size-distorted in finite samples. Such distortions vanish as the sample size increases, but they may be sizeable when the number of observations is not large. Better control of the type I error frequency can be achieved by using critical values obtained from a parametric bootstrap resampling scheme instead of asymptotic critical values since the test statistics' exact null distributions may be poorly approximated by their asymptotic counterpart when n is not large. Bootstrap-based testing inferences can be performed as follows. For $i = 1,2$:

1. Using the original sample $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$, compute ζ_i.
2. Obtain a sample of size n, say $\boldsymbol{Y}_b^* = (Y_1^*, \ldots, Y_n^*)^\top$, by independently sampling from the beta law with μ_i and ϕ_i replaced, respectively, with $\hat{\mu}_i = g_1^{-1}(\hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_p x_{ip})$ and $\hat{\phi}_i = g_2^{-1}(\hat{\delta}_1 + \hat{\delta}_2 z_{i2} + \cdots + \hat{\delta}_q z_{iq})$, $i = 1, \ldots, n$.
3. Using $\boldsymbol{Y}_b^*$, compute $\zeta_{i,b}^*$.
4. Execute steps (2) and (3) B times.
5. Reject the null hypothesis of correct model specification at significance level α if ζ_i exceeds the $1-\alpha$ quantile of $\zeta_{i,1}^*, \ldots, \zeta_{i,B}^*$.

The use of bootstrap resampling as outlined above may considerably reduce size distortions of the ζ_1 and ζ_2 tests since they are now based on critical values obtained from estimates of the test statistics' exact null distributions.

The standard formulation of the information matrix tests presented above involves testing K restrictions. (Recall that $K = k(k+1)/2$ restrictions are tested because the matrices $A(\boldsymbol{\theta}_0)$ and $B(\boldsymbol{\theta}_0)$ are symmetric.) It is possible, nonetheless, to focus on a given subset of parameters by computing the test statistics for $K_s \leq K$ restrictions. When $K_s = K$, all restrictions are considered; when $K_s < K$, focus is placed on a subset of restrictions. For instance, the practitioner may choose to only consider restrictions related to the mean or to the precision submodel, the corresponding number of restrictions being $P = p(p+1)/2$ and $Q = q(q+1)/2$, respectively. Here, we stress the usefulness of the matrix expression we developed for $\nabla \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y})$. For example, if one wishes to only test restrictions related to the mean submodel, it suffices to consider $C_i^{\boldsymbol{\beta},\boldsymbol{\beta}}$ and its derivatives with respect to $\boldsymbol{\beta}$. In that case, $\boldsymbol{d}_i(\boldsymbol{\theta}) = \text{vech}(C_i^{\boldsymbol{\beta},\boldsymbol{\beta}})$, and it follows that $\nabla \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y}) = (X * X)^\top \Lambda_{\boldsymbol{\beta},\boldsymbol{\beta}}^{\boldsymbol{\beta}} X$. Similarly, when the restrictions only relate to the precision submodel, it suffices to use $C_i^{\boldsymbol{\delta},\boldsymbol{\delta}}$ and its derivatives with respect to $\boldsymbol{\delta}$. Here, $\boldsymbol{d}_i(\boldsymbol{\theta}) = \text{vech}(C_i^{\boldsymbol{\delta},\boldsymbol{\delta}})$ and $\nabla \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y}) = (Z * Z)^\top \Lambda_{\boldsymbol{\delta},\boldsymbol{\delta}}^{\boldsymbol{\delta}} Z$.

Under fixed precision, $\phi_1 = \cdots = \phi_n = \phi$, g_2 is the identity link function, and $\eta_{2i} = \delta_1 \; \forall i$. The model is thus a particular case of the varying precision model. When computing the quantities used in ζ_1 and ζ_2, one must set $d\phi_i / d\eta_{2i} = 1$. As a consequence, $d(d\phi_i / d\eta_{2i}) / d\phi_i = 0$, etc.

2.3.2 On the validity of beta regression misspecification tests

We will now discuss the validity of the tests of correct beta regression model specification presented in the previous subsection. In particular, we will show that $\sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{D} \mathcal{N}_K(\mathbf{0}, V(\boldsymbol{\theta}_0))$ and we will prove the consistency of the former estimators of the asymptotic covariance matrix of $\sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n)$. The proofs of all results are presented in Appendix C. The following assumptions are made:

(A1) $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_P(n^{-1/2})$.

(A2) All regressors are uniformly bounded, that is, there exists $M > 0$ such that $|x_{ir}| < M$ and $|z_{iR}| < M$ for all $i = 1, \ldots, n$, $r = 1, \ldots, p$ and $R = 1, \ldots, q$.

(A3) The regression functions are uniformly bounded above and below. This means that there exists an interval $[\mu_L, \mu_U] \subset (0,1)$ such that $\mu_i \in [\mu_L, \mu_U]$ for all $i = 1, \ldots, n$. Also, there exists an interval $[\phi_L, \phi_U] \subset (0, \infty)$ such that $\phi_i \in [\phi_L, \phi_U]$ for all $i = 1, \ldots, n$.

(A4) The sequence $(\dot{A}_i(\boldsymbol{\theta}_0))$ is Cesàro convergent to a nonsingular matrix $A(\boldsymbol{\theta}_0)$, in the sense that there exists a nonsingular matrix $A(\boldsymbol{\theta}_0)$ such that the sequence of matrices $n^{-1}\sum_{i=1}^{n} \dot{A}_i(\boldsymbol{\theta}_0)$ converges to $A(\boldsymbol{\theta}_0)$ as $n \to \infty$.

(A5) The sequence $(\mathbb{E}(\partial \boldsymbol{d}_i(\boldsymbol{\theta})/\partial \boldsymbol{\theta}))$ is Cesàro convergent to a matrix $\nabla \boldsymbol{D}(\boldsymbol{\theta}_0)$ for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

(A6) The sequence of covariance matrices of $\boldsymbol{d}_i(\boldsymbol{\theta}_0) - \nabla \boldsymbol{D}(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0)^{-1} \nabla \ell(\boldsymbol{\theta}_0, Y_i)$ is Cesàro convergent to a positive definite matrix $V(\boldsymbol{\theta}_0)$ under correct model specification.

(A7) The sequence of covariance matrices of $\boldsymbol{d}_i(\boldsymbol{\theta}_0)$ is Cesàro convergent to a matrix $\Phi(\boldsymbol{\theta}_0)$ under correct model specification.

We note that Assumption (A1) holds whenever $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is asymptotically normally distributed.

At the outset, we establish two results: Lemmas 1 and 2.

Lemma 1. *$\ell(\mu_i, \phi_i; y_i)$ and all its partial derivatives of any order with respect to components of $\boldsymbol{\theta}$ have finite moments of all orders and all of them are uniformly bounded in i.*

Lemma 2. *Let $(\boldsymbol{W}_{n,i})$ be a double sequence of independent random vectors of same dimension k with $\mathbb{E}(\boldsymbol{W}_{n,i}) = \boldsymbol{\mu}_{n,i}$ and $\mathrm{Var}(\boldsymbol{W}_{n,i}) = \Sigma_{n,i}$, where $\Sigma_{n,i}$ is a positive definite matrix $\forall n$ and $\forall i$. Assume that there exist $\delta > 0$ and $\Delta > 0$ such that $\mathbb{E}(\|\boldsymbol{W}_{n,i}\|^{2+\delta}) < \Delta$ $\forall i$. Let $\overline{\boldsymbol{W}}_n = n^{-1}\sum_{i=1}^{n} \boldsymbol{W}_{n,i}$, $\bar{\boldsymbol{\mu}}_n = n^{-1}\sum_{i=1}^{n} \boldsymbol{\mu}_{n,i}$ and $\bar{\Sigma}_n = n^{-1}\sum_{i=1}^{n} \Sigma_{n,i}$. If $\bar{\Sigma}_n$ converges to a positive definite matrix V, as $n \to \infty$, then,*

$$\sqrt{n}(\overline{\boldsymbol{W}}_n - \bar{\boldsymbol{\mu}}_n) \xrightarrow{D} \mathcal{N}_k(\mathbf{0}, V).$$

We can now state our first main result: a central limit theorem for the vector $\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n)$. The proof of this result uses arguments that are similar to those in White (1982).

Theorem 1. *Under Assumptions (A1)–(A6) and correct model specification,*

$$\sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{D} \mathcal{N}_K(\mathbf{0}, V(\boldsymbol{\theta}_0)).$$

We will now move to our second main result which relates to the consistency of $V_{n1}(\hat{\boldsymbol{\theta}}_n)$.

Theorem 2. *Under Assumptions (A1)–(A7) and correct model specification,*

$$V_{n1}(\hat{\boldsymbol{\theta}}_n) \xrightarrow{P} V(\boldsymbol{\theta}_0).$$

The expression obtained for $V(\boldsymbol{\theta}_0)$ may seem unusual, but we observe that $A(\boldsymbol{\theta}_0)$ is a negative definite matrix.

The consistency of $V_{n2}(\hat{\boldsymbol{\theta}}_n)$ can now be easily established.

Theorem 3. *Under Assumptions (A1)–(A7) and correct model specification,*

$$V_{n2}(\hat{\boldsymbol{\theta}}_n) \xrightarrow{P} V(\boldsymbol{\theta}_0).$$

2.4 NUMERICAL EVIDENCE

We will now report Monte Carlo simulation results on the tests' finite sample performance. Data generation is performed under the null and alternative hypotheses in order to evaluate the tests' null and non-null behavior, respectively. Parameter estimates were obtained by numerically maximizing the beta regression log-likelihood function using the BFGS quasi-Newton algorithm with analytical first derivatives. Starting values used in the iterative optimization scheme are computed as described on pages 349 and 350 of Ferrari, Espinheira and Cribari-Neto (2011). The null hypothesis is that the beta regression model is correctly specified and the alternative hypothesis is that it is misspecified. Inference is based on the following tests: ζ_1, ζ_{1B}, ζ_2, ζ_{2B} and RESET (ζ_R). The ζ_{1B} and ζ_{2B} tests employ bootstrap critical values. We use the implementation of the RESET denoted by (PEREIRA; CRIBARI-NETO, 2014) as R_μ, since it is the overall best performer in their numerical evaluations. RESET test inferences are carried out using the likelihood ratio test. The number of Monte Carlo and bootstrap replications are 5000 and 500, respectively.

The sample sizes and significance levels are, respectively, $n \in \{100, 250, 500, 1000\}$ and $\alpha \in \{0.01, 0.05, 0.1\}$ for the size simulations and $\alpha \in \{0.05, 0.1\}$ for the power simulations. The simulations were performed using the R statistical computing environment; see R Core Team (2023).

At the outset, we consider the following varying precision beta regression model: $\log(\mu_i/(1-\mu_i)) = \beta_1 + \beta_2 x_{i2}$ and $\log(\phi_i) = \delta_1 + \delta_2 z_{i2}$, $i = 1, \ldots, n$. The true parameter values are $\beta_1 = 1.5$, $\beta_2 = 1.2$, $\delta_1 = 1.5$ and $\delta_2 = 2$. The values of x_{i2} and z_{i2} were obtained as random draws from $\mathcal{U}(-0.5, 0.5)$ and $\mathcal{U}(1, 1.5)$ distributions, respectively. In our first set of simulations, the fitted and true models coincide. Table 9 contains the null rejection rates of the ζ_{1B}, ζ_{2B} and ζ_R tests. All entries are percentages. We do not report results on the ζ_1 and ζ_2 tests since they are very oversized; their null rejection rates at the 5% significance level and $n = 500$ exceed 37% and 44%, respectively. Recall that these tests employ asymptotic critical values, that they do not employ data resampling, and that their test statistics contain estimators of the *asymptotic* covariance matrix of $\sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n; \boldsymbol{Y})$. By contrast, the ζ_{1B} and ζ_{2B} tests, which employ parametric bootstrap critical values, display very good control of the type I error frequency. As in Horowitz (1994), the use of critical values obtained through bootstrap resampling yields empirical sizes that are very close to the corresponding significance levels, whereas the empirical and nominal sizes may differ greatly when asymptotic critical values are employed. The ζ_R test displays good control of the type I error frequency.

We will now move to simulations in which the true data generating process differs from the fitted model, i.e., the latter is incorrectly specified. We thus report the estimated powers of the tests. We will consider eight scenarios in which the specification of the fitted model is in error. We will refer to them as S1, S2, S3, S4, S5, S6, S7 and S8. The non-null rejection rates are presented in Table 10. All entries are percentages.

In the first scenario with model misspecification (S1), a mean regressor present in the true model is not included in the fitted model. The data generating mechamism is $\log(\mu_i/(1-\mu_i)) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3}$ and $\log(\phi_i) = \delta_1 + \delta_2 z_{i2}$, with $\beta_1 = 1.5$, $\beta_2 = -1.2$, $\beta_3 = -0.75$, $\delta_1 = 1.2$, $\delta_2 = 2$. The values of x_{i2}, x_{i3} and z_{i2} are obtained, respectively, as random draws from the $\mathcal{U}(-1.5, 1.5)$, $\mathcal{LN}(0, 0.5)$ and $\mathcal{U}(1, 1.5)$ distributions. The fitted model does not contain x_3 as a mean regressor; it is thus incorrectly specified. The simulation results show that the powers of the two information matrix tests increase

Table 9 – Null rejection rates (%).

n	ζ_{1B}	ζ_{2B}	ζ_R
	$\alpha = 10\%$		
100	10.0	10.4	11.2
250	9.8	10.0	9.6
500	10.1	9.9	10.9
1000	10.4	10.3	10.2
	$\alpha = 5\%$		
100	5.1	5.5	5.7
250	4.9	5.1	4.8
500	5.4	5.3	5.4
1000	5.3	5.3	5.1
	$\alpha = 1\%$		
100	1.3	1.2	1.2
250	1.2	1.3	0.9
500	1.2	1.3	1.0
1000	1.1	1.2	1.0

Source: Author

with the sample size. That is, their type II error frequencies diminish as the sample size increases. The results also show that these tests can reliably detect that the model specification is in error, especially when the sample size is not small. When $n = 250$, the powers of the two information matrix tests at $\alpha = 10\%$ exceed 95%. By contrast, the RESET test (ζ_R) is not consistent in this setting. Notably, the RESET test displays unit asymptotic power under missing mean covariate in Pereira and Cribari-Neto (2014). The main difference between their simulations and ours is that they generate the values of all regressors using a small sample size and replicate these values for larger sample sizes, which implies that the values of the means (μ_i) and precisions (ϕ_i) are also replicated. This is, however, unrealistic, since in practical applications there is no replication of the values of regressors across different sample sizes. If we proceed in that manner, the RESET test becomes consistent. We generated 125 values for each covariate and replicated them once, three times and seven times for $n \in \{125, 250, 500, 1000\}$. The RESET test's powers at $\alpha = 10\%$ ($\alpha = 5\%$) become 47.4%, 81.6%, 99.1% and 100.0% (30.0%, 67.3%, 96.9% and 100.0%) for these sample sizes. Finally, we note that the RESET test may also be inconsistent in the classical regression model when an important regressor is not in the fitted model; see Leung and Yu (2000).

The second scenario (S2) is of neglected nonlinearity in the mean submodel. The true data generating process is $\log(\mu_i/(1-\mu_i)) = \beta_1 + \exp(\beta_2 x_{i2})$ and $\log(\phi_i) = \delta_1 + \delta_2 z_{i2}$, with $\beta_1 = -1$, $\beta_2 = -2.3$, $\delta_1 = 1.5$, $\delta_2 = 2$. The values of x_{i2} and z_{i2} were selected as in the size simulations. The estimated model uses $\beta_2 x_{i2}$ instead of $\exp(\beta_2 x_{i2})$, and is thus incorrectly specified. The simulation results show that all estimated powers increase with n. Here, ζ_{2B} outperforms ζ_{1B} for $n > 100$. For instance, the ζ_{1B} and ζ_{2B} non-null rejection rates for $n = 250$ and $\alpha = 10\%$ are 65.0% and 82.0%, respectively. It is noteworthy that the RESET test is considerably more powerful than the information matrix tests in this setting. This is not surprising, since the neglected nonlinearity in the mean submodel greatly impacts powers of such a predictor.

In Scenario S3, we generate samples according to the Kumaraswamy regression model (MITNIK; BAEK, 2013) and fit the beta model. A noteworthy difference is that in Kumaraswamy model μ_i is the median (not the mean) of the ith response. The regression structure in both models is $\log(\mu_i/(1-\mu_i)) = \beta_1 + \beta_2 x_{i2}$ and $\log(\phi_i) = \delta_1 + \delta_2 z_{i2}$, with $\beta_1 = 1$, $\beta_2 = 3$, $\delta_1 = 3.5$, $\delta_2 = 2.5$. The values of x_{i2} were obtained as random $\mathcal{U}(0,1)$ draws and $z_{i2} = x_{i2}^2$. We note that the tests become progressively more powerful as the sample size increases. Furthermore, ζ_{1B} is slightly more powerful than ζ_{2B}. For instance, for $\alpha = 10\%$ and $n = 500$, the estimated powers of ζ_{1B} and ζ_{2B} are 93.4% and 92.0%, respectively. The RESET test is not consistent. It displays non-null rejection rates that are close to the corresponding significance levels for all sample sizes. For example, the test's power for $n = 500$ and $\alpha = 10\%$ is 8.7%. In order to verify whether the RESET test lacks consistency under other laws, we performed simulations in which the beta regression model is fitted to data generated using the unit Weibull model of Mazucheli, Menezes and Ghitany (2018); again, the only difference between the true and fitted models lies in the response distribution. The results are not presented for brevity. Again, the RESET test was not consistent. It thus seems that, unlike the information matrix tests, the RESET test is not capable of identifying model misspecification that is solely related to the response distribution.

In the fourth misspecification scenario (S4), the mean link function is incorrectly specified. The true data generating process is $\log(-\log(1-\mu_i)) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3}$ and $\log(\phi_i) = \delta_1 + \delta_2 z_{i2}$, with $\beta_1 = -1$, $\beta_2 = -1$, $\beta_3 = 1.5$, $\delta_1 = 3.7$ and $\delta_2 = 1.7$. The fitted model, however, uses the logit mean link function. The values of x_{i2} and z_{i2} are obtained as random $\mathcal{U}(-0.5, 0.5)$ draws and those of x_{i3} come from standard uniform draws. Here, ζ_{1B} outperforms ζ_{2B}. For example, the estimated powers of ζ_{1B} and ζ_{2B} for $n = 250$ and $\alpha = 5\%$ are, respectively, 58.6% and 39.7%. The overall best performer in this setting is the RESET test. It is considerably more powerful than the information matrix tests, especially for $n \leq 250$.

Similarly to the previous scenario, here (Scenario S5) the mean link function is misspecified. The link function in the true model is probit whereas the fitted model used the logit link. Unlike the previous scenario, here both link functions are symmetric. The true model is $\Phi^{-1}(\mu_i) = \beta_1 + \beta_2 x_{i2}$ and $\log(\phi_i) = \delta_1 + \delta_2 z_{i2}$, with $\beta_1 = 1.5$, $\beta_2 = 1.2$, $\delta_1 = 1.2$ and $\delta_2 = 2.0$, Φ denoting the standard normal distribution function. The only difference between the fitted and true models is that the former uses the logit link. The values of x_{i2} and z_{i2} are obtained as random $\mathcal{U}(-1.5, 1.5)$ and $\mathcal{U}(1.0, 1.5)$ draws, respectively. Unlike in the previous scenario, here the information matrix tests are more powerful than the RESET test. For example, the non-null rejection rates of ζ_{1B}, ζ_{2B} and ζ_R when $n = 250$ and $\alpha = 10\%$ are, respectively, 53.5%, 52.9% and 35.8%.

The precision submodel specification is in error in Scenario S6. The data generating process is $\log(\mu_i/(1-\mu_i)) = \beta_1 + \beta_2 x_{i2}$ and $\log(\phi_i)) = \delta_1 + \delta_2 x_{i2}$. The precision submodel is, however, incorrectly specified. The model fitted to the data uses $\log(\phi_i) = \delta_1 + \delta_2 z_{i2}$. The values of x_{i2} and z_{i2} are obtained as realizations from the standard normal and standard uniform distributions, respectively, $\beta_1 = 0.25$, $\beta_2 = -0.5$, $\delta_1 = 2$ and $\delta_2 = 0.5$. The results show that the ζ_{1B} is the best performer. For example, the estimated powers of ζ_{1B} and ζ_{2B} when $n = 250$ and $\alpha = 10\%$ are, respectively, 87.8% and 83.6%. Additionally, the information matrix tests are substantially more powerful than the RESET test. For instance, when $n = 250$ and $\alpha = 5\%$ the estimated powers of the ζ_{1B}, ζ_{2B} and ζ_R tests are 75.8%, 67.6% and 36.3%, respectively.

The seventh situation in which there is model misspecification (Scenario S7) involves fitting a beta regression model to data subject to inflation. The data are generated using an inflated beta regression with inflation at zero. Data inflation occurs with probability 0.05. The model is $\log(\mu_i/(1-\mu_i)) = \beta_1 + \beta_2 x_{i2}$ and $\log(\phi_i) = \delta_1 + \delta_2 z_{i2}$. The values of the two covariates were obtained as random $\mathcal{U}(0, 1)$ draws and the parameter values are $\beta_1 = -1$, $\beta_2 = 2$, $\delta_1 = 1$, $\delta_2 = 1.5$. The inflated values are replaced by $0.5/n$ prior to fitting the beta model. The only source of model misspecification is that the discrete nature of the data inflation mechanism is neglected. The results show that the information matrix tests are able to reliably detect the model misspecification, especially for $n \geq 250$. They are considerably more powerful than the RESET test. As in the first scenario, the RESET test is not consistent, i.e., it does not display unit asymptotic power. To exemplify, consider $n = 250$ and $\alpha = 5\%$. The powers of the ζ_{1B}, ζ_{2B} and ζ_R tests are 97.7%, 97.3% and 8.0%, respectively.

In Scenario S8, a beta regression model is fitted, but the true underlying data generating mechanism is a mixture of beta laws. The true density function of y_i is $\lambda_m f_{1i} + (1-\lambda_m) f_{2i}$, $\lambda_m \in (0,1)$, where f_{1i} and f_{2i} are beta densities with means μ_{1i} and μ_{2i}, respectively, and common precision ϕ_i. Here, $\log(\mu_{ji}/(1-\mu_{ji})) = \beta_{1,j} + \beta_{2,j} x_i$ and $\log(\phi_i) = \delta_1 + \delta_2 z_i$, $j = 1,2$. Also, $\lambda_m = 0.9$, $\beta_{1,1} = 1$, $\beta_{2,1} = 1$, $\beta_{1,2} = 0.25$, $\beta_{2,2} = -0.5$, $\delta_1 = 2.0$ and $\delta_2 = 3.0$. The fitted mean submodel is $\log(\mu_i/(1-\mu_i)) = \beta_1 + \beta_2 x_i$ and the precision submodel is as in the true model. The values of x_i and z_i were obtained as random draws from the standard uniform distribution. The reported results show that the ζ_{1B} and ζ_{2B} tests display high powers as early as when $n = 100$. For example, with $n = 100$ and $\alpha = 10\%$ the estimated powers of ζ_{1B} and ζ_{2B} are 77.2% and 75.8%, respectively. Once again, the RESET test is not consistent.

An attractive feature of information matrix tests is that they can be performed using a subset of the restrictions implied by the information matrix equality. It is typically possible to achieve higher power by doing so. To illustrate that, consider Scenario S2 (neglected nonlinearity), $\alpha = 10\%$ and $n \in \{100, 250, 500\}$. The powers of ζ_{1b} (ζ_{2b}) when we only consider restrictions related to the mean submodel are 25.5%, 82.2% and 97.9% (27.2%, 83.2% and 98.1%); in Scenario S4 (true link function: cloglog), we obtain 35.5%, 73.5% and 95.1% (32.3%, 74.4% and 95.6%); finally, in Scenario S5 (true link function: probit) the non-null rejection rates become 45.7%, 72.1% and 90.1% (44.7%, 73.7% and 90.8%). In all cases, the tests' powers are increased.

The simulation results presented above show that the information matrix tests are capable to detecting several sources of model misspecification in beta regressions, especially when the sample size is not small. They also show that the RESET test lacks consistency under some forms of incorrect model specification. It is, however, very powerful for detecting neglected nonlinearity. Since the information matrix tests and the RESET test have distinct strengths when it comes to detecting model misspecification, they should

Table 10 – Non-null rejection rates (%): Scenarios S1 through S8.

n	ζ_{1B}	ζ_{2B}	ζ_R	ζ_{1B}	ζ_{2B}	ζ_R
	S1			S2		
	$\alpha = 10\%$					
100	19.0	15.8	12.6	20.1	19.8	100.0
250	96.0	95.2	0.4	65.0	82.0	100.0
500	94.1	95.5	39.9	90.4	93.4	100.0
1000	100.0	100.0	6.7	100.0	100.0	100.0
	$\alpha = 5\%$					
100	10.4	7.9	5.5	12.3	11.8	100.0
250	91.7	90.0	0.1	49.3	70.1	100.0
500	88.1	89.9	21.5	78.7	84.8	100.0
1000	99.9	99.9	2.7	100.0	100.0	100.0
	S3			S4		
	$\alpha = 10\%$					
100	25.2	21.2	8.2	19.8	17.9	84.9
250	64.2	60.1	7.7	73.4	55.5	98.5
500	93.4	92.0	8.7	96.5	90.5	100.0
1000	100.0	100.0	9.0	100.0	100.0	100.0
	$\alpha = 5\%$					
100	15.1	11.4	4.0	10.4	9.9	76.8
250	50.1	45.1	3.5	58.6	39.7	96.5
500	87.0	84.4	4.1	93.0	81.8	99.9
1000	99.9	99.8	4.4	100.0	100.0	100.0
	S5			S6		
	$\alpha = 10\%$					
100	21.6	20.0	14.7	36.8	26.4	21.5
250	53.5	52.9	35.8	87.8	83.6	46.1
500	86.7	86.7	62.0	100.0	99.9	53.3
1000	99.64	99.86	81.50	100.0	100.0	99.4
	$\alpha = 5\%$					
100	12.3	11.2	7.8	22.1	14.2	14.0
250	37.1	35.6	23.3	75.8	67.6	36.4
500	73.7	72.4	46.9	99.9	99.6	53.3
1000	98.62	99.26	70.48	100.0	100.0	99.0
	S7			S8		
	$\alpha = 10\%$					
100	56.6	55.1	11.9	77.2	75.8	8.4
250	99.0	98.6	14.3	99.8	99.8	7.5
500	100.0	100.0	15.4	100.0	100.0	7.4
1000	100.0	100.0	17.0	100.0	100.0	7.6
	$\alpha = 5\%$					
100	43.0	37.9	6.3	65.0	62.6	4.0
250	97.7	97.3	8.0	99.4	99.4	3.0
500	100.0	100.0	8.8	100.0	100.0	3.5
1000	100.0	100.0	10.8	100.0	100.0	3.7

Source: Author

be used in complementary fashion in practical applications.

2.5 EMPIRICAL APPLICATIONS

We will now present and discuss two empirical applications of the proposed misspecification tests. They will showcase the usefulness of such tests.

2.5.1 Proportion of body fat

In the first empirical application, the interest lies in modeling the proportion of body fat in the arms (variable: ARMS) and abdomen (variable: ANDROID). Measurements were made on 298 people in a public hospital located in the capital of the state of Paraná, Brazil, in 2018. The values of the following explanatory variables were also recorded: age (in years) and body mass index – BMI (in kg/m^2). Each individual was classified according to gender (female, male) and the level of physical activity (sedentary, insufficiently active, active); in what follows, we will use a dummy variable for the former and two dummy variables for the latter. BMI is a parameter adopted by the World Health Organization to classify weight-related health patterns, such as malnutrition and obesity. In general, BMI correlates positively with the proportion of body fat in obese individuals. The stratification according to physical activity levels was based on the International Physical Activity Questionnaire (IPAQ), which allows estimation of the weekly time spent in physical activities of different intensities; for details on the questionnaire, see Benedetti *et al.* (2007) and Matsudo *et al.* (2001). The source of the data is Petterle *et al.* (2021). Mazucheli *et al.* (2021) and Mazucheli *et al.* (2022) modeled the ARMS variable using unit Birnbaum-Saunders and Vasicek regression models, respectively. We will use, in our beta regression analysis, the same covariates as them.

We note that Deurenberg, Weststrate and Seidell (1991) developed predictive formulas for fat proportion measurements. The authors observed a positive correlation between age and fat proportion in a sample of 1229 individuals, 521 males and 708 females of different age groups. This correlation was higher in women.

We will perform beta regression analyses for ARMS and ANDROID. The mean and median (minimum and maximum) values of ARMS are, respectively, 0.2661 and 0.2610 (0.0420 and 0.5470), the standard deviation being 0.1113. The corresponding figures for

ANDROID are 0.3787 and 0.3970 (0.0720 and 0.5800), and 0.1102. There are 150 female and 148 male individuals. The minimum, maximum and mean ages are 18, 87 and 46, respectively, and the standard deviation of the ages is 19.8792. The mean BMI is 24.7200 and the standard deviation is 3.1507. The number of sedentary, insufficiently active and active individuals are, respectively, 60, 76 and 162.

At the outset, we consider the dependent variable ARMS. The following fixed precision beta regression model was fitted to the data, after preliminary investigation that showed no evidence of variable precision:

$$g_1(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6},$$

where x_{i2} is age, x_{i3} is BMI, x_{i4} equals 1 for men and 0 for women, x_{i5} equals 1 for insufficiently active individuals and 0 otherwise, and x_{i6} equals 1 for active individuals and 0 otherwise. Five link functions were used, namely logit, probit, loglog, cloglog, and cauchit.

As noted by White (1982), it is appropriate to ignore restrictions that are identically equal to zero or linear combinations of other restrictions. There are seven parameters in the above beta regression model: six regression coefficients and ϕ. Hence, $k = 7$. The maximum number of restrictions that can be tested is thus $k(k+1)/2 = 28$. However, four restrictions related to the intercepts and the dummy variables must be disregarded, since three are not unique ($C_i^{(r=4,s=1)} = C_i^{(r=4,s=4)}$, $C_i^{(r=5,s=1)} = C_i^{(r=5,s=5)}$, and $C_i^{(r=6,s=1)} = C_i^{(r=6,s=6)}$) and one is identically null ($C_i^{(r=6,s=5)} = 0$). We thus test 24 restrictions.

Table 11 contains the information matrix tests' p-values (expressed as percentages) along with the AIC and BIC values and the value of the pseudo-R^2 of Nagelkerke (1991) (R^2). The misspecification tests were performed using 500 bootstrap replications. The best results are in boldface. The model with the largest p-values is the loglog model. Interestingly, it also has the highest pseudo-R^2 values and the lowest AIC and BIC values. It is also noteworthy that the only model rejected at the usual significance levels is the cauchit model.

We performed the RESET test for the models with the five link functions. All models are rejected at the 10% significance level; the loglog model is the only model not rejected at 5%. This inference contrasts with that obtained from the information matrix tests, where only the cauchit model is rejected. Given the discrepancy between the

Table 11 – Information matrix tests' p-values (%), information criteria values, pseudo-R^2 values; dependent variable: ARMS.

link function	ζ_{1B}	ζ_{2B}	AIC	BIC	R^2
logit	13.40	19.00	−907.49	−881.61	0.78
probit	16.20	25.60	−914.13	−888.25	0.78
loglog	**27.20**	**41.60**	**−923.41**	**−897.53**	**0.79**
cloglog	12.60	16.00	−897.86	−871.98	0.77
cauchit	0.60	1.20	−854.59	−828.71	0.74

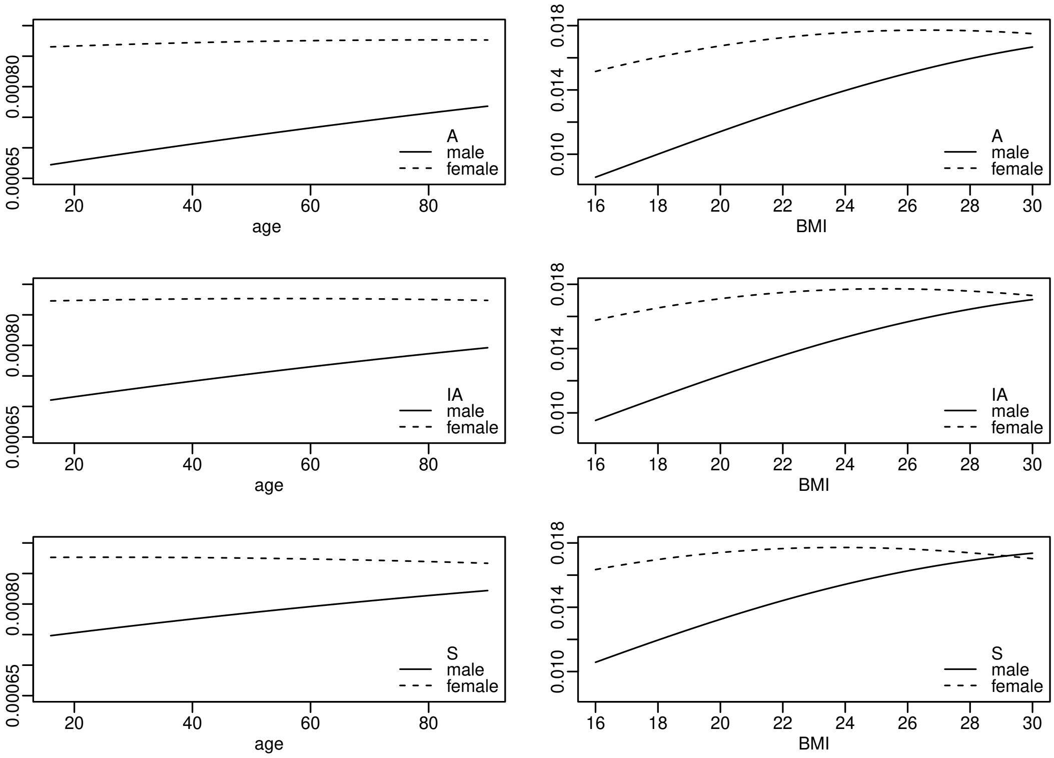
Source: Author

two sets of inferences, we turn to the analysis of the residuals of the estimated models. First, we constructed residual quantile-quantile plots with simulated envelopes. The number of simulations is 100 and the envelopes correspond to the 0.025 and 0.975 residual quantiles. We use the Pearson residual (r_i) since it is based on a comparison between y_i and $\hat{\mu}_i = g_1^{-1}(\hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_6 x_{i6})$, and our interest lies in using the latter as a representation of the former. The total number of residuals outside the envelopes for the logit, probit, cloglog, loglog and cauchit models are, respectively, 4, 0, 8, 0 and 21. (These plots and all other residual plots are not presented for brevity.) Second, for each fitted model, we plotted r_i against $\hat{\eta}_{1i}$. As noted by Ferrari and Cribari-Neto (2004), a detectable trend in this plot is suggestive of mean link function misspecification. The only plot with a detectable trend is that of the cauchit model: there is a noticeable rise followed by a decline. That is, there is a visible quadratic trend in the cauhit residual plot. To be sure, we estimated, for each link function, linear regressions of r_i on (i) $\hat{\eta}_{1i}$ and (ii) $\hat{\eta}_{1i}^2$. The only model for which such trends are statistically significant at the usual significance levels according to z-tests is the cauchit model. Third, we produced worm plots for the fitted models using quantiles residuals. For details on worm plots and quantile residuals, see Buuren and Fredriks (2001) and Dunn and Smyth (1996), respectively. Such plots convey no clear evidence against the correct specification of any model since, for the five models, nearly all points fall within the two semicircles of reference. Additionally, for all models, the coefficients of the cubic fits are not indicative of misfit. In all cases, the absolute values of such coefficients are smaller than 0.10 (intercept), 0.10 (linear term), 0.05 (quadratic term) and 0.03 (cubic term); see Buuren and Fredriks (2001) for details. The coefficient of the quadratic term for the cauchit model is −0.0441, only slightly smaller than 0.05 in absolute value, i.e., it falls below the misfit threshold by a very slim margin. In summary, the conclusion drawn from the residual analyses is more in line with the inference reached via information matrix tests than with that from the RESET test.

Using the loglog link function, we obtain the following estimates for $\beta_1, \ldots, \beta_6$ (standard errors in parentheses): -1.2595 (0.0748), 0.0024 (0.0006), 0.0482 (0.0034), -0.4926 (0.0185), -0.0706 (0.0272), and -0.1353 (0.0260). Also, $\hat{\phi} = 68.4009$. All regression coefficients are different from zero at the 1% significance level according to individual z tests. Notice that $\hat{\beta}_2$ and $\hat{\beta}_3$ are positive, thus implying that, all else equal, mean arms fat proportion increases with age and BMI. Additionally, the estimated coefficients of all three dummy variables are negative. It thus follows that mean arms fat proportion tends to be lower for men; it is also lower for non-sedentary individuals. Regarding the level of physical activity, we note that $\hat{\beta}_6$ (related to active individuals) is nearly twice as small as $\hat{\beta}_5$ (insufficiently active individuals); that is, the negative impact of physical activity on mean arms fat proportion is quite strong.

We computed the impacts of x_2 (age) and x_3 (BMI) on the mean proportion of arms fat for male and female individuals. When estimating the impact of x_2, we set the value of x_3 at its median value, and vice-versa. Separate impacts were obtained for sedentary, insufficiently active, and active individuals. More specifically, we computed $\partial\mu_i/\partial x_{i2} = \partial g_1^{-1}(\eta_{1i})/\partial x_{i2}$ (the impact of age on the mean response) and $\partial\mu_i/\partial x_{i3} = \partial g_1^{-1}(\eta_{1i})/\partial x_{i3}$ (the impact of BMI on the mean response), where g_1 is the loglog link, and then replaced all regression coefficients in these derivatives by their maximum likelihood estimates. The estimated impacts of the two covariates (age and BMI) on the mean response are presented in the six panels of Figure 13. Several important conclusions can be drawn from the estimated impacts presented in this figure. First, all estimated impacts are positive. Second, in all cases, the impacts are more intense for women than for men (except for the impact of BMI when such a covariate assumes very large values). Third, the impact of age becomes progressively stronger as men become older; by contrast, it is nearly constant for women. The convergence of the two curves (male and female) as age increases is slow. Fourth, for all ages, increased physical activity is more beneficial for men than for women; the impact curve for male individuals shift down as the level of physical activities increases (from S to IA, and then to A; bottom to top). Fifth, the impact of BMI is strictly increasing for men, but not for women; for the latter, it slowly increases, peaks, and then slowly decreases. Sixth, the BMI impacts are nearly the same for men and women when BMI is large (around 29 and higher). Seventh, when BMI is small, its impact is much stronger for women than for men. Eighth, the impact of BMI increases fairly quickly for men as the value of BMI increases.

Figure 13 – Estimated impacts of age (left panels) and BMI (right panels) on the mean proportion of arms fat; the bottom, middle and top panels are for sedentary (S), insufficiently active (IA), and active (A) individuals.



Source: Author

We fitted the same model for the dependent variable ANDROID. The p-values of ζ_{1B} (ζ_{2B}) for the logit, probit, loglog, cloglog, and cauchit models are 0.20% (2.00%), 0.60% (1.80%), 1.20% (3.00%), $< 0.01\%$ (0.40%), and $< 0.01\%$ (0.20%), respectively. All five models are rejected by the two tests at the 5% significance level, the rejection occurring at 1% for the cloglog and cauchit models. Furthermore, ζ_{1B} rejects the probit and logit models at 1%. The maximal pseudo-R^2 is 0.67. The five models are rejected by the RESET test at 1%.

In conclusion, the information matrix tests indicate that there is not evidence against the correct specification of four out of the five fitted models when ARMS is the dependent variable; the cauchit model is rejected by the misspecification tests. A better distinction between the logit, probit, loglog and cloglog models through the information matrix tests would require a larger sample size. By contrast, according to the tests, there is strong evidence against all fitted models using ANDROID as the response variable.

2.5.2 Proportion of religious disbelievers

We will now consider a second empirical application: the beta regression modeling of the proportion of religious disbelievers in different nations performed by Cribari-Neto and Souza (2013). The authors used data on 124 countries and the following model:

$$\begin{aligned}\text{loglog}(\mu_i) &= \beta_1 + \beta_2\text{IQ}_i + \beta_3\text{IQ}_i^2 + \beta_4\text{MUSL}_i + \beta_5\text{INCOME}_i + \beta_6\text{OPENESS}_i\\ \log(\phi_i) &= \delta_1 + \delta_2\text{IQ}_i.\end{aligned}$$

The response variable is the proportion of atheists in the general population and the covariates are: (i) IQ: average intelligence quotient of the population, (ii) MUSL: dummy variable that equals 1 if the majority of the population is Muslim and 0 otherwise, (iii) INCOME: per capita Gross National Income adjusted for purchasing power parity and (iv) OPENESS: logarithm of the ratio between the volume of foreign trade (sum of total imports and exports) and the Gross Domestic Product. Their analysis focuses on the impact of average intelligence on the mean proportion of the population who do not hold religious beliefs. After fitting the model, the authors plotted such an impact against IQ similarly to what we did in the previous empirical application; see Figure 3 in their paper. The use of intelligence quotient squared as a mean regressor caused the impact curve to be bell-shaped: the impact of average intelligence becomes progressively stronger, peaks at around $\text{IQ} = 107$, and then gradually weakens. For all values of IQ, the impact is positive which implies that the mean proportion of religious disbelievers increase with average intelligence. We will now investigate whether their model is correctly specified using information matrix testing inference.

As noted above, the model used by Cribari-Neto and Souza (2013) uses the loglog mean link function. We will also consider three alternative links: logit, probit, and cloglog. We do not consider the cauchit model since it yields a considerably smaller pseudo-R^2 and two mean regressors, MUSL and INCOME, lose statistical significance at 5%. The information matrix tests are performed using 500 bootstrap replications. Since there are eight parameters ($k = 8$), it is possible to test up to 36 restrictions. There are four superfluous restrictions that relate to the intercepts, IQ and IQ^2; they are not unique ($C_i^{(r=1,s=4)} = C_i^{(r=4,s=4)}$, $C_i^{(r=1,R=2)} = C_i^{(r=2,R=1)}$, $C_i^{(r=1,s=3)} = C_i^{(r=2,s=2)}$ and $C_i^{(r=3,R=1)} = C_i^{(r=2,R=2)}$). We then test 32 restrictions. Table 12 contains the tests

p-values (expressed as percentages) along with AIC, BIC and pseudo-R^2 values. The best results are in boldface. They all favor the loglog model (largest p-values, smallest AIC and BIC values, largest pseudo-R^2). These results are taken as supporting evidence for the model used by Cribari-Neto and Souza (2013). We also note that the loglog model is the only model that is not rejected by the RESET test at the usual significance levels.

Table 12 – Information matrix tests' p-value (%), information criteria values, pseudo-R^2 values; dependent variable: proportion of atheists.

link function	ζ_{1B}	ζ_{2B}	AIC	BIC	R^2
logit	10.60	8.80	−509.66	−487.10	0.74
probit	13.80	6.00	−514.37	−491.81	0.75
loglog	**14.60**	**14.60**	**−518.98**	**−496.42**	**0.76**
cloglog	1.20	0.20	−502.56	−479.99	0.73

Source: Author

A key advantage of information matrix tests is that they allow one to focus on just a few restrictions, that is, one may only consider a few selected indicators. As noted earlier, the main focus of the empirical analysis in Cribari-Neto and Souza (2013) is on the impact of IQ on the mean response. We then performed the misspecification tests by only considering restrictions on the covariates related to average intelligence: (i) IQ and IQ^2 in the mean submodel and (ii) IQ in the precision submodel. By doing so we test six restrictions ($C_i^{(r=2,s=2)}$, $C_i^{(r=3,s=2)}$, $C_i^{(r=3,s=3)}$, $C_i^{(r=2,R=2)}$, $C_i^{(r=3,R=2)}$, $C_i^{(R=2,S=2)}$). We present in Table 13 the ζ_{1B} and ζ_{2B} tests' p-values (expressed as percentages). The only model that is not rejected by the information matrix tests at the 5% significance level is the loglog model. This is further evidence in favor of the model used by Cribari-Neto and Souza (2013). It also showcases the usefulness of the information matrix tests since they allow practitioners to test the correct specification of key aspects of their models. Notably, it is not possible to use the RESET test here since it cannot be used with focus on a subset of the covariates.

Table 13 – Information matrix tests' p-value (%) for six restrictions related to IQ and IQ^2.

link function	ζ_{1B}	ζ_{2B}
logit	2.00	1.20
probit	3.80	3.40
loglog	7.80	8.60
cloglog	0.60	< 0.01

Source: Author

A distinctive feature of the model used by Cribari-Neto and Souza (2013) is the use of average intelligence squared (IQ^2) as a mean regressor which leads to the bell-shaped impact curve shown in Figure 3 of their paper. We will use the information matrix tests to verify whether the model remains correctly specified when such a regressor is absent. The total number of parameters becomes $k = 7$, and it is then possible to test up to 28 restrictions. There are non-unique restrictions related to the intercept and the dummy variable ($C_i^{r=1,s=3}$,$C_i^{r=3,s=3}$) and to IQ and the intercept ($C_i^{r=1,S=2}$,$C_i^{r=2,S=1}$). We then test 26 restrictions. Both tests reject the correct specification of the model at 5%. The model is rejected by the RESET test at 1%. It is then clear that the correct specification of the model used by Cribari-Neto and Souza (2013) requires the inclusion of average intelligence squared in the mean submodel. This is further evidence in favor of their results.

Next, we will investigate whether the loglog model remains correctly specified when a couple of atypical cases are removed from the data, namely: Vietnam and the United States. Vietnam is a rather atypical country, as it has the largest response value and a value of IQ between the median and the 3rd quartile. Its proportion of atheists is 0.81, which is considerably larger than that of the nation with the second-highest proportion of religious disbelievers, Japan (0.65). As for the United States, its proportion of religious disbelievers is smaller than other high-income countries with high average intelligence. The country even contains largely religious regions known as the Bible Belt and the Mormon Corridor. The results obtained after these two cases were removed from the data are presented in Table 14 (again, p-values are expressed as percentages). We test 32 restrictions and the best results are in boldface. Now the only model that is not rejected at the 10% significance level is the loglog model. The same conclusion is reached using the RESET test. This is further evidence in favor of such a model.

Table 14 – Information matrix tests' p-value (%), information criteria values, pseudo-R^2 values; dependent variable: proportion of atheists; Vietnam and US removed from the data.

link function	ζ_{1B}	ζ_{2B}	AIC	BIC	R^2
logit	0.40	1.00	-532.14	-509.71	0.78
probit	7.40	7.80	-537.83	-515.40	0.79
loglog	**24.80**	**30.40**	$\mathbf{-543.07}$	$\mathbf{-520.63}$	**0.80**
cloglog	< 0.01	0.60	-524.55	-502.11	0.76

Source: Author

We also performed the information tests based on only six restrictions (those related to IQ and IQ^2) using the incomplete dataset. The loglog model is the only model that is not rejected by the information matrix tests at 10%. (The ζ_{1B} and ζ_{2B} p-values are 12.20% and 13.00%, respectively.) All other models are rejected at 5%. These results reinforce the correct specification of the loglog model. Again, it is not possible to use the RESET test to draw conclusions on the correct specification of a subset of regressors.

Finally, we note that the data used in this empirical were modeled by Rauber, Cribari-Neto and Bayer (2020) through a beta regression with a parametric mean link function. The authors argue that it is possible to achieve a better fit using such a model. It is not our intention here to search for the best fitting model. Instead, our goal was to determine whether the model used in the initial modeling of such data is correctly specified. The information matrix tests indicate that the model used that analysis is not misspecified.

2.6 CONCLUDING REMARKS

The beta regression model is used with dependent variables that assume values in the standard unit interval, $(0, 1)$, such as rates, proportions and concentration indices. It has been widely used by practitioners in a wide range of fields. The model comprises two submodels, one for the response mean and another for the precision, each involving a link function and a linear predictor with covariates and regression coefficients. Model misspecification can stem from the use of an incorrect mean link function, from using a precision link function that is not adequate, from leaving out an important independent variable from one of the linear predictors, from neglecting existing nonlinearities and so on.

It is of paramount importance to determine whether a fitted beta regression model is correctly specified prior to drawing inferences and conclusions from it. This is typically done through residual analysis, which involves some level of subjectivity. In this paper, we introduced two information matrix misspecification tests that can be used to that end. The null hypothesis is that the fitted beta regression model is not misspecified and the alternative hypothesis is that the model specification is in error. The test statistics we present are based on the information matrix equality, which fails to hold when the model specification is not correct. It is possible to test the overall adequacy of the model by considering all restrictions or to focus on a set of selected restrictions that are associated with key aspects of the model. We proved the validity of the tests when used with beta regressions. We also presented simulation evidence that showed that the tests perform reliably when coupled with bootstrap resampling. In particular, it is possible to achieve good control of the type I error frequency by using data resampling. Our numerical evidence also showed that the tests are able to reliably detect model misspecification, especially when the sample size is not small.

Two empirical applications were presented and discussed. They showcased the usefulness of the proposed misspecification tests in beta regression analyses. The applications relate to physiological biometrics (proportion of body fat) and environmental biometrics (proportion of religious disbelievers).

There are several directions for future research. First, it would be interesting to extend our results to cover some variants of the beta regression model, such as the model that uses a parametric mean link function. Second, alternative formulations of the information matrix test can be considered. Third, it would be of value to extend our results for dynamic beta models, i.e., for models used in time series analysis.

3 NEW STRATEGIES FOR DETECTING ATYPICAL OBSERVATIONS BASED ON THE INFORMATION MATRIX EQUALITY

3.1 INTRODUCTION

It is often of interest in empirical analyses to identify atypical observations. They may disproportionately influence the model fit and should be individually examined. In regression modeling, this detection is typically accomplished using residuals, Cook's distances or measures of local influence.

Our chief goal in this chapter is to introduce new strategies for atypical data points detection to be used whenever parameter estimation is performed via maximum likelihood. They are based on the information matrix equality, which is known to hold under correct model specification. It is possible to test if the model is correctly specified by using what is known as 'the information matrix test'; see, e.g., White (1982). We use the information matrix equality in a different fashion. Based on its sample counterpart, we create different measures of the degree of adequate model specification; the closer they are to zero, the better the model specification. The proposed criteria are then used to quantify the degree of unusualness of each observation in the sample. To accomplish that, we compare the values of the measures of adequate model specification without each observation in the sample to those obtained using the complete data.

The strategies we propose for detecting atypical data points have several novel features, e.g.: (i) they are based on a concept not yet explored in the literature, namely: the degree of adequate model specification, (ii) they embrace a new definition of atypical observations, which are the cases that disproportionately alter the degree of model adequacy, (iii) they can be used not only in regression analyses, but also when fitting probability distributions, (iv) they only require knowledge of first- and second-order log-likelihood derivatives.

Additionally, based on the information matrix equality, we introduce a modified version of Cook's generalized distance and a new criterion for atypical cases detection that employs the two Cook's distances (standard and modified).

A word of caution is in order. No atypical cases detection strategy is uniformly superior to all others, and it is not our desire or ambition to propose strategies that achieve this goal. Different measures of unusualness of observations carry different information and are often used in a complementary way. Our goal is to propose some atypical cases detection mechanisms that employ criteria that differ from those used so far, that can be easily implemented, and that are tailored to models in which inference is performed by maximum likelihood.

The chapter is organized in the following manner. The new approaches for detecting atypical observations are presented in Section 3.2. In Section 3.3, some empirical applications using Gaussian and beta regressions are presented and discussed. They use Gaussian and beta regressions. Finally, Section 3.4 contains some concluding remarks.

3.2 NEW ATYPICAL DATA POINTS DETECTION STRATEGIES

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ be a vector of independent random variables such that $Y_i \sim f(\boldsymbol{\theta}_i)$, where f is a probability density function with respect to the Lebesgue measure on an interval or the counting measure on some discrete set which is indexed by a k_1-vector of parameters, $\boldsymbol{\theta}_i$, for $i = 1, \ldots, n$. It is common to reduce the number of parameters by specifying models for the components of $\boldsymbol{\theta}_i$ which are then taken to be functions of a k-vector of parameters, say $\boldsymbol{\theta}$. By doing so, the number of parameters is reduced from $k_1 \times n$ to k $(k < n)$, and inferences are made on $\boldsymbol{\theta}$. When the random variables are identically distributed, $k_1 = k$ and $\boldsymbol{\theta}_i = \boldsymbol{\theta}$ $\forall i$.

In our setup, the parameter vector $\boldsymbol{\theta}$ is estimated by maximum likelihood. The (total) log-likelihood function for $\boldsymbol{Y}$ with observed values $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ is $\ell \equiv \ell(\boldsymbol{\theta}) \equiv \ell(\boldsymbol{\theta}; \boldsymbol{y}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}; y_i)$, where $\ell_i(\boldsymbol{\theta}; y_i)$ is the ith individual log-likelihood function, i.e., the log-density for the ith observation seen as a function of $\boldsymbol{\theta}$. The maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \mathbb{R}^k} \ell(\boldsymbol{\theta})$. Oftentimes, it cannot be expressed in closed-form, and point estimates are obtained by numerically maximizing $\ell(\boldsymbol{\theta})$ using, say, a Newton or quasi-Newton nonlinear optimization algorithm.

In regression analysis, it is commonly assumed that the probability distribution of each Y_i involves two parameters, say μ_i and ϕ_i. It is also usual to define link functions, say g_1 and g_2, such that $g_1(\mu_i) = \eta_{1i} = \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ and $g_2(\phi_i) = \eta_{2i} = \delta_1 z_{i1} + \cdots + \delta_q z_{iq}$ $(k = p + q < n)$. Oftentimes, $x_{i1} = z_{i1}$ $\forall i$. Here, η_{1i} and η_{2i} are linear predictors that contain covariates ($x_{i1}, \ldots, x_{ip}$ and $z_{i1}, \ldots, z_{iq}$, respectively) and regression coefficients ($\beta_1, \ldots, \beta_p$ and $\delta_1, \ldots, \delta_q$, respectively). The regressors in each submodel must be linearly independent. Generally, μ_i is the mean or median (or, possibly, a given quantile) of Y_i

and ϕ_i is a precision or dispersion parameter; under constant precision, $z_{i1} = 1\ \forall i$ and $\delta_2 = \cdots = \delta_q = 0$, i.e., $q = 1$. The two link functions and their inverses are often required to be strictly monotonic and twice continuously differentiable. For instance, in beta regressions, Y_i is assumed to be beta-distributed with mean μ_i and precision ϕ_i so that $\mathrm{Var}(Y_i) = \mu_i(1-\mu_i)/(1+\phi_i)$, $g_1 : (0,1) \to \mathbb{R}$, and $g_2 : \mathbb{R}_+ \to \mathbb{R}$. Here, $\boldsymbol{\theta}_i = (\mu_i, \phi_i)^\top$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\delta}^\top)^\top$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_q)^\top \in \mathbb{R}^q$; for details on such a model, see Cribari-Neto and Zeileis (2010) and Douma and Weedon (2019).

Let $\boldsymbol{U} \equiv \boldsymbol{U}(\boldsymbol{\theta})$ be the score function, i.e., $\boldsymbol{U} = \partial\ell/\partial\boldsymbol{\theta} = \sum_{i=1}^{n} \partial\ell_i(\boldsymbol{\theta}; y_i)/\partial\boldsymbol{\theta}$. Fisher's expected information matrix for a single observation is $\dot{B}_i(\boldsymbol{\theta}) = \mathbb{E}(\partial\ell(\boldsymbol{\theta}; Y_i)/\partial\boldsymbol{\theta} \times \partial\ell(\boldsymbol{\theta}; Y_i)/\partial\boldsymbol{\theta}^\top)$. It is commonly assumed that $A(\boldsymbol{\theta}) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \dot{A}_i(\boldsymbol{\theta})$ and $B(\boldsymbol{\theta}) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \dot{B}_i(\boldsymbol{\theta})$ exist, where $\dot{A}_i(\boldsymbol{\theta}) = \mathbb{E}(\partial^2\ell(\boldsymbol{\theta}; Y_i)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top)$. When the model is correctly specified, $B(\boldsymbol{\theta}_0) = -A(\boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_0$ is the true parameter value. This equation is known as 'the information matrix equality'. Hence, when the model specification is correct, $A(\boldsymbol{\theta}_0) + B(\boldsymbol{\theta}_0) = O_{k\times k}$, where $O_{k\times k}$ denotes the $k \times k$ matrix of zeros. White (1982) developed a model misspecification test known as 'the information matrix test' in which the null hypothesis of correct model specification is tested against the alternative hypothesis that the model specification is in error, i.e., $\mathcal{H}_0 : A(\boldsymbol{\theta}_0) + B(\boldsymbol{\theta}_0) = O_{k\times k}$ and $\mathcal{H}_1 : A(\boldsymbol{\theta}_0) + B(\boldsymbol{\theta}_0) \neq O_{k\times k}$. Instead of considering a dichotomous classification between well-specified and poorly specified models, we will use the information matrix equality to define measures of the degree of model adequacy in such a way that the closer those measures are to zero, the better the model adequacy.

Let $A_i(\boldsymbol{\theta}; Y_i) = \partial^2\ell(\boldsymbol{\theta}; Y_i)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ and $A_n(\boldsymbol{\theta}; \boldsymbol{Y}) = n^{-1}\sum_{i=1}^{n} A_i(\boldsymbol{\theta}; Y_i)$. Also, let $B_i(\boldsymbol{\theta}; Y_i) = \partial\ell(\boldsymbol{\theta}; Y_i)/\partial\boldsymbol{\theta} \times \partial\ell(\boldsymbol{\theta}; Y_i)/\partial\boldsymbol{\theta}^\top$ and $B_n(\boldsymbol{\theta}; \boldsymbol{Y}) = n^{-1}\sum_{i=1}^{n} B_i(\boldsymbol{\theta}; Y_i)$. Notice that $A_n(\boldsymbol{\theta}; \boldsymbol{Y})$ and $B_n(\boldsymbol{\theta}; \boldsymbol{Y})$ are the sample counterparts of $A(\boldsymbol{\theta})$ and $B(\boldsymbol{\theta})$, respectively.

At the outset, we propose measuring the distances between (i) $-A_n(\hat{\boldsymbol{\theta}}; \boldsymbol{Y})$ and $B_n(\hat{\boldsymbol{\theta}}; \boldsymbol{Y})$, and (ii) $-A_n^{-1}(\hat{\boldsymbol{\theta}}; \boldsymbol{Y})$ and $B_n^{-1}(\hat{\boldsymbol{\theta}}; \boldsymbol{Y})$. We note that the latter two matrices are estimators of the asymptotic covariance of $\hat{\boldsymbol{\theta}}$, $\mathrm{Cov}(\hat{\boldsymbol{\theta}})$. These distances can be viewed as measures of *the degree of model adequacy*, in the sense that they involve the sample counterparts of matrices that are expected to coincide in the population when the postulated model is correctly specified.

The distance between $-A_n(\hat{\boldsymbol{\theta}}; \boldsymbol{Y})$ and $B_n(\hat{\boldsymbol{\theta}}; \boldsymbol{Y})$ and that between $-A_n^{-1}(\hat{\boldsymbol{\theta}}; \boldsymbol{Y})$ and $B_n^{-1}(\hat{\boldsymbol{\theta}}; \boldsymbol{Y})$ have interesting interpretations in the canonical full rank exponential family.

Suppose $Y_1,\ldots,Y_n$ are independent and identically distributed (i.i.d.) random variables with common probability density function f with respect to the Lebesgue measure on an interval or the counting measure on a discrete set in the form

$$f(y;\boldsymbol{\xi}) = h(y)\exp\left(\sum_{j=1}^{N}\xi_j T_j(y) - \gamma(\boldsymbol{\xi})\right),$$

where $T_1,\ldots,T_N$ are linearly independent functions, $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_N)^\top$ is an unknown vector parameter and γ is a strictly convex function with continuous first and second derivatives. The score vector based on $Y_1,\ldots,Y_n$ is $\sum_{i=1}^{n}(\boldsymbol{T}(Y_i) - \nabla\gamma(\boldsymbol{\xi}))$, where ∇ is the gradient operator and $\boldsymbol{T} = (T_1,\ldots,T_N)^\top$. The MLE $\hat{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$ solves $\nabla\gamma(\hat{\boldsymbol{\xi}}) = n^{-1}\sum_{i=1}^{n}\boldsymbol{T}(Y_i)$. The Hessian matrix of the log-density with respect to $\boldsymbol{\xi}$ is simply $-H\gamma(\boldsymbol{\xi})$, where H represents the Hessian operator. Using the information equality, we conclude that the covariance matrix of $\boldsymbol{T}$ is $\mathrm{Var}(\boldsymbol{T}) = H\gamma(\boldsymbol{\xi})$. Notice that $-A_n(\hat{\boldsymbol{\xi}};\boldsymbol{Y}) = H\gamma(\hat{\boldsymbol{\xi}})$ is the MLE of $\mathrm{Var}(\boldsymbol{T})$ based on the common distribution that was assumed for the random variables, while $B_n(\hat{\boldsymbol{\xi}};\boldsymbol{Y}) = n^{-1}\sum_{i=1}^{n}(\boldsymbol{T}(Y_i) - \nabla\gamma(\hat{\boldsymbol{\xi}}))(\boldsymbol{T}(Y_i) - \nabla\gamma(\hat{\boldsymbol{\xi}}))^\top$ is the nonparametric moment estimator of the same covariance matrix. The distance between $-A_n(\hat{\boldsymbol{\xi}};\boldsymbol{Y})$ and $B_n(\hat{\boldsymbol{\xi}};\boldsymbol{Y})$ is a measure of the proximity between these two covariance matrix estimators. If it is small, we may conclude that the postulated distribution is an adequate model. Equivalently, the distance between $-A_n^{-1}(\hat{\boldsymbol{\xi}};\boldsymbol{Y})$ and $B_n^{-1}(\hat{\boldsymbol{\xi}};\boldsymbol{Y})$ measures the proximity between the MLE and a nonparametric moment estimator of the precision matrix (the inverse of the covariance matrix).

Our proposal is to measure the distance between the sample counterparts of the matrices that define the information matrix equality or their inverses. This is done for a given observed sample $\boldsymbol{y}$ in order to identify data points that are atypical. Cases that disproportionately impact the distance between the two matrices are taken to be atypical. That is, data points that substantially alter the degree of model adequacy are classified as *atypical cases*.

We will measure the distances between (i) $C_{a_1,n}(\hat{\boldsymbol{\theta}};\boldsymbol{Y}) = A_n(\hat{\boldsymbol{\theta}};\boldsymbol{Y}) + B_n(\hat{\boldsymbol{\theta}};\boldsymbol{Y})$ and $O_{k\times k}$ and (ii) $C_{a_2,n}(\hat{\boldsymbol{\theta}};\boldsymbol{Y}) = A_n^{-1}(\hat{\boldsymbol{\theta}};\boldsymbol{Y}) + B_n^{-1}(\hat{\boldsymbol{\theta}};\boldsymbol{Y})$ and $O_{k\times k}$. These distances are viewed as measures of the degree of model adequacy with smaller values indicating better model adequacy. As noted earlier, it is possible to test the null hypothesis that $A(\boldsymbol{\theta}_0) + B(\boldsymbol{\theta}_0) = O_{k\times k}$ (correct model specification) using the information matrix test (WHITE, 1982). Our goal, however, is different. For an observed sample $\boldsymbol{y}$, we seek to measure the distance between $-A_n^{-1}(\hat{\boldsymbol{\theta}};\boldsymbol{y})$ and $B_n^{-1}(\hat{\boldsymbol{\theta}};\boldsymbol{y})$, and then evaluate how such a distance is impacted by each observation in the sample, i.e., by each component of $\boldsymbol{y}$.

We need a metric for the distance between two symmetric matrices. Consider the vector space $\mathbb{R}^{r\times s}$ of $r\times s$ matrices. Different norms can be defined in such a space. A well-known norm is that of the maximum on the unit sphere induced by the norms in $\mathbb{R}^r$ and $\mathbb{R}^s$. Suppose we use the norms $\|\ \|_a$ and $\|\ \|_b$ in $\mathbb{R}^r$ and $\mathbb{R}^s$, respectively. The norm of a matrix $M\in\mathbb{R}^{r\times s}$ can be defined as $\|M\|_z=\max_{\|\boldsymbol{w}\|_b=1}\|M\boldsymbol{w}\|_a$, where $\boldsymbol{w}$ is a column vector of dimension s. It is easy to see that $\|M\|_z=\max_{\boldsymbol{w}\neq\mathbf{0}_s}\|M\boldsymbol{w}\|_a/\|\boldsymbol{w}\|_b$, where $\mathbf{0}_s$ an s-vector of zeros. When $\|\ \|_a$ and $\|\ \|_b$ are the Euclidean norms in the corresponding spaces, the norm $\|M\|_z$ will be the largest singular value of M. That is, in the case of Euclidean norms, $\|M\|_z$ corresponds to the largest square root of the eigenvalues of $M^\top M$ (which are guaranteed to be non-negative real numbers). If M is a symmetric s-dimensional matrix, this largest singular value is the maximum of the absolute values of the eigenvalues of M (which are assuredly real numbers). That is, if M is an s-dimensional symmetric matrix, we can define $\|M\|_z=\max_{1\leq j\leq s}|\lambda_j|$, where $\lambda_1,\ldots,\lambda_s$ are the eigenvalues of M. In this case, in the vector subspace of $s\times s$ symmetric matrices, we can define the distance between two matrices as the norm of the difference, i.e., the distance between two symmetric matrices M_1 and M_2 can be defined as $\|\Delta\|_z$, where $\Delta=M_1-M_2$. For details on norms of matrices, see Horn and Johnson (2012).

We note that $\|\Delta\|_z\in[0,\infty)$. If desired, a distance defined in the interval $[0,1)$ can be obtained using the following result. Let $d:M\times M\to\mathbb{R}$ be a distance on a non-empty set M and let $h:[0,\infty)\to[0,\infty)$ be an increasing concave function such that $h(0)=0$. Then, it follows that $\varphi:M\times M\to\mathbb{R}$ defined by $\varphi(x,y)=h(d(x,y))$ is also a distance. E.g., $\|\Delta\|_z/(\|\Delta\|_z+1)$.

Recall that we wish to measure the distance between $C_{a_1,n}(\hat{\boldsymbol{\theta}};\boldsymbol{Y})$ and $O_{k\times k}$ and between $C_{a_2,n}(\hat{\boldsymbol{\theta}};\boldsymbol{Y})$ and $O_{k\times k}$ for an observed sample $\boldsymbol{y}$. This can be accomplished by using

$$m_1=\|C_{a_1,n}(\hat{\boldsymbol{\theta}};\boldsymbol{y})\|_z \quad\text{and}\quad m_2=\|C_{a_2,n}(\hat{\boldsymbol{\theta}};\boldsymbol{y})\|_z,$$

respectively. The closer these measures are to zero, the better the degree of model adequacy.

We also suggest measuring the aforementioned distances using Euclidian norms.

We propose to use

$$m_3 = \|\mathrm{vech}(C_{a_1,n}(\hat{\boldsymbol{\theta}};\boldsymbol{y}))\|_2 \quad \text{and} \quad m_4 = \|\mathrm{vech}(C_{a_2,n}(\hat{\boldsymbol{\theta}};\boldsymbol{y}))\|_2,$$

where $\| \, \|_2$ denotes the Euclidian norm and vech is the operator that, when applied to a square matrix, returns the vector formed by its lower triangular portion (including the diagonal). Like the previous metric, smaller values are indicative of superior model adequacy.

The distance measures $m_1,\ldots,m_4$ are not invariant to regressor rescaling. If such an invariance is important, an alternative is to use a multiplicative specification. If $-A(\theta_0) = B(\theta_0)$, it follows that $-A^{-1}(\theta_0)B(\theta_0) = I_k$, where I_k is the k-dimensional identity matrix. We could consider measuring the distance between $-A_n^{-1}(\hat{\boldsymbol{\theta}};\boldsymbol{y})B_n(\hat{\boldsymbol{\theta}};\boldsymbol{y})$ and I_k. The former matrix, however, is not guaranteed to be symmetric, although it obtained as the product of two symmetric matrices. As a consequence, its eigenvalues are not guaranteed to be real numbers. We will then work with

$$C_{m,n}(\hat{\boldsymbol{\theta}};\boldsymbol{y}) = P_n^{-1}(\hat{\boldsymbol{\theta}};\boldsymbol{y})B_n(\hat{\boldsymbol{\theta}};\boldsymbol{y})(P_n^{-1}(\hat{\boldsymbol{\theta}};\boldsymbol{y}))^\top,$$

where $P_n(\hat{\boldsymbol{\theta}};\boldsymbol{y})$ is obtained from the Choleski decomposition of $-A_n(\hat{\boldsymbol{\theta}};\boldsymbol{y})$, i.e., the former is a lower triangular matrix such that $P_n(\hat{\boldsymbol{\theta}};\boldsymbol{y})(P_n(\hat{\boldsymbol{\theta}};\boldsymbol{y}))^\top = -A_n(\hat{\boldsymbol{\theta}};\boldsymbol{y})$. The Choleski decomposition is used because the triangular structure that is obtained makes matrix inversion easier. We propose measuring the distance between $C_{m,n}(\hat{\boldsymbol{\theta}};\boldsymbol{y})$ and I_k using

$$m_5 = \|C_{m,n}(\hat{\boldsymbol{\theta}};\boldsymbol{y}) - I_k\|_z \quad \text{and} \quad m_6 = \|\mathrm{vech}(C_{m,n}(\hat{\boldsymbol{\theta}};\boldsymbol{y}) - I_k)\|_2.$$

Let $m_{j,i}$ be the value of m_j when observation i is not in the sample, $i = 1,\ldots,n$ and $j = 1,\ldots,6$. We define the following measures of the sensitivity of the degree of model adequacy to observation i:

$$s_{j,i} = \frac{m_{j,i}}{m_j}.$$

The quantities $m_{1,i}$ through $m_{6,i}$ measure the impact that observation i exerts on the degree of model adequacy. They can be therefore used to identify atypical cases, i.e., to single out data points that disproportionately impact the degree of model adequacy.

Since the distributions of $m_{1,i},\ldots,m_{6,i}$ obtained from the random vector $\boldsymbol{Y}$ are, in general, difficult to obtain, we will adopt, for each observed sample $\boldsymbol{y}$, an ad hoc rule for atypical cases detection. Let $q_{\tau,j}$ be the τth quantile of $s_{j,i}$ and

$$\mathcal{I}_{r,j} = [v_j - z_{r,j}(q_{0.500,j} - q_{0.125,j}), v_j + z_{r,j}(q_{0.875,j} - q_{0.50,j})],$$

$r = 1,2$ and $j = 1,\ldots,6$. Notice that there are two intervals for each sensitivity measure defined so far: $\mathcal{I}_{1,j}$ and $\mathcal{I}_{2,j}$. Observation i is to be taken as atypical if $s_{i,j} \notin \mathcal{I}_{1,j}$ or, alternatively, if $s_{i,j} \notin \mathcal{I}_{2,j}$. We recommend,based on experimentation with several data sets, using $v_j = 1.0$ for $j = 1,\ldots,6$. We also recommend using $z_{1,1} = \cdots = z_{1,4} = 3.75$, $z_{1,5} = z_{1,6} = 2.5$, $z_{2,1} = \cdots = z_{2,4} = 7.50$, and $z_{2,5} = z_{2,6} = 5.0$. It should be noted that, for each j, $\mathcal{I}_{2,j}$ is wider than $\mathcal{I}_{1,j}$. The two detection intervals account for any existing skewness in the sample distribution of $s_{i,j}$. One may use $\mathcal{I}_{2,j}$ when more conservative atypical case detection is desired, or when the number of cases identified using $\mathcal{I}_{1,j}$ is deemed to be excessive. The proposed detection intervals may, of course, be tailored by users to suit their specific needs.

The measures $m_{1,i},\ldots,m_{6,i}$ introduced above are somewhat related to Cook's generalized distance:

$$D_i = (n-1)(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^\top(-A_n(\hat{\boldsymbol{\theta}}_{(i)};\boldsymbol{y}))(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}),$$

where $\hat{\boldsymbol{\theta}}_{(i)}$ and $A_{n-1}^{(i)}$ are, respectively, the maximum likelihood estimate of $\boldsymbol{\theta}$ and the quantity A_n computed when case i is not in the sample; see, e.g., Díaz-García and González-Farías (2004) and Flora, LaBrish and Chalmers (2012). For an empirical analysis based on this measure, see Cordeiro *et al.* (2021). Like $m_{1,i},\ldots,m_{4,i}$, Cook's generalized distance is also based on individual case deletion and uses an estimate of $-A_n(\boldsymbol{\theta};\boldsymbol{Y})$. A common rule-of-thumb is that observation i is taken to be atypical if D_i exceeds one.

Exploring the fact that under correct model specification $A(\boldsymbol{\theta}_0) = B(\boldsymbol{\theta}_0)$, we introduce the following modified Cook's generalized distance:

$$D_i^m = \frac{n-1}{2}(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^\top(-A_{n-1}^{(i)}(\hat{\boldsymbol{\theta}}_{(i)};\boldsymbol{y}) + B_{n-1}^{(i)}(\hat{\boldsymbol{\theta}}_{(i)};\boldsymbol{y}))(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}),$$

where $B_{n-1}^{(i)}$ corresponds to the quantity B_n calculated without the ith observation. We note that D_i^m is useful for determining whether a given observation is influential for the model fit, but it may not be able to capture its full impact. This is because the withdrawal of observation i may make $B_{n-1}^{(i)}(\hat{\boldsymbol{\theta}}_{(i)};\boldsymbol{y}) - A_{n-1}^{(i)}(\hat{\boldsymbol{\theta}}_{(i)};\boldsymbol{y})$ smaller, which, of course, does not necessarily mean an improvement in the model fit.

As an example, assume that $Y_1, \ldots, Y_n$ is an i.i.d. sample, and we want to test whether each Y_i is normally distributed with unit variance. It is not difficult to see that

$$D_i^m = \frac{n-1}{2n^2}(\bar{Y}^{(i)} - Y_i)^2(1 + \hat{v}^{(i)}),$$

where $\bar{Y}^{(i)}$ and $\hat{v}^{(i)}$ denote, respectively, the sample mean and the variance moment estimator, both computed after withdrawing observation i from the sample. Notice that values of $\hat{v}^{(i)}$ that are progressively smaller than one (the assumed value of the variance) increasingly worsen the model fit, and yet they progressively reduce the value of D_i^m. We would expect, however, the value of D_i^m to increase (not decrease) when case i worsens the model fit.

This motivates us to work with the differences $D_i^m - D_i$. Notice that $D_i^m \approx D_i$ whenever $-A_n(\hat{\boldsymbol{\theta}}_{(i)}; \boldsymbol{y}) \approx B_n(\hat{\boldsymbol{\theta}}_{(i)}; \boldsymbol{y})$, i.e., whenever observation i does not considerably alter the degree of adequate model specification. The distance between D_i^m and D_i is expected to grow as case i becomes more impactful to the degree of model adequacy. Like $m_{5,i}$ and $m_{6,i}$, the two Cook's distances are invariant to regressor rescaling.

For the above example of unit variance normal fit, we obtain

$$D_i^m - D_i = \frac{n-1}{2n^2}(\bar{Y}^{(i)} - Y_i)^2(1 - \hat{v}^{(i)}),$$

which seems to be a reasonable measure for the unusualness of case i. The three terms in the above expression can be easily interpreted. The fraction $(n-1)/2n^2$, which decays to zero as n increases, indicates that, when the number of observations is large, the impact of each individual observation will tend to be small. The term $(\bar{Y}^{(i)} - Y_i)^2$ relates to how far apart are the ith case and the overall mean of all other cases; larger values of $(\bar{Y}^{(i)} - Y_i)^2$ will lead to larger values of $D_i^m - D_i$. The term $1 - \hat{v}^{(i)}$ is the one that best describes the impact of observation i on the degree of model adequacy. If $\hat{v}^{(i)}$ becomes very different from 1, the absolute value of $D_i^m - D_i$ will tend to be large. This is coherent with the fact that a value of $\hat{v}^{(i)}$ that is very different from one yields evidence against using a unit variance distribution to represent the data.

Using the aforementioned difference, we introduce a new measure of cases atypicalness which is invariant to regressor rescaling:

$$s_{7,i} = D_i^m - D_i.$$

Observation i is considered atypical if $s_{7,i} \notin \mathcal{I}_{1,7}$ or, alternatively, if $s_{7,i} \notin \mathcal{I}_{2,7}$. These two intervals are as before with $v_7 = 0.0$, $z_{1,7} = 4.0$ and $z_{2,7} = 8.0$.

3.3 EMPIRICAL APPLICATIONS

In what follows, we will present applications of the proposed mechanisms of atypical data points detection. Atypical cases detection based on $s_{j,i}$ is performed using $\mathcal{I}_{1,j}$, for $j = 1, \ldots, 7$. Log-likelihood maximization is performed using the BFGS quasi-Newton algorithm with analytic first derivatives. All computations were performed using the Ox matrix programming language; see Doornik (2021). We performed the information matrix tests ζ_{1B} and ζ_{2B} developed in previous chapter for the empirical applications that use beta regressions. For the application in Subsection 3.3.3, we considered all restrictions when performing the tests. For those in Subsections 3.3.4 and 3.3.5, since the sample sizes are small, we first performed the tests by only considering restrictions related to mean submodel, and then carried out the tests by only considering restrictions related to the precision submodel. In all three applications, the correct model specification is not rejected at the usual significance levels.

3.3.1 Per capita spending on public schools

The interest here lies in modeling the relationship between statewide per capita spending on public schools (Y) and per capita income (x_2) in the US in 1979; the latter is scaled by 10^{-4}. Wisconsin is not considered due to missing data, and Washington, DC is included in the dataset. Hence, $n = 50$. These data were analyzed by Cribari-Neto (2004) and Cribari-Neto and Pereira (2019). Unlike what was done in their empirical analyses, we will consider an additional covariate, namely: $x_{i3} = x_{i2} \times d_i$, where d_i equals one for Southern states and zero otherwise. We will use the Gaussian linear regression model with multiplicative heteroskedasticity proposed by Harvey (1976). In particular, we assume that $Y_1, \ldots, Y_n$ are independent random variables such that $Y_i \sim \mathcal{N}(\mu_i, \phi_i)$. We use the following model:

$$\mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3},$$
$$\log(\phi_i) = \delta_1 + \delta_2 x_{i2}.$$

Parameter estimation was done by numerically maximizing the model's log-likelihood function and, except for the intercepts, all regression coefficients are different from zero at the 5% significance level according to individual z-tests; δ_1 is non-null at 10%. In particular, the p-values of the tests of $\mathcal{H}_0 : \beta_3 = 0$ and $\mathcal{H}_0 : \delta_2 = 0$ against two-sided alternative hypotheses are, respectively, 0.0175 and 0.0003. It is noteworthy that $\hat{\beta}_3 < 0$ which implies that per capita spending on public schools increases with per capita income at a slower rate in the South.

In Appendix D we give simple expressions for $A_n(\boldsymbol{\theta};\boldsymbol{y})$ and $B_n(\boldsymbol{\theta};\boldsymbol{y})$ in the Gaussian linear regression model with multiplicative heteroskedasticity. Using these expressions, we proceed to detect atypical data points in the empirical application at hand. Table 15 presents the atypical cases detected using the different detection strategies. Notice that $s_{1,i}$ and $s_{3,i}$ only single out case 2 (Alaska). According to $s_{2,i}$ and $s_{4,i}$, observations 2, 7, and 50 (Alaska, Connecticut, and Wyoming) are atypical. $s_{5,i}$ detects cases 1, 4, 19, 23, 24, 26, 31, 44 and 45 (Alabama, Arkansas, Maine, Minnesota, Mississippi, Montana, New Mexico, Utah, and Vermont), $s_{6,i}$ singles out cases 2, 19, 24, 31, 44 and 45 (Alaska, Maine, Mississippi, New Mexico, Utah, and Vermont) whereas $s_{7,i}$ detects cases 2, 26, 31, and 44 (Alaska, Montana, New Mexico, and Utah). Finally, based on D_i we conclude that observation 2 (Alaska) is atypical and D_i^m singles out cases 2 and 44 (Alaska and Utah) as atypical.

Table 15 – Atypical cases detection, per capita spending on public schools in the US.

Criterion	Atypical cases singled out
$s_{1,i}$	2
$s_{2,i}$	2, 7, 50
$s_{3,i}$	2
$s_{4,i}$	2, 7, 50
$s_{5,i}$	1, 4, 19, 23, 24, 26, 31, 44, 45
$s_{6,i}$	2, 19, 24, 31, 44, 45
$s_{7,i}$	2, 26, 31, 44
D_i	2
D_i^m	2, 44

Source: Author

Alaska (case 2) is influential for the test inference that δ_2 is non-null, i.e., for concluding that there is heteroskedasticity. When this observation is not in the data, the z test p-value becomes 0.1622, and the null hypothesis of constant dispersion (homoskedasticity) is not rejected at the 10% significance level. Except for $s_{5,i}$, all measures detect this case as atypical.

New Mexico (case 31) and Utah (case 44) are influential for the conclusion that the regression slope differs for Southern states. Without these cases in the data, the

p-values of $\mathcal{H}_0 : \beta_3 = 0$ become, respectively, 0.0658 and 0.0579. As consequence, the null hypothesis is no longer rejected at 5%. When both cases are removed from the data, the p-value of the test becomes 0.2100, and $\mathcal{H}_0$ is not rejected at 10%. Recall that the p-value of that same test is 0.0175 when all observations are used. Case 31 was detected by $s_{5,i}$, $s_{6,i}$ and $s_{7,i}$ whereas case 44 was detected by these measures and also by D_i^m.

3.3.2 Statewide per pupil spending

The amount of money allocated to public schools varies significantly across different states and is influenced by various factors. The funding that schools receive is directly related to their per student spending, which is impacted by factors such as teacher salaries and benefits. Additionally, there are several other factors that contribute to per pupil spending. In most states, instructional employee salaries and benefits make up at least 50% of the total per pupil spending. Administrative expenses and support staff costs are also included in the overall expenditure.

The interest here lies in modeling per pupil spending (Y) as a function of per capita income (x_2) in the 50 states and the District of Columbia (DC). Hence, $n = 51$. The source for the data on spending per pupil in 2023 is Education Data Initiative. The source for per capita income data in 2021 is the United States Census Bureau. As in the previous analysis, we use the Gaussian linear regression model with multiplicative heteroskedasticity. Here,

$$\mu_i = \beta_1 + \beta_2 x_{i2},$$
$$\log(\phi_i) = \delta_1 + \delta_2 x_{i2}.$$

Maximum likelihood parameter estimation was carried out numerically maximizing the model's log-likelihood function. Except for the mean submodel intercept, all parameters are non-null at the 1% significance level according to individual z tests. In Table 16 we present the cases identified as atypical by the different detection strategies. According to $s_{1,i}$ and $s_{3,i}$ [$s_{2,i}$ and $s_{4,i}$], only case 32 (New York) [case 36 (Oklahoma)] is atypical. Notice that $s_{5,i}$ singles out cases 20, 21, 32, and 48 (Maryland, Massachusetts, New York, and DC) whereas $s_{6,i}$ singles out observations 7, 21, 24, 32, 36, and 48 (Connecticut, Massachusetts, Mississippi, New York, Oklahoma, and DC). Observations 18, 32, 44, and 49 (Louisiana, New York, Utah, and West Virginia) are taken to be unusual when detection is based on $s_{7,i}$. Finally, D_i and D_i^m only single out case 32 (New York) as atypical.

There are two observations in the sample that, when individually removed, cause the estimate of δ_2 to change considerably. The relative change in $\hat{\delta}_2$ due to the removal of New York (DC) from the sample is -31.37% (25.28%). These two observations are thus influential. The only measures that simultaneously indentified the two cases as atypical were $s_{5,i}$ and $s_{6,i}$. We also note $s_{1,i}$, $s_{3,i}$, $s_{7,i}$, D_i, and D_i^m singled out New York as atypical.

Table 16 – Atypical cases detection, per pupil spending in the US.

Criterion	Atypical cases singled out
$s_{1,i}$	32
$s_{2,i}$	36
$s_{3,i}$	32
$s_{4,i}$	36
$s_{5,i}$	20, 21, 32, 48
$s_{6,i}$	7, 21, 24, 32, 36, 48
$s_{7,i}$	18, 32, 44, 49
D_i	32
D_i^m	32

Source: Author

3.3.3 Proportion of religious disbelievers worldwide

We consider the beta regression analysis presented in Cribari-Neto and Souza (2013). The interest lies in modeling the proportion of atheists (Y) in a cross-section of countries. The regressors are: (i) x_2: dummy variable that equals 1 if the majority of the population is Muslim and 0 otherwise, (ii) x_3: per capita Gross National Income adjusted for purchasing power parity, (iii) x_4: logarithm of the ratio between the volume of foreign trade (sum of total imports and exports) and the Gross Domestic Product and (iv) x_5: average intelligence quotient of the population. Using data on 124 countries ($n = 124$), the authors fitted the following varying precision beta regression model:

$$\begin{aligned}\text{loglog}(\mu_i) &= \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i5}^2,\\ \log(\phi_i) &= \delta_1 + \delta_2 x_{5i}.\end{aligned}$$

Their focus was on estimating the impact of changes in average intelligence on the mean proportion of religious disbelievers. These data were also analyzed by Cribari-Neto, Silva and Vasconcellos (2023), Guedes, Cribari-Neto and Espinheira (2020) and Rauber, Cribari-Neto and Bayer (2020). A subset of the data was modeled in Guedes, Cribari-Neto and Espinheira (2021). For details on the relationship between intelligence and religiosity, see Zuckerman, Silberman and Hall (2013). At the outset, we note that this is a challenging application for atypical cases detection due to the very high sample correlation between x_5 and x_5^2.

The model parameters were estimated by maximum likelihood. All regression coefficients are non-null according to individual z tests at the 5% significance level.

In Appendix E we provide expressions for the matrices $A_n(\boldsymbol{\theta};\boldsymbol{y})$ and $B_n(\boldsymbol{\theta};\boldsymbol{y})$ which can be used in varying precision beta regressions. We will use these expressions in the computations that follows.

Our interest lies in detecting atypical observations in the aforementioned data. The cases identified as atypical based on the different approaches are listed in Table 17.

Table 17 – Atypical cases detection, data on the prevalence of religious disbelievers worldwide.

Criterion	Atypical cases singled out
$s_{1,i}$	27, 66, 77, 78, 122
$s_{2,i}$	14, 20, 22, 27, 33, 38, 66, 77, 78, 97, 122
$s_{3,i}$	27, 66, 77, 78, 122
$s_{4,i}$	14, 20, 22, 27, 33, 38, 52, 57, 66, 77, 78, 97, 98, 122
$s_{5,i}$	16, 22, 27, 31, 33, 57, 77, 78, 122
$s_{6,i}$	16, 22, 27, 31, 57, 77, 78, 122
$s_{7,i}$	16, 27, 31, 57, 66, 71, 77, 78
D_i	77, 122
D_i^m	77, 78, 122

Source: Author

All nine criteria single out case 77 and eight out of the nine criteria single out observation 122 as atypical. They correspond to Mozambique and Vietnam, respectively. The latter has the highest response value which is considerably larger than those of the next two countries with the highest shares of religious disbelievers (Sweden and the Czech Republic, respectively). Its average intelligence and relative volume of foreign trade are, as expected, high. However, Vietnam has a low per capita income, and that goes against the positive correlation between the response and this covariate. It is thus atypical because it has the largest fraction of religious disbelievers even though it is a low-income country.

There are four (three) observations in the sample for which the sum (maximum) of the absolute percent discrepancies of the estimated slopes exceed 40% (20%): Botswana, Liberia, Mozambique and Vietnam (Botswana, Liberia and Vietnam). Botswana and Liberia are cases 14 and 66, respectively. These two cases were only simultaneously singled out by $s_{2,i}$ and $s_{4,i}$. We note that Botswana noticeably impacts the estimate of β_3, which increases by over 25% when this case is not in the sample.

Liberia (case 66) is a very influential data point: it noticeably impacts the estimate of β_4, which increases by nearly 40% when the reduced sample is used. More importantly, per capita income loses statistical significance at 5%. It thus seems that the statistical significance of per capita income when used in conjunction with average intelligence and the relative volume of foreign trade is greatly impacted by a single observation, namely: Liberia. We note that the values of the response and also of average intelligence and per capita income are small for Liberia, but its relative volume of foreign trade is quite large (it is in the upper quartile). Since the former three variables positively correlate with the latter, Liberia displays an atypical pattern. Also, it exerts considerable impact on the resulting inferences. It is noteworthy that Liberia was identified as an atypical data point by six out of the nine criteria. It was singled out by Cook's generalized distance.

In summary, cases 14, 66, 77 and 122 (Botswana, Liberia, Mozambique and Vietnam) are influential for the model fit and corresponding inferences. The only detection strategies that were able to identify all of them as atypical data points are $s_{2,i}$ and $s_{4,i}$. Additionally, we note that case 78 (Namibia) is singled out as atypical by all measures except for Cook's generalized distance (D_i). It has a noticeable impact on the statistical significance of x_3: this covariate loses statistical significance at 5% when case 78 is not in the data since the p-value of the corresponding z-test becomes 0.0788.

An advantage of atypical cases detection based on $s_{j,i}$, $j = 1, \ldots, 7$, is that it allows one to focus on specific aspects of the model, such as a subset of regressors. To exemplify, we will focus on the regressors x_5 and x_5^2 of the mean submodel and x_5 of the precision submodel. In this way, the matrices $C_{a_1,n}(\hat{\boldsymbol{\theta}};\boldsymbol{Y})$, $C_{a_2,n}(\hat{\boldsymbol{\theta}};\boldsymbol{Y})$ and $C_{m,n}(\hat{\boldsymbol{\theta}};\boldsymbol{Y})$ have dimension 3×3. There are two countries that, when they are individually removed from the sample, lead to aggregate absolute percentage changes in the relevant point estimates in excess of 20%, namely: Mozambique and Vietnam. We note that all seven detection measures singled out Mozambique (case 77) and six of those seven measures singled out Vietnam (case 122).

3.3.4 Proportion of religious disbelievers in the United States

We will now consider the beta regression analysis in Souza and Cribari-Neto (2018). The authors modeled the statewide proportions of atheists in the United States (US). This is their response variable (Y). There are 50 observations. The covariates are average intelligence quotient (x_2), a dummy variable that equals one if the state belongs to the Extended Bible Belt (defined as the Bible Belt plus Utah) and zero otherwise (x_3), percentage of Hispanic or Latino population (x_4), an income index based on personal earnings (x_5), percentage of the total population living in urban areas (x_6). The beta regression model they fitted is

$$\text{cloglog}(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 (x_{i5} \times x_{i6}),$$
$$\log(\phi_i) = \delta_1 + \delta_2 x_{i2}.$$

Maximum likelihood estimation of the regression coefficients was carried out numerically. All coefficients are different from zero at the 5% significance level according to individual z-tests. Table 18 contains the atypical cases identified by each criterion.

Table 18 – Atypical cases detection, data on the prevalence of religious disbelievers in the US.

Criterion	Atypical cases singled out
$s_{1,i}$	—
$s_{2,i}$	48
$s_{3,i}$	—
$s_{4,i}$	48
$s_{5,i}$	1, 18, 24, 42
$s_{6,i}$	1, 18, 24, 42, 45
$s_{7,i}$	1, 19, 42, 45
D_i	1, 45
D_i^m	1, 19, 45, 48

Source: Author

We note that $s_{1,i}$ and $s_{3,i}$ do not identify any observations as atypical and that according to $s_{2,i}$ and $s_{4,i}$ the only atypical observation is case 48 (West Virginia). $s_{5,i}$ singles out observations 1, 18, 24 and 42 (Alabama, Louisiana, Mississippi, and Tennessee); $s_{6,i}$ additionally identifies case 45 (Vermont). Data points 1, 19, 42 and 45 (Alabama, Maine, Tennessee, and Vermont) are atypical when detection is based on $s_{7,i}$. D_i^m singles out cases 1, 19, 45 and 48 (Alabama, Maine, Vermont, and West Virginia) whereas D_i only singles out cases 1 and 45 (Alabama and Vermont).

Alabama (case 1) is quite influential for the inferences drawn from the fitted model. The point estimates are considered altered when this case is not in the sample; e.g., the estimates of β_1, β_2 and β_4 (δ_1 and δ_2) decrease by nearly 16%, 45% and 23% (increase by over 62% and nearly 77%), respectively. Additionally, x_2 loses statistical significance at 10%.

Case 45 (Vermont) has a sizeable impact on the the estimates of δ_1 and δ_2 which decrease by over 17% and over 22% when this data point is not in the sample. We no longer conclude that the regression coefficients β_4, β_5 and β_7 are non-null at 5%.

West Virginia (case 48) has a large impact on the testing inferences. The covariates x_5, x_6 and their interaction ($x_5 \times x_6$) lose statistical significance at 10%. Additionally, x_2 loses significance at 5%. Some point estimates are also considerably impacted; e.g., the estimates of β_5, β_6 and β_7 decrease, respectively, by approximately 18%, 29% and 21% when the data do not include observation 48. It is clear that West Virginia is quite influential. It was detected as atypical by $s_{2,i}$, $s_{4,i}$ and D_i^m.

Based on the above diagnostic analysis, we consider the reduced beta regression model $\text{cloglog}(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$ and $\log(\phi_i) = \delta_1 + \delta_2 x_{i2}$. The likelihood ratio test favors this model, since its p-value for testing $\mathcal{H}_0 : \beta_5 = \beta_6 = \beta_7 = 0$ in the larger model is 0.4320. The reduced model is also favored by the Akaike and Bayesian information criteria (AIC and BIC). Interestingly, the estimate of β_2 is over 25% larger when the reduced model is used (0.0459 vs 0.0362). That is, the impact of average intelligence on the prevalence of religious disbelievers is strengthened when a more parsimonious model is used. We arrived at such a reduced model through a diagnostic analysis that identified atypical and influential data points.

3.3.5 Reading accuracy

We will now use a dataset analyzed by Smithson and Verkuilen (2006). There are 44 observations on reading accuracy of dyslexic and non-dyslexic Australian children. The response (Y) are reading accuracy indices and the independent variables are: a dummy variable that equals 1 for dyslexics and -1 for non-dyslexics (x_2), nonverbal IQ converted to z-scores (x_3), and an interaction variable ($x_4 = x_2 \times x_3$). These data were also analyzed by Bayer and Cribari-Neto (2017), Cribari-Neto and Queiroz (2014), Espinheira, Ferrari and Cribari-Neto (2008), Guedes, Cribari-Neto and Espinheira (2020) and Grün, Kosmidis and Zeileis (2012). In particular, Cribari-Neto and Queiroz (2014) proposed modeling these data using the following beta regression model:

$$\text{logit}(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4},$$
$$\log(\phi_i) = \delta_1 + \delta_2 x_{i2} + \delta_3 x_{i3} + \delta_4 x_{i3}^2.$$

We estimated the above model and computed the measures of atypical cases detection. Table 19 presents the atypical data points detected using the different approaches.

Table 19 – Atypical cases detection, data on reading accuracy.

Criterion	Atypical cases singled out
$s_{1,i}$	32, 33
$s_{2,i}$	26, 32, 35
$s_{3,i}$	32, 33
$s_{4,i}$	26, 32, 35
$s_{5,i}$	31, 32, 33, 39
$s_{6,i}$	32, 33
$s_{7,i}$	14, 19, 24, 32, 23
D_i	32, 33, 38, 39
D_i^m	13, , 38, 39

Source: Author

There are only two observations for which the sum of the absolute percent discrepancies of the estimated slopes exceeds 40%: 24 and 32. The latter was identified by all measures. The former was only identified by $s_{7,i}$.

A remark on case 32 is in order. It is influential since the estimates for δ_2 and δ_3 change by 15.67% and 11.87%, respectively, when it is not in the sample. This case corresponds to a dyslexic child with high IQ z-score (0.7090, larger than the third quartile) but with very low reading accuracy index (0.5405, smaller than the first quartile). This case thus goes against the positive correlation between y_i and x_{i3} which equals, respectively, 0.5679 and 0.6056 with and without case 32 in the data. All nine detection strategies were able to identify this observation as atypical.

Case 24 is also noteworthy. It corresponds to a non-dyslexic child with below average reading accuracy and IQ z-score close to the third quartile. The estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ increase, respectively, by nearly 11% and nearly 17% when case 24 is not in the data. The value predicted by the model for this case is quite distant from the observed value (observed, y_{24}: 0.6466; predicted, $\hat{\mu}_{24}$: 0.9349). The atypical nature of observation 24 translates into an inaccurate model prediction because it impacts parameter estimates of the mean submodel. This atypical case was detected by D_i^m and $s_{7,i}$. Interestingly, unlike Cook's generalized distance, both measures based on it that we introduced (D_i^m and $s_{7,i}$) were able to single out case 24.

3.4 CONCLUDING REMARKS

We introduced a new approach for identifying atypical observations in empirical analyses that are based on maximum likelihood inference. We defined measures of adequate model specification in a way that smaller values are indicative of better model specification. Such measures follow from the information matrix equality, which holds when the model is correctly specified. This equality is the basis for information matrix misspecification tests which are commonly used to determine whether the specification of a given model is in error. Our approach considers that there are different degrees of adequate model specification which are coupled to the distance between the sample counterparts of the matrices that define the information matrix equality. We introduced several measures of model specification adequacy and showed that they can be used to identify atypical data points. Such points are those that disproportionately alter the degree of model adequacy when removed from the data.

We presented empirical applications involving beta and Gaussian regression models. Overall, the proposed detection strategies were able to identify influential observations, i.e., cases that substantially impact the resulting inferences.

4 CONCLUSION

This PhD dissertation focused on the information matrix equality, which is known to hold when the model is correctly specified. In this context, our focus was on modeling double bounded random variables. At first, we presented three information matrix test statistics for univariate beta models. We showed results from a set of Monte Carlo simulations, carried out to evaluate the tests' finite sample performance. We also presented an application of the proposed tests to state and county COVID-19 mortality data in the United States.

In the following, we developed information matrix tests for the varying precision beta regression model. The null hypothesis is that of correct specification of the fitted beta regression model. It is tested against the alternative hypothesis that the model specification is in error. We obtained two information matrix test statistics. They use different estimators of the covariance matrix of a given random vector. We proved the consistency of both covariance matrix estimators in the class of beta regressions. We also presented the results of extensive Monte Carlo simulations. They showed that the tests display good control of the type I error frequency when bootstrap resampling is used. Different sources of model misspecification were considered when the data were generated under the alternative hypothesis. The numerical evidence we presented showed that the two information matrix tests can reliably detect that the fitted model is incorrectly specified, especially when the sample size is not small. Two empirical applications were presented and discussed.

Using the sample counterparts of the matrices that define the information matrix equality, we introduced a new concept, namely the degree of adequate model specification. We argued that it equals the distance between two suitably defined symmetric matrices. More specifically, we proposed three definitions of the degree of adequate model specification and used two matrix norms to quantify each of them. We then defined measures of sensitivity of the degree of adequate model specification to each observation in the sample. Such measures can be used to detect atypical data points. We also introduced a modified version of Cook's generalized distance. The proposed criteria were used in empirical applications that employ Gaussian and beta regressions. The results showed that they are capable of reliably detecting atypical cases in the data.

REFERENCES

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.

BAYER, F. M.; TONDOLO, C. M.; MÜLLER, F. M. Beta regression control chart for monitoring fractions and proportions. **Computers & Industrial Engineering**, v. 119, p. 416–426, 2018.

BENEDETTI, T. R. B.; ANTUNES, P. C.; nEZ, C. R. R.-A.; MAZO, G. Z.; PETROSKI, E. L. Reprodutibilidade e validade do Questionário Internacional de Atividade Física (IPAQ) em homens idosos. **Revista Brasileira de Medicina do Esporte**, v. 13, n. 1, p. 11–16, 2007.

BRAUN, H. A simple method for testing goodness of fit in the presence of nuisance parameters. **Journal of the Royal Statistical Society B**, v. 42, n. 1, p. 53–63, 1980.

BURY, K. **Statistical Distributions in Engineering**. New York: Cambridge University Press, 1999.

BUUREN, S. van; FREDRIKS, M. Worm plot: A simple diagnostic device for modelling growth reference curves. **Statistics in Medicine**, v. 20, n. 8, p. 1259–1277, 2001.

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

CAPANU, M.; PRESNELL, B. Misspecification tests for binomial and beta-binomial models. **Statistics in Medicine**, v. 27, n. 14, p. 2536–2554, 2008.

CHESHER, A. The information matrix test: Simplified calculation via a score test interpretation. **Economics Letters**, v. 13, n. 1, p. 45–48, 1983.

CHESHER, A.; SPADY, R. Asymptotic expansions of the information matrix test statistic. **Econometrica**, v. 56, n. 3, p. 787–815, 1991.

CHUA, K. C.; ONG, S. H. Test of misspecification with application to negative binomial distribution. **Computational Statistics**, v. 28, n. 3, p. 993–1009, 2013.

CORDEIRO, G. M.; FIGUEIREDO, D.; SILVA, L.; ORTEGA, E. M. M.; PRATAVIERA, F. Explaining COVID-19 mortality rates in the first wave in Europe. **Model Assisted Statistics and Applications**, v. 16, p. 211–221, 2021.

CRIBARI-NETO, F. On the corrections to information matrix tests. **Econometric Reviews**, v. 16, n. 1, p. 39–53, 1997.

CRIBARI-NETO, F. Asymptotic inference under heteroskedasticity of unknown form. **Computational Statistics & Data Analysis**, v. 45, n. 2, p. 215–233, 2004.

CRIBARI-NETO, F. A beta regression analysis of COVID-19 mortality in Brazil. **Infectious Disease Modelling**, v. 8, n. 2, p. 309–317, 2023.

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. Disponível em: <http://dx.doi.org/10.1080/02664763.2014.993368>.

CRIBARI-NETO, F.; PEREIRA, I. F. S. Testing inference in heteroskedastic linear regressions: a comparison of two alternative approaches. **Journal of Statistical Computation and Simulation**, v. 89, n. 8, p. 1437–1465, 2019.

CRIBARI-NETO, F.; QUEIROZ, M. P. F. On testing inference in beta regressions. **Journal of Statistical Computation and Simulation**, v. 84, n. 1, p. 186–203, 2014. Disponível em: <http://dx.doi.org/10.1080/00949655.2012.700456>.

CRIBARI-NETO, F.; SCHER, V. T.; BAYER, F. M. Beta autoregressive moving average model selection with application to modeling and forecasting stored hydroelectric energy. **International Journal of Forecasting**, 2022. Forthcoming.

CRIBARI-NETO, F.; SILVA, J. J. Santana-e; VASCONCELLOS, K. L. P. **Beta regression misspecification tests**. [S.l.], 2023. (UFPE Working Paper).

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

CRIBARI-NETO, F.; ZEILEIS, A. Beta regression in R. **Journal of Statistical Software**, v. 34, n. 2, p. 1–24, 2010.

DAVIDSON, R.; MACKINNON, J. G. Graphical methods for investigating the size and power of hypothesis tests. **The Manchester School**, v. 66, n. 1, p. 1–26, 1998.

DEURENBERG, P.; WESTSTRATE, J. A.; SEIDELL, J. C. Body mass index as a measure of body fatness: Age-and sex-specific prediction formulas. **British Journal of Nutrition**, v. 65, n. 2, p. 105–114, 1991.

DHAENE, G.; HOORELBEKE, G. The information matrix test with bootstrap-based covariance matrix estimation. **Economics Letters**, v. 82, n. 3, p. 341–347, 2004.

DÍAZ-GARCÍA, J. A.; GONZáLEZ-FARÍAS, G. A note on Cook's distance. **Journal of Statistical Planning and Inference**, v. 120, n. 1-2, p. 119–136, 2004.

DOORNIK, J. A. **An Object-oriented Matrix Language Ox 9**. Richmond: Timberlake Consultants Press, 2021.

DOUMA, J. C.; WEEDON, J. T. Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. **Methods in Ecology and Evolution**, v. 10, n. 9, p. 1412–1430, 2019.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, v. 5, n. 3, p. 236–244, 1996.

ELMER, J.; JONES, B. L.; NAGIN, D. S. Using the beta distribution in group-based trajectory models. **BMC Medical Research Methodology**, v. 18, n. 1, p. 152, 2018.

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. Influence diagnostics in beta regression. **Computational Statistics & Data Analysis**, v. 52, n. 9, p. 4417–4431, 2008.

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. Erratum: 32, 2017, 1777.

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.; ESPINHEIRA, P. L.; CRIBARI-NETO, F. Diagnostic tools in beta regression with varying dispersion. **Statistica Neerlandica**, v. 65, n. 3, p. 337–351, 2011.

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.

FLORA, D. B.; LABRISH, C.; CHALMERS, R. P. Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. **Frontiers in Psychology**, v. 3, p. 55, 2012.

FURNO, M. The information matrix test in the linear regression with ARMA errors. **Journal of the Italian Statistical Society**, v. 5, n. 3, p. 369–385, 1996.

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.

GUEDES, A. C.; CRIBARI-NETO, F.; ESPINHEIRA, P. L. Modified likelihood ratio tests for unit gamma regressions. **Journal of Applied Statistics**, v. 47, n. 9, p. 1562–1586, 2020.

GUEDES, A. C.; CRIBARI-NETO, F.; ESPINHEIRA, P. L. Bartlett-corrected tests for varying precision beta regressions with application to environmental biometrics. **PLOS ONE**, v. 16, n. 6, p. e0253349, 2021.

HALL, A. The information matrix test for the linear model. **Review of Economic Studies**, v. 54, n. 2, p. 257–263, 1987.

HARVEY, A. C. Estimating regression models with multiplicative heteroscedasticity. **Econometrica**, v. 44, n. 3, p. 461–465, 1976.

HORN, R. A.; JOHNSON, C. R. **Matrix Analysis**. 2nd. ed. New York: Cambridge University Press, 2012. ISBN 978-0521839402.

HOROWITZ, J. L. Bootstrap-based critical values for the information matrix test. **Journal of Econometrics**, v. 61, n. 2, p. 395–411, 1994.

JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. **Univariate Continuous Distributions**. 2nd. ed. New York: Wiley, 1995. v. 2.

JONES, M. C. Kumaraswamy distribution: A beta-type distribution with some tractability advantages. **Statistical Methodology**, v. 6, n. 1, p. 70–81, 2009.

JØRGENSEN, B. **The Theory of Dispersion Models**. London: Chapman and Hall, 1997.

LANCASTER, T. The covariance matrix of the information matrix test. **Econometrica**, v. 52, n. 4, p. 1051–1053, 1984.

LEUNG, S. F.; YU, S. How effective are the RESET tests for omitted variables? **Communications in Statistics - Theory and Methods**, v. 29, n. 4, p. 879–902, 2000.

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.

MATSUDO, S.; ARAÚJO, T.; MATSUDO, V.; ANDRADE, D.; ANDRADE, E.; OLIVEIRA, L.; BRAGGION, G. Questionário internacional de atividade física (IPAQ): Estudo de validade e reprodutibilidade no Brasil. **Revista Brasileira de Atividade Física & Saúde**, v. 6, n. 2, p. 5–18, 2001.

MAZUCHELI, J.; ALVES, B.; KORKMAZ, M. c.; LEIVA, V. Vasicek quantile and mean regression models for bounded data: New formulation, mathematical derivations, and numerical applications. **Mathematics**, v. 10, n. 9, p. 1389, 2022.

MAZUCHELI, J.; LEIVA, V.; ALVES, B.; MENEZES, A. F. B. A new quantile regression for modeling bounded data under a unit Birnbaum-Saunders distribution with applications in medicine and politics. **Symmetry**, v. 13, n. 4, p. 682, 2021.

MAZUCHELI, J.; MENEZES, A. F. B.; GHITANY, M. E. The unit-Weibull distribution and associated inference. **Journal of Applied Probability and Statistics**, v. 13, n. 2, p. 1–22, 2018.

MITNIK, P. A.; BAEK, S. The Kumaraswamy distribution: Median-dispersion reparameterizations for regression modeling and simulation-based estimation. **Statistical Papers**, v. 54, n. 1, p. 177–192, 2013.

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**. 2nd. ed. New York: Springer, 2006.

OGUAMANAM, D. C. D.; MARTIN, H. R.; HUISSOON, J. P. On the application of the beta distribution to gear damage analysis. **Applied Acoustics**, v. 45, n. 3, p. 247–261, 1995.

ORME, C. The calculation of the information matrix test for binary data models. **The Manchester School of Economic & Social Studies**, v. 56, n. 4, p. 370–376, 1988.

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. Erratum: 55, 2011, 2445.

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

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. Disponível em: <http://dx.doi.org/10.1080/03610918.2012.712183>.

PETTERLE, R. R.; BONAT, W. H.; SCARPIN, C. T.; JONASSON, T.; BORBA, V. Z. C. Multivariate quasi-beta regression models for continuous bounded data. **International Journal of Biostatistics**, v. 17, n. 1, p. 39–53, 2021.

PRO, G.; GILBERT, P. A.; BALDWIN, J. A.; BROWN, C. C.; YOUNG, S.; ZALLER, N. Multilevel modeling of county-level excessive alcohol use, rurality, and COVID-19 case fatality rates in the US. **PLOS ONE**, v. 16, n. 6, p. e0253466, 2021.

PROKHOROV, A.; SCHEPSMEIER, U.; ZHU, Y. Generalized information matrix tests for copulas. **Econometric Reviews**, v. 38, n. 9, p. 1024–1054, 2019.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2023. Disponível em: <https://www.R-project.org/>.

RAMSEY, J. B. Tests for specification errors in classical linear least-squares regression analysis. **Journal of the Royal Statistical Society B**, v. 31, n. 2, p. 350–371, 1969.

RAUBER, C.; CRIBARI-NETO, F.; BAYER, F. M. Improved testing inferences for beta regressions with parametric mean link function. **AStA Advances in Statistical Analysis**, v. 104, n. 4, p. 687–717, 2020.

RIBEIRO, T. F.; CORDEIRO, G. M.; PEÑA-RAMÍREZ, F. A.; GUERRA, R. R. A new quantile regression for the covid-19 mortality rates in the united states. **Computational and Applied Mathematics**, Springer, v. 40, n. 7, p. 1–16, 2021.

ROCHA, A. V.; CRIBARI-NETO, F. Beta autoregressive moving average models. **TEST**, v. 18, p. 529–545, 2009. Erratum: 26, 2017, 451-459.

ROCHA, A. V.; CRIBARI-NETO, F. Erratum to: Beta autoregressive moving average models. **TEST**, v. 26, p. 451–459, 2017.

SCHER, V. T.; CRIBARI-NETO, F.; PUMI, G.; BAYER, F. M. Goodness-of-fit tests for βARMA hydrological time series modeling. **Environmetrics**, v. 31, n. 3, p. e2607, 2020.

SILVA, J. J. Santana-e; CRIBARI-NETO, F.; VASCONCELLOS, K. L. P. Beta distribution misspecification tests with application to Covid-19 mortality rates in the United States. **PLOS ONE**, v. 17, n. 9, p. e0274781, 2022.

SMITHSON, M.; VERKUILEN, J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. **Psychological Methods**, v. 11, n. 1, p. 54–71, 2006.

SOUZA, T. C.; CRIBARI-NETO, F. Intelligence, religiosity and homosexuality non-acceptance: Empirical evidence. **Intelligence**, v. 52, p. 63–70, 2015.

SOUZA, T. C.; CRIBARI-NETO, F. Intelligence and religious disbelief in the United States. **Intelligence**, v. 68, p. 48–57, 2018.

SULAIMAN, M. Y.; OO, W. M. H.; WAHAB, M. A.; ZAKARIA, A. Application of beta distribution model to Malaysian sunshine data. **Renewable Energy**, v. 18, n. 4, p. 573–579, 1999.

SWEARINGEN, C. J.; TILLEY, B. C.; ADAMS, R. J.; RUMBOLDT, Z.; NICHOLAS, J. S.; BANDYOPADHYAY, D.; WOOLSONE, R. F. Application of beta regression to analyze ischemic stroke volume in NINDS rt-PA clinical trials. **Neuroepidemiology**, v. 37, n. 2, p. 73–82, 2011.

WHITE, H. Maximum likelihood estimation of misspecified models. **Econometrica**, v. 50, n. 1, p. 1–25, 1982. Erratum: 51, 1983, 513.

WHITE, H. **Asymptotic Theory for Econometricians**. 2nd. ed. New York: Academic Press, 2001. ISBN 0127466525.

WILEY, J. A.; HERSCHKORN, S.; PADIAN, N. S. Heterogeneity in the probability of HIV transmission per sexual contact: The case of male-to-female transmission in penile-vaginal intercourse. **Statistics in Medicine**, v. 8, n. 1, p. 93–102, 1989.

ZHANG, B. An information matrix test for logistic regression models based on case-control data. **Biometrika**, v. 88, n. 4, p. 921–932, 2001.

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 – INFORMATION MATRIX TEST QUANTITIES FOR THE BETA DISTRIBUTION

We present below the quantities required to compute the information matrix test statistics for the beta model. It is possible to show that

$$
\begin{aligned}
&w = \psi'(\mu\phi) + \psi'((1-\mu)\phi), \quad c = \phi[\mu w - \psi'((1-\mu)\phi)],\\
&m = \psi''(\mu\phi) - \psi''((1-\mu)\phi), \quad \frac{\partial \mu^*}{\partial \mu} = \phi\psi'(\mu\phi) + \phi\psi'((1-\mu)\phi) = \phi w,\\
&\frac{\partial \mu^*}{\partial \phi} = \mu\psi'(\mu\phi) - (1-\mu)\psi'((1-\mu)\phi) = \frac{c}{\phi},\\
&\frac{\partial^2 \mu^*}{\partial \phi^2} = \mu^2\psi''(\mu\phi) - (1-\mu)^2\psi''((1-\mu)\phi),\\
&\frac{\partial \mu^\dagger}{\partial \mu} = -\phi\psi'((1-\mu)\phi), \quad \frac{\partial \mu^\dagger}{\partial \phi} = (1-\mu)\psi'((1-\mu)\phi) - \psi'(\phi),\\
&\frac{\partial w}{\partial \mu} = \phi\psi''(\mu\phi) - \phi\psi''((1-\mu)\phi) = \phi m,\\
&\frac{\partial w}{\partial \phi} = \mu\psi''(\mu\phi) + (1-\mu)\psi''((1-\mu)\phi), \quad \frac{\partial c}{\partial \mu} = \phi\left(w + \phi\frac{\partial w}{\partial \phi}\right),\\
&\frac{\partial c}{\partial \phi} = \frac{\partial \mu^*}{\partial \phi} + \phi\frac{\partial^2 \mu^*}{\partial \phi^2}.
\end{aligned}
$$

Additionally, we obtain, after some algebra,

$$
\nabla \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y}) = \frac{1}{n}\sum_{t=1}^{n}
\begin{bmatrix}
\frac{\partial d_1(\boldsymbol{\theta}; Y_t)}{\partial \mu} & \frac{\partial d_1(\boldsymbol{\theta}; Y_t)}{\partial \phi}\\
\frac{\partial d_2(\boldsymbol{\theta}; Y_t)}{\partial \mu} & \frac{\partial d_2(\boldsymbol{\theta}; Y_t)}{\partial \phi}\\
\frac{\partial d_3(\boldsymbol{\theta}; Y_t)}{\partial \mu} & \frac{\partial d_3(\boldsymbol{\theta}; Y_t)}{\partial \phi}
\end{bmatrix},
$$

where

$$
\begin{aligned}
\frac{\partial d_1(\boldsymbol{\theta}; Y_t)}{\partial \mu} &= \phi^2\left[-\frac{\partial w}{\partial \mu} - 2(Y_t^* - \mu^*)\left(\frac{\partial \mu^*}{\partial \mu}\right)\right],\\
\frac{\partial d_1(\boldsymbol{\theta}; Y_t)}{\partial \phi} &= 2\phi\left[(Y_t^* - \mu^*)^2 - w\right] + \phi^2\left[-2(Y_t^* - \mu^*)\left(\frac{\partial \mu^*}{\partial \phi}\right) - \frac{\partial w}{\partial \phi}\right],\\
\frac{\partial d_2(\boldsymbol{\theta}; Y_t)}{\partial \mu} &= -\frac{\partial \mu^*}{\partial \mu} - \frac{\partial c}{\partial \mu} - \phi\frac{\partial \mu^*}{\partial \mu}\left[\mu(Y_t^* - \mu^*) + (Y_t^\dagger - \mu^\dagger)\right]\\
&\quad + \phi(Y_t^* - \mu^*)\left[(Y_t^* - \mu^*) - \mu\frac{\partial \mu^*}{\partial \mu} - \frac{\partial \mu^\dagger}{\partial \mu}\right],\\
\frac{\partial d_2(\boldsymbol{\theta}; Y_t)}{\partial \phi} &= -\frac{\partial \mu^*}{\partial \phi} - \frac{\partial c}{\partial \phi} + (Y_t^* - \mu^*)\left[\mu(Y_t^* - \mu^*) + (Y_t^\dagger - \mu^\dagger)\right]
\end{aligned}
$$

$$
\begin{aligned}
&\quad - \phi \left(\frac{\partial \mu^*}{\partial \phi} \right) \left[\mu (Y_t^* - \mu^*) + (Y_t^\dagger - \mu^\dagger) \right] - \phi \, (Y_t^* - \mu^*) \left[\mu \frac{\partial \mu^*}{\partial \phi} + \frac{\partial \mu^\dagger}{\partial \phi} \right], \\
\frac{\partial d_3(\boldsymbol{\theta}; Y_t)}{\partial \mu} &= -\frac{c}{\phi} - \frac{\mu}{\phi} \frac{\partial c}{\partial \mu} + \left[\psi'((1-\mu)\phi) + (1-\mu)\phi \psi''((1-\mu)\phi) \right], \\
&\quad + 2 \left[\mu (Y_t^* - \mu^*) + (Y_t^\dagger - \mu^\dagger) \right] \left[(Y_t^* - \mu^*) - \mu \frac{\partial \mu^*}{\partial \phi} - \frac{\partial \mu^\dagger}{\partial \phi} \right], \\
\frac{\partial d_3(\boldsymbol{\theta}; Y_t)}{\partial \phi} &= \frac{\mu c}{\phi^2} - \frac{\mu}{\phi} \frac{\partial c}{\partial \phi} - (1-\mu)^2 \psi''((1-\mu)\phi) + \psi''(\phi) \\
&\quad - 2 \left[\mu (Y_t^* - \mu^*) + (Y_t^\dagger - \mu^\dagger) \right] \left[\mu \frac{\partial \mu^*}{\partial \phi} + \frac{\partial \mu^\dagger}{\partial \phi} \right].
\end{aligned}
$$

Recall that we only use the first two rows of $\nabla \boldsymbol{D}_n(\boldsymbol{\theta}; \boldsymbol{Y})$ (evaluated at $\hat{\boldsymbol{\theta}}$) in the information matrix test statistics.

APPENDIX B – INFORMATION MATRIX TEST QUANTITIES FOR THE BETA REGRESSION MODEL

We present below the quantities required to compute the information matrix test statistics for the varying precision beta regression model. We have

$$\frac{d\mu_i}{d\eta_{1i}} = \frac{1}{g'(\mu_i)}, \quad \frac{d}{d\mu_i}\frac{d\mu_i}{d\eta_{1i}} = \frac{-g_1''(\mu_i)}{(g_1'(\mu_i))^2}, \quad \frac{d}{d\mu_i}\left(\frac{d\mu_i}{d\eta_{1i}}\right)^2 = \frac{-2g_1''(\mu_i)}{(g_1'(\mu_i))^3},$$
$$\frac{d^2}{d\mu_i^2}\frac{d\mu_i}{d\eta_i} = \frac{-g_1'''(\mu_i)g_1'(\mu_i) + 2(g_1''(\mu_i))^2}{(g_1'(\mu_i))^3}, \quad \frac{d\phi_i}{d\eta_{2i}} = \frac{1}{g_2'(\phi_i)}, \quad \frac{d}{d\phi_i}\frac{d\phi_i}{d\eta_{2i}} = \frac{-g_2''(\phi_i)}{(g_1'(\phi_i))^2},$$
$$\frac{d}{d\phi_i}\left(\frac{d\phi_i}{d\eta_{2i}}\right)^2 = \frac{-2g_2''(\phi_i)}{(g_2'(\phi_i))^3}, \quad \frac{d^2}{d\phi_i^2}\frac{d\phi_i}{d\eta_{2i}} = \frac{-g_2'''(\phi_i)g_2'(\phi_i) + 2(g_2''(\phi))^2}{(g_2'(\phi_i))^3}.$$

Let $w_i = \psi'(\mu_i\phi_i) + \psi'((1-\mu_i)\phi_i)$, $c_i = \phi_i[\mu_i w_i - \psi'((1-\mu_i)\phi_i)]$ and also

$$p_i = (1-\mu_i)^2\psi'((1-\mu_i)\phi_i) + \mu^2\psi'(\mu_i\phi_i) - \psi'(\phi_i), \quad a_i = \frac{d}{d\mu_i}\frac{d\mu_i}{d\eta_{1i}}\left(\frac{d\mu_i}{d\eta_{1i}}\right)^2,$$
$$b_i = \frac{d\mu_i}{d\eta_{1i}}\left[\left(\frac{d^2}{d\mu_i^2}\frac{d\mu_i}{d\eta_{1i}}\right)\frac{d\mu_i}{d\eta_{1i}} + \left(\frac{d}{d\mu_i}\frac{d\mu_i}{d\eta_{1i}}\right)^2\right], \quad m_i = \psi''(\mu_i\phi_i) - \psi''((1-\mu_i)\phi_i),$$
$$t_i = \left(\frac{d}{d\phi_i}\frac{d\phi_i}{d\eta_{2i}}\right)\left(\frac{d\phi_i}{d\eta_{2i}}\right)^2,$$
$$v_i = \frac{d\phi_i}{d\eta_{1i}}\left[\left(\frac{d^2}{d\phi_i^2}\frac{d\phi_i}{d\eta_{2i}}\right)\frac{d\phi_i}{d\eta_{2i}} + \left(\frac{d}{d\phi_i}\frac{d\phi_i}{d\eta_{2i}}\right)^2 - 2p_i\left(\frac{d\phi_i}{d\eta_{2i}}\right)^2\right],$$

where ψ' and ψ'' are the trigamma and tetragamma functions, respectively. The following derivatives are needed for obtaining $\boldsymbol{d}_i(\boldsymbol{\theta})$ and its Jacobian matrix:

$$\frac{\partial\mu_i^*}{\partial\mu_i} = \phi_i w_i, \quad \frac{\partial\mu_i^*}{\partial\phi_i} = \frac{c_i}{\phi_i}, \quad \frac{\partial\mu_i^\dagger}{\partial\mu_i} = -\phi_i\psi'((1-\mu_i)\phi_i),$$
$$\frac{\partial\mu_i^\dagger}{\partial\phi_i} = (1-\mu_i)\psi'((1-\mu_i)\phi_i) - \psi'(\phi_i),$$
$$\frac{\partial^2\mu_i^*}{\partial\phi_i^2} = \mu_i^2\psi''(\mu_i\phi_i) - (1-\mu_i)^2\psi''((1-\mu_i)\phi_i), \quad \frac{\partial w_i}{\partial\mu_i} = \phi_i m_i,$$
$$\frac{\partial w_i}{\partial\phi_i} = \mu_i\psi''(\mu_i\phi_i) + (1-\mu_i)\psi''((1-\mu_i)\phi_i), \quad \frac{\partial m_i}{\partial\mu_i} = \phi_i\psi'''(\mu_i\phi_i) + \phi_i\psi'''((1-\mu_i)\phi_i),$$
$$\frac{\partial m_i}{\partial\phi_i} = \mu_i\psi'''(\mu_i\phi_i) - (1-\mu_i)\psi'''((1-\mu_i)\phi_i), \quad \frac{\partial c_i}{\partial\mu_i} = \phi_i\left(w_i + \phi_i\frac{\partial w_i}{\partial\phi_i}\right),$$
$$\frac{\partial c_i}{\partial\phi_i} = \frac{\partial\mu_i^*}{\partial\phi_i} + \phi_i\frac{\partial^2\mu_i^*}{\partial\phi_i^2}, \quad \frac{\partial p_i}{\partial\mu_i} = 2\frac{\partial\mu_i^*}{\partial\phi_i} + \phi_i\frac{\partial^2\mu_i^*}{\partial\phi_i^2},$$
$$\frac{\partial p_i}{\partial\phi_i} = (1-\mu_i)^3\psi''((1-\mu_i)\phi_i) + \mu_i^3\psi''(\mu_i\phi_i) - \psi''(\phi_i).$$

Also, let

$$e_i = \frac{\partial \mu_i^*}{\partial \mu_i} - \frac{\partial c_i}{\partial \mu_i}, \quad h_i = \mu_i \frac{\partial \mu_i^*}{\partial \mu_i} + \frac{\partial \mu_i^\dagger}{\partial \mu_i}, \quad r_i = \mu_i \frac{\partial \mu_i^*}{\partial \phi_i} + \frac{\partial \mu_i^*}{\partial \phi_i}.$$

The partial derivatives of the matrix $C_i(\boldsymbol{\theta})$ are

$$\begin{aligned}
\frac{\partial C_i^{(r,s)}}{\partial \beta_t} =& \Bigg\{ \phi_i^2 \left[-3w_i + 2(y_i^* - \mu_i^*)\right] a_i - \phi_i^3 \left(\frac{d\mu_i}{d\eta_{1i}}\right)^3 \left[m_i + 2(y_i^* - \mu_i^*) w_i\right] \\
& + \phi_i (y_i^* - \mu_i^*) b_i \Bigg\} x_{ir} x_{is} x_{it}, \\
\frac{\partial C_i^{(r,R)}}{\partial \beta_s} =& \frac{d\phi_i}{d\eta_{2i}} \frac{d\mu_i}{d\eta_{1i}} \Bigg[\left(\frac{d}{d\mu_i}\frac{d\mu_i}{d\eta_{1i}}\right) \Bigg\{ \phi_i (y_i^* - \mu_i^*) \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)\right] \\
& + y_i^* - \mu_i^* - c_i \Bigg\} + \frac{d\mu_i}{d\eta_{1i}} \Bigg(\phi_i \Bigg\{ - \frac{\partial \mu_i^*}{\partial \mu_i} \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)\right] \\
& + (y_i^* - \mu_i^*)(y_i^* - \mu_i^* - c_i) \Bigg\} - e_i \Bigg) \Bigg] x_{ir} z_{iR} x_{is}, \\
\frac{\partial C_i^{(R,S)}}{\partial \beta_r} =& \frac{d\mu_i}{d\eta_{1i}} \Bigg\{ -\frac{\partial p_i}{\partial \mu_i} \left(\frac{d\phi_i}{d\eta_{2i}}\right)^2 + \frac{d}{d\phi_i}\frac{d\phi_i}{d\eta_{2i}} \left(\frac{d\phi_i}{d\eta_{2i}}\right) (y_i^* - \mu_i^* - h_i) \\
& + 2\left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)\right] (y_i^* - \mu_i^* - h_i) \left(\frac{d\phi_i}{d\eta_{2i}}\right)^2 \Bigg\} z_{iR} z_{iS} x_{ir}, \\
\frac{\partial C_i^{(r,s)}}{\partial \delta_R} =& \frac{d\phi_i}{d\eta_{2i}} \Bigg\{ \left(\frac{d\mu_i}{d\eta_{1i}}\right)^2 \left[-2\phi_i w_i - \phi_i^2 \frac{\partial w_i}{\partial \phi_i} + 2\phi_i (y_i^* - \mu_i^*)^2 - 2\phi_i c_i (y_i^* - \mu_i^*)\right] \\
& + \frac{d\mu_i}{d\eta_{1i}} \left(\frac{d}{d\mu_i}\frac{d\mu_i}{d\eta_{1i}}\right) (y_i^* - \mu_i^* - c_i) \Bigg\} x_{ir} x_{is} z_{iR}, \\
\frac{\partial C_i^{(r,R)}}{\partial \delta_S} =& \frac{d\mu_i}{d\eta_{1i}} \frac{d\phi_i}{d\eta_{2i}} \Bigg(\left(\frac{d}{d\phi_i}\frac{d\phi_i}{d\eta_{2i}}\right) \Bigg\{ \phi_i (y_i^* - \mu_i^*) \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)\right] \\
& + (y_i^* - \mu_i^*) - c_i \Bigg\} + \frac{d\phi_i}{d\eta_{2i}} \Bigg\{ (y_i^* - \mu_i^*) \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)\right] \\
& - c_i \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)\right] - \phi_i (y_i^* - \mu_i) r_i - \frac{c_i}{\phi_i} - \frac{\partial c_i}{\partial \phi_i} \Bigg\} \Bigg) x_{ir} z_{iR} z_{iS}, \\
\frac{\partial C_i^{(R,S)}}{\partial \delta_T} =& \Bigg(- \frac{\partial p_i}{\partial \phi_i} \left(\frac{d\phi_i}{d\eta_{2i}}\right)^3 + t_i \left\{ -3p_i + 2\left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)\right]^2 \right\} \\
& + \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)\right] v_i \Bigg) z_{iR} z_{iS} z_{iT}.
\end{aligned}$$

The above derivatives are used to construct the matrix $\nabla \boldsymbol{D}_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \nabla \boldsymbol{d}_i(\boldsymbol{\theta})$.

APPENDIX C – PROOFS OF THE VALIDITY OF TWO BETA REGRESSION INFORMATION MATRIX TESTS

Proof of Lemma 1

Proof. Let Y be a beta distributed random variable and $W_1 = \log(Y)$. It is easy to see, from the moment generating function of W_1, that $\mathbb{E}(\|W_1\|^r) < \infty$ for all integers $r > 0$. Since $1-Y$ is also beta distributed, the same holds for $W_2 = \log(1-Y)$. It follows that Y_i^* and $Y_i^\dagger$ have finite moments of all orders.

Let G be a given partial derivative of $\ell(\mu_i, \phi_i; Y_i)$ of any order with respect to a component of $\boldsymbol{\theta}$. Then, it follows from the expression for $\ell(\mu_i, \phi_i; Y_i)$ that G is a linear combination of Y_i^* and $Y_i^\dagger$. Thus, G has finite moments of all orders.

Also, the expression for $\mathbb{E}(G^r)$, for a given integer $r > 0$, involves products of regressors and a continuous function of (μ_i, ϕ_i). Thus, from Assumptions (A2) and (A3), there exists a positive constant $K_{G,r}$ such that $|\mathbb{E}(G^r)| < K_{G,r}$ for all i. $\square$

Proof of Lemma 2

Proof. Let $\boldsymbol{v}$ be a k-dimensional vector with $\|\boldsymbol{v}\| = 1$. Consider the sequence of univariate random variables $Z_{n,i} = \boldsymbol{v}^\top \boldsymbol{W}_{n,i}$. Then, $\gamma_{n,i} = \mathbb{E}(Z_{n,i}) = \boldsymbol{v}^\top \boldsymbol{\mu}_{n,i}$ and $\mathrm{Var}(Z_{n,i}) = \sigma_{n,i}^2 = \boldsymbol{v}^\top \Sigma_{n,i} \boldsymbol{v}$, with $\sigma_{n,i} > 0\ \forall n$ and $\forall i$, since $\Sigma_{n,i}$ is positive definite $\forall n$ and $\forall i$. From the Cauchy-Schwartz inequality, $\|Z_{n,i}\| \leq \|\boldsymbol{W}_{n,i}\|$, therefore, $\mathbb{E}(\|Z_{n,i}\|^{2+\delta}) \leq \mathbb{E}(\|\boldsymbol{W}_{n,i}\|^{2+\delta}) < \Delta\ \forall i$.

Define $\bar{Z}_n = n^{-1}\sum_{i=1}^n Z_{n,i} = \boldsymbol{v}^\top \overline{\boldsymbol{W}}_n$, $\bar{\gamma}_n = \mathbb{E}(\bar{Z}_n) = n^{-1}\sum_{i=1}^n \gamma_{n,i} = \boldsymbol{v}^\top \bar{\boldsymbol{\mu}}_n$ and $\bar{\sigma}_n^2 = \mathrm{Var}(\sqrt{n}\bar{Z}_n) = n^{-1}\sum_{i=1}^n \sigma_{n,i}^2$. Notice that

$$\bar{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \boldsymbol{v}^\top \Sigma_{n,i} \boldsymbol{v} = \boldsymbol{v}^\top \bar{\Sigma}_n \boldsymbol{v}.$$

Since $\bar{\Sigma}_n$ converges to a positive definite matrix V as $n \to \infty$, then, $\bar{\sigma}_n^2 = \boldsymbol{v}^\top \bar{\Sigma}_n \boldsymbol{v} \to \boldsymbol{v}^\top V \boldsymbol{v} > 0$. That means we can take $\delta' > 0$ such that $\bar{\sigma}_n^2 > \delta'$ for sufficiently large n.

From Theorem 5.11 in White (2001),

$$\sqrt{n}\frac{\bar{Z}_n - \bar{\gamma}_n}{\bar{\sigma}_n} \xrightarrow{D} \mathcal{N}(0,1).$$

Therefore,

$$\sqrt{n}\frac{\bar{Z}_n - \bar{\gamma}_n}{\sqrt{\boldsymbol{v}^\top V \boldsymbol{v}}} = \sqrt{n}\frac{\bar{Z}_n - \bar{\gamma}_n}{\bar{\sigma}_n}\sqrt{\frac{\boldsymbol{v}^\top \bar{\Sigma}_n \boldsymbol{v}}{\boldsymbol{v}^\top V \boldsymbol{v}}} \xrightarrow{D} \mathcal{N}(0,1).$$

Hence, $\sqrt{n}(\bar{Z}_n - \bar{\gamma}_n) \xrightarrow{D} \mathcal{N}(0, \boldsymbol{v}^\top V \boldsymbol{v})$, for all unit vectors $\boldsymbol{v}$. In summary, for all unit vectors $\boldsymbol{v}$,

$$\sqrt{n}\boldsymbol{v}^\top(\overline{\boldsymbol{W}}_n - \bar{\boldsymbol{\mu}}_n) = \sqrt{n}(\bar{Z}_n - \bar{\gamma}_n) \xrightarrow{D} \mathcal{N}(0, \boldsymbol{v}^\top V \boldsymbol{v}),$$

the limiting distribution being the distribution of $\boldsymbol{v}^\top \mathcal{N}_k(\mathbf{0}, V)$ From the Cramér-Wold theorem, we conclude that

$$\sqrt{n}(\overline{\boldsymbol{W}}_n - \bar{\boldsymbol{\mu}}_n) \xrightarrow{D} \mathcal{N}_k(\mathbf{0}, V),$$

as we wanted. □

Proof of Theorem 3

Proof. For the beta regression log-likelihood function ℓ, we have

$$\nabla\ell(\hat{\boldsymbol{\theta}}_n) = \nabla\ell(\boldsymbol{\theta}_0) + nA_n(\boldsymbol{\theta}_0; \boldsymbol{Y})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \frac{n}{2}\boldsymbol{r},$$

where $\boldsymbol{r}$, the remainder of the multivariate Taylor expansion, is a vector whose ith component is $(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top J_{i,n}(\boldsymbol{\theta}^\diamond_{i.n}; \boldsymbol{Y})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, $J_{i,n}$ being the Jacobian matrix of the ith row of $A_n(\boldsymbol{\theta}_0, \boldsymbol{Y})$ and $\boldsymbol{\theta}^\diamond_{i,n}$, for each i, is a vector such that $\|\boldsymbol{\theta}^\diamond_{i,n} - \boldsymbol{\theta}_0\| \leq \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|$. Since $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_P(n^{-1/2})$, whereas $J_{i,n}(\boldsymbol{\theta}^\diamond_i; \boldsymbol{Y}) = O_P(1)$, it follows that $\boldsymbol{r} = O_P(n^{-1})$. That is,

$$\nabla\ell(\hat{\boldsymbol{\theta}}_n) = \nabla\ell(\boldsymbol{\theta}_0) + nA_n(\boldsymbol{\theta}_0; \boldsymbol{Y})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + O_P(1).$$

The left hand side of the above equation is, of course, equal to zero. Thus,

$$\mathbf{0} = \nabla\ell(\boldsymbol{\theta}_0) + nA_n(\boldsymbol{\theta}_0; \boldsymbol{Y})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + O_P(1).$$

That is, $\nabla\ell(\boldsymbol{\theta}_0) + nA_n(\boldsymbol{\theta}_0; \boldsymbol{Y})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_P(1)$. Therefore,

$$n^{-1/2}\nabla\ell(\boldsymbol{\theta}_0) + n^{1/2}A_n(\boldsymbol{\theta}_0; \boldsymbol{Y})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_P(n^{-1/2}) = o_P(1),$$

i.e., $n^{-1/2}\nabla\ell(\boldsymbol{\theta}_0) + n^{1/2}A_n(\boldsymbol{\theta}_0; \boldsymbol{Y})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{P} \mathbf{0}$. This can be rewritten as

$$n^{-1/2}\nabla\ell(\boldsymbol{\theta}_0) + n^{1/2}\Big(A_n(\boldsymbol{\theta}_0; \boldsymbol{Y}) - n^{-1}\sum_{i=1}^n \dot{A}_i(\boldsymbol{\theta}_0)\Big)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$
$$+ n^{-1/2}\sum_{i=1}^n \dot{A}_i(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{P} \mathbf{0}.$$

From Lemma 1, $A_n(\boldsymbol{\theta}_0; \boldsymbol{Y})$ is an average of independent random variables with uniformly bounded variances. Then, it follows from Kolmogorov's first strong law of large numbers that $A_n(\boldsymbol{\theta}_0;\boldsymbol{Y}) - n^{-1}\sum_{i=1}^n \dot{A}_i(\boldsymbol{\theta}_0)$ converges almost surely to zero. The second term is, therefore, $o_P(1)$. Hence,

$$n^{-1/2}\nabla\ell(\boldsymbol{\theta}_0) + n^{-1/2}\sum_{i=1}^n \dot{A}_i(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{P} \mathbf{0}.$$

That is,

$$n^{-1/2}\nabla\ell(\boldsymbol{\theta}_0) + n^{1/2}\Big(n^{-1}\sum_{i=1}^n \dot{A}_i(\boldsymbol{\theta}_0) - A(\boldsymbol{\theta}_0)\Big)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$
$$+ n^{1/2}A(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{P} \mathbf{0}.$$

Assumptions (A1) and (A4) imply that the second term in the above expression is $o_P(1)$. We arrive at $n^{-1/2}\nabla\ell(\boldsymbol{\theta}_0) + n^{1/2}A(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{P} \mathbf{0}$. From this result, we immediately obtain

$$n^{-1/2}\nabla\boldsymbol{D}(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)^{-1}\nabla\ell(\boldsymbol{\theta}_0) + n^{1/2}\nabla\boldsymbol{D}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{P} \mathbf{0}. \qquad \text{(C.1)}$$

We now consider the vector $\boldsymbol{D}_n$. A first order Taylor series expansion yields

$$\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{D}_n(\boldsymbol{\theta}_0) + \nabla\boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \qquad \text{(C.2)}$$

where $\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \le \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|$. Notice that $\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_P(1)$, since $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_P(1)$.

Consider now the matrix $\nabla\boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n)$. Let $(M)_{.j}$ denote the jth column of a general matrix M. We consider a Taylor expansion of $(\nabla\boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n))_{.j}$ in the form

$$(\nabla\boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n))_{.j} = (\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0))_{.j} + K_{j,n}(\boldsymbol{\theta}^{\diamond}_{j,n};\boldsymbol{Y})(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

where $K_{j,n}$ is the Jacobian matrix of $(\nabla\boldsymbol{D}_n)_{.j}$ and $\|\boldsymbol{\theta}^{\diamond}_{j,n} - \boldsymbol{\theta}_0\| \le \|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|$. Since $K_{j,n} = O_P(1)$ and $\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_P(1)$, we conclude that $(\nabla\boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n))_{.j} - (\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0))_{.j} = o_P(1)$, for each j. That is, $\nabla\boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n) - \nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0) = o_P(1)$. Now,

$$\nabla\boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n) = \nabla\boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n) - \nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0) + \nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0) - \mathbb{E}(\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0)) + \mathbb{E}(\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0)).$$

We have seen that $\nabla\boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n) - \nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0) = o_P(1)$. Also, from Lemma 1, all entries of $\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0)$ are averages of independent random variables with uniformly bounded variances. We then conclude, from Kolgororov's first law of large numbers, that $\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0) - \mathbb{E}(\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0))$ converges to a zero matrix almost surely. Finally, Assumption (A5) states that $\mathbb{E}(\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0))$ converges to $\nabla\boldsymbol{D}(\boldsymbol{\theta}_0)$. We arrive at $\nabla\boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n) \xrightarrow{P} \nabla\boldsymbol{D}(\boldsymbol{\theta}_0)$. Since $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_P(n^{-1/2})$, we easily obtain

$$\sqrt{n}(\nabla\boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n) - \nabla\boldsymbol{D}(\boldsymbol{\theta}_0))(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{P} \mathbf{0}.$$

We combine the result above with (C.1) to obtain

$$n^{-1/2}\nabla \boldsymbol{D}(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)^{-1}\nabla \ell(\boldsymbol{\theta}_0) + n^{1/2}\nabla \boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{P} \mathbf{0}. \tag{C.3}$$

We will now turn to the quantities $\boldsymbol{d}_i(\boldsymbol{\theta}_0) - \nabla \boldsymbol{D}(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)^{-1}\nabla \ell(\boldsymbol{\theta}_0; Y_i)$. Under the null hypothesis, they have zero mean vector. Also, from Lemma 1, they have finite and uniformly bounded moments of all orders. From this fact and Assumption (A6), we conclude that all conditions of Lemma 2 are satisfied. We obtain

$$\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0) - n^{-1/2}\nabla \boldsymbol{D}(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)^{-1}\nabla \ell(\boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}_K(\mathbf{0}, V(\boldsymbol{\theta}_0)).$$

By combining this result with (C.3), we arrive at

$$\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0) + \sqrt{n}\nabla \boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}_K(\mathbf{0}, V(\boldsymbol{\theta}_0)).$$

We can now return to Equation (C.2). We have

$$\sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n) = \sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0) + \sqrt{n}\nabla \boldsymbol{D}_n(\tilde{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).$$

Thus, we proved that, under the null hypothesis,

$$\sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{D} \mathcal{N}_K(\mathbf{0}, V(\boldsymbol{\theta}_0)).$$

□

Proof of Theorem 4

Proof. We begin by considering a first order Taylor expansion of the jth column of V_{n1} in the form

$$(V_{n1}(\hat{\boldsymbol{\theta}}_n))_{.j} = (V_{n1}(\boldsymbol{\theta}_0))_{.j} + W_{j,n}(\boldsymbol{\theta}^{\diamond}_{j,n}; \boldsymbol{Y})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

where $W_{j,n}$ is the Jacobian matrix of $(V_{n1})_{.j}$ and $\|\boldsymbol{\theta}^{\diamond}_{j,n} - \boldsymbol{\theta}_0\| \leq \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|$. Since $W_{j,n} = O_P(1)$ and $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_P(1)$, then, $W_{j,n}(\boldsymbol{\theta}^{\diamond}_{j,n}; \boldsymbol{Y})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = o_P(1)$. Given that this holds for all j, we conclude that $V_{n1}(\hat{\boldsymbol{\theta}}_n) - V_{n1}(\boldsymbol{\theta}_0) \xrightarrow{P} O_{K\times K}$.

If suffices, then, to show that $V_{n1}(\boldsymbol{\theta}_0) \xrightarrow{P} V(\boldsymbol{\theta}_0)$. We will show that this convergence is, in fact, almost sure, under our assumptions. We begin by noticing that

$$\begin{aligned} V_{n1}(\boldsymbol{\theta}_0) =& \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{d}_i(\boldsymbol{\theta}_0)\nabla\ell(\boldsymbol{\theta}_0;Y_i)^\top A_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1} \\ &\times \nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top) - \frac{1}{n}\sum_{i=1}^{n}\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0;\boldsymbol{Y})A_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}\nabla\ell(\boldsymbol{\theta}_0;Y_i)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top \\ &+ \frac{1}{n}\sum_{i=1}^{n}(\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0;\boldsymbol{Y})A_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}\nabla\ell(\boldsymbol{\theta}_0;Y_i) \\ &\times \nabla\ell(\boldsymbol{\theta}_0;Y_i)^\top A_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top). \end{aligned}$$

We will consider below each one of the four terms in the above expression separately.

1 We write

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top =& \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top - \mathbb{E}(\boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top)) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top). \end{aligned}$$

From Lemma 1, the first term in the right hand side of the above expression is an average of independent random matrices with entries having uniformly bounded variances and zero mean. Then, it follows from Kolmogorov's first strong law of large numbers that it converges almost surely to a zero matrix. Assumption (A7) guarantees that the second term in the right hand side of the expression converges to $\Phi(\boldsymbol{\theta}_0)$. We thus conclude that left-hand side of the above expression converges almost surely to $\Phi(\boldsymbol{\theta}_0)$.

2 As for the second term in the expression of $V_{n1}(\boldsymbol{\theta}_0)$, we have

$$\begin{aligned} &\Big(n^{-1}\sum_{i=1}^{n}\boldsymbol{d}_i(\boldsymbol{\theta}_0)\nabla\ell(\boldsymbol{\theta}_0;Y_i)^\top\Big)A_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top = \\ &\Big(n^{-1}\sum_{i=1}^{n}(\boldsymbol{d}_i(\boldsymbol{\theta}_0)\nabla\ell(\boldsymbol{\theta}_0;Y_i)^\top - \mathbb{E}(\boldsymbol{d}_i(\boldsymbol{\theta}_0)\nabla\ell(\boldsymbol{\theta}_0;Y_i)^\top))\Big) \\ &\times A_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top + \Big(n^{-1}\sum_{i=1}^{n}\mathbb{E}(\boldsymbol{d}_i(\boldsymbol{\theta}_0)\nabla\ell(\boldsymbol{\theta}_0;Y_i)^\top)\Big) \\ &\times A_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}\nabla\boldsymbol{D}_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top. \end{aligned}$$

From Lemma 1 and Kolmogorov's first strong law of large numbers, the first term in the above sum converges almost surely to a zero matrix. Also, $A_n(\boldsymbol{\theta}_0;\boldsymbol{Y})$ converges almost surely to $A(\boldsymbol{\theta}_0)$ and $\nabla \boldsymbol{D}_n(\boldsymbol{\theta}_0; \boldsymbol{Y})$ converges almost surely to $\nabla \boldsymbol{D}(\boldsymbol{\theta}_0)$. From Lancaster (1984) and Assumption (A5), the average in the second term converges to $-\nabla \boldsymbol{D}(\boldsymbol{\theta}_0)$. It then follows that the second term in the expression of $V_{n1}(\boldsymbol{\theta}_0)$ converges almost surely to $\nabla \boldsymbol{D}(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0)^{-1} \nabla \boldsymbol{D}(\boldsymbol{\theta}_0)^\top$.

3 It is immediate, then, that the third term in the expression of $V_{n1}(\boldsymbol{\theta}_0)$ converges almost surely to $\nabla \boldsymbol{D}(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0)^{-1} \nabla \boldsymbol{D}(\boldsymbol{\theta}_0)^\top$.

4 The fourth term in the expression of $V_{n1}(\boldsymbol{\theta}_0)$ is

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \nabla \boldsymbol{D}_n(\boldsymbol{\theta}_0; \boldsymbol{Y}) A_n(\boldsymbol{\theta}_0; \boldsymbol{Y})^{-1} \nabla \ell(\boldsymbol{\theta}_0; Y_i) \nabla \ell(\boldsymbol{\theta}_0; Y_i)^\top \\
&A_n(\boldsymbol{\theta}_0; \boldsymbol{Y})^{-1} \nabla \boldsymbol{D}_n(\boldsymbol{\theta}_0; \boldsymbol{Y})^\top = \nabla \boldsymbol{D}_n(\boldsymbol{\theta}_0; \boldsymbol{Y}) A_n(\boldsymbol{\theta}_0; \boldsymbol{Y})^{-1} \\
&\times \left(n^{-1} \sum_{i=1}^{n} \nabla \ell(\boldsymbol{\theta}_0; Y_i) \nabla \ell(\boldsymbol{\theta}_0; Y_i)^\top \right) A_n(\boldsymbol{\theta}_0; \boldsymbol{Y})^{-1} \nabla \boldsymbol{D}_n(\boldsymbol{\theta}_0; \boldsymbol{Y})^\top \\
&= \nabla \boldsymbol{D}_n(\boldsymbol{\theta}_0; \boldsymbol{Y}) A_n(\boldsymbol{\theta}_0; \boldsymbol{Y})^{-1} B_n(\boldsymbol{\theta}_0; \boldsymbol{Y}) A_n(\boldsymbol{\theta}_0; \boldsymbol{Y})^{-1} \nabla \boldsymbol{D}_n(\boldsymbol{\theta}_0; \boldsymbol{Y})^\top.
\end{aligned}
$$

Under the null hypothesis, $\mathbb{E}(A_i(\boldsymbol{\theta}_0) + B_i(\boldsymbol{\theta}_0)) = O_{k\times k} \forall i$. We then conclude, from Kolmogorov's first strong law of large numbers and Assumption (A4) that $B_n(\boldsymbol{\theta}_0; \boldsymbol{Y})$ converges almost surely to $-A(\boldsymbol{\theta}_0)$. The fourth term in the expression of $V_{n1}(\boldsymbol{\theta}_0)$ thus converges almost surely to

$$
-\nabla \boldsymbol{D}(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0)^{-1} \nabla \boldsymbol{D}(\boldsymbol{\theta}_0)^\top.
$$

We conclude that $V_{n1}(\boldsymbol{\theta}_0)$ converges to $\Phi(\boldsymbol{\theta}_0) + \nabla \boldsymbol{D}(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0)^{-1} \nabla \boldsymbol{D}(\boldsymbol{\theta}_0)^\top$ almost surely. We thus proved that $V_{n1}(\hat{\boldsymbol{\theta}}_n) \xrightarrow{P} \Phi(\boldsymbol{\theta}_0) + \nabla \boldsymbol{D}(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0)^{-1} \nabla \boldsymbol{D}(\boldsymbol{\theta}_0)^\top$.

We will now turn to the expression of $V(\boldsymbol{\theta}_0)$. Under the null hypothesis, it is the limit when $n \to \infty$ of

$$
\begin{aligned}
&n^{-1} \sum_{i=1}^{n} \Big[\mathbb{E}\Big(\boldsymbol{d}_i(\boldsymbol{\theta}_0) - \nabla \boldsymbol{D}(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0)^{-1} \nabla \ell(\boldsymbol{\theta}_0; Y_i) \Big) \\
&\times \Big(\boldsymbol{d}_i(\boldsymbol{\theta}_0) - \nabla \boldsymbol{D}(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0)^{-1} \nabla \ell(\boldsymbol{\theta}_0; Y_i) \Big)^\top \Big] \\
&= n^{-1} \sum_{i=1}^{n} \mathbb{E}(\boldsymbol{d}_i(\boldsymbol{\theta}_0) \boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top) - \nabla \boldsymbol{D}(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0)^{-1} \\
&\times \left(n^{-1} \sum_{i=1}^{n} \mathbb{E}(\nabla \ell(\boldsymbol{\theta}_0; Y_i) \boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top) \right) - \left(n^{-1} \sum_{i=1}^{n} \mathbb{E}(\boldsymbol{d}_i(\boldsymbol{\theta}_0) \nabla \ell(\boldsymbol{\theta}_0; Y_i)^\top) \right) \\
&\times A(\boldsymbol{\theta}_0)^{-1} \nabla \boldsymbol{D}(\boldsymbol{\theta}_0)^\top + \nabla \boldsymbol{D}(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0)^{-1} \left(n^{-1} \sum_{i=1}^{n} \mathbb{E}(\nabla \ell(\boldsymbol{\theta}_0; Y_i) \nabla \ell(\boldsymbol{\theta}_0; Y_i)^\top) \right) \\
&\times A(\boldsymbol{\theta}_0)^{-1} \nabla \boldsymbol{D}(\boldsymbol{\theta}_0)^\top.
\end{aligned}
$$

From what we have seen before, the limit of the above expression is

$$\Phi(\boldsymbol{\theta}_0)+\nabla\boldsymbol{D}(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)^{-1}\nabla\boldsymbol{D}(\boldsymbol{\theta}_0)^\top$$

and, hence, this is the expression for $V(\boldsymbol{\theta}_0)$. We thus arrive then at the desired result. □

Proof of Theorem 5

Proof. Similarly to the proof of the former theorem, we conclude from a first order Taylor expansion that $V_{n2}(\hat{\boldsymbol{\theta}}_n)-V_{n2}(\boldsymbol{\theta}_0)\xrightarrow{P}O_{K\times K}$.

We will complete the proof by showing that $V_{n2}(\boldsymbol{\theta}_0)\xrightarrow{P}V(\boldsymbol{\theta}_0)$. We have

$$\begin{aligned}
&V_{n2}(\boldsymbol{\theta}_0)=n^{-1}\sum_{i=1}^{n}\boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top+n^{-1}\sum_{i=1}^{n}(\boldsymbol{d}_i(\boldsymbol{\theta}_0)\nabla\ell(\boldsymbol{\theta}_0;Y_i)^\top\\
&\times B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top)+n^{-1}\sum_{i=1}^{n}(L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}\\
&\times\nabla\ell(\boldsymbol{\theta}_0;Y_i)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top)+n^{-1}\sum_{i=1}^{n}(L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}\nabla\ell(\boldsymbol{\theta}_0;Y_i)\\
&\times\nabla\ell(\boldsymbol{\theta}_0;Y_i)^\top B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top)=n^{-1}\sum_{i=1}^{n}\boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top\\
&+\left(n^{-1}\sum_{i=1}^{n}\boldsymbol{d}_i(\boldsymbol{\theta}_0)\nabla\ell(\boldsymbol{\theta}_0;Y_i)^\top\right)B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top\\
&+L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}\left(n^{-1}\sum_{i=1}^{n}\nabla\ell(\boldsymbol{\theta}_0;Y_i)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top\right)\\
&+L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}\left(n^{-1}\sum_{i=1}^{n}\nabla\ell(\boldsymbol{\theta}_0;Y_i)\nabla\ell(\boldsymbol{\theta}_0;Y_i)^\top\right)\\
&\times B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top=n^{-1}\sum_{i=1}^{n}\boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top-(L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})\\
&\times B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top)-L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top\\
&+L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top\\
&=n^{-1}\sum_{i=1}^{n}\boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)^\top-L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^{-1}L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})^\top.
\end{aligned}$$

It is now clear from the proof of the previous theorem that, under the null hypothesis, the first term in the last sum above converges almost surely to $\Phi(\boldsymbol{\theta}_0)$, $L_n(\boldsymbol{\theta}_0;\boldsymbol{Y})$ converges almost surely to $-\nabla\boldsymbol{D}(\boldsymbol{\theta}_0)$ and $B_n(\boldsymbol{\theta}_0;\boldsymbol{Y})$ converges almost surely to $-A(\boldsymbol{\theta}_0)$. The last expression above thus converges almost surely to

$$\Phi(\boldsymbol{\theta}_0)+\nabla\boldsymbol{D}(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)^{-1}\nabla\boldsymbol{D}(\boldsymbol{\theta}_0)^\top$$

which is the expression for $V(\boldsymbol{\theta}_0)$. That completes the proof. □

APPENDIX D – SAMPLE MATRICES FOR ATYPICAL CASES DETECTION IN GAUSSIAN LINEAR REGRESSIONS

In what follows we will obtain expressions for $A_n(\boldsymbol{\theta};\boldsymbol{y})$ and $B_n(\boldsymbol{\theta};\boldsymbol{y})$ in the Gaussian linear regression model with multiplicative heteroskedasticity. Let $\mu_i = \boldsymbol{x}_i^\top\boldsymbol{\beta}$ and $\phi_i = \exp(\boldsymbol{z}_i^\top\boldsymbol{\delta})$, where $\boldsymbol{x}_i = (x_{i1},\ldots,x_{ip})^\top$ and $\boldsymbol{z}_i = (z_{i1},\ldots,z_{iq})^\top$. Also, let $X = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_n]^\top$ and $Z = [\boldsymbol{z}_1 \cdots \boldsymbol{z}_n]^\top$ be the matrices of mean and dispersion regressors, respectively. The i-th log-likelihood function for $Y_1,\ldots,Y_n$ with observed values $y_1,\ldots,y_n$

$$\ell_i(\boldsymbol{\theta};y_i) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\boldsymbol{z}_i^\top\boldsymbol{\delta} - \frac{1}{2}\exp(-\boldsymbol{z}_i^\top\boldsymbol{\delta})(y_i - \boldsymbol{x}_i^\top\boldsymbol{\delta})^2.$$

The derivatives of $\ell_i(\boldsymbol{\theta};y_i)$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ is

$$\begin{aligned}\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}} &= \exp(-\boldsymbol{z}_i^\top\boldsymbol{\delta})(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})\boldsymbol{x}_i,\\ \frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\delta}} &= \frac{1}{2}\boldsymbol{z}_i\Big[\exp(-\boldsymbol{z}_i^\top\boldsymbol{\delta})(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2 - 1\Big].\end{aligned}$$

The blocks of $B_i(\boldsymbol{\theta};y_i) = \partial\ell_i(\boldsymbol{\theta};y_i)/\partial\boldsymbol{\theta} \times \partial\ell_i(\boldsymbol{\theta};y_i)/\partial\boldsymbol{\theta}^\top$ can be written as

$$\begin{aligned}\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}}\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}^\top} &= \exp(-2\boldsymbol{z}_i^\top\boldsymbol{\delta})(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2\boldsymbol{x}_i\boldsymbol{x}_i^\top,\\ \frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}}\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\delta}^\top} &= \frac{1}{2}\exp(-\boldsymbol{z}_i^\top\boldsymbol{\delta})(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})\Big[\exp(-\boldsymbol{z}_i^\top\boldsymbol{\delta})(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2\\ &\quad - 1\Big]\boldsymbol{x}_i\boldsymbol{z}_i^\top,\\ \frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\delta}}\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\delta}^\top} &= \frac{1}{4}\Big[\exp(-\boldsymbol{z}_i^\top\boldsymbol{\delta})(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2 - 1\Big]^2\boldsymbol{z}_i\boldsymbol{z}_i^\top.\end{aligned}$$

Similarly, the blocks of $A_i(\boldsymbol{\theta};y_i) = \partial^2\ell_i(\boldsymbol{\theta};y_i)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ can be expressed as

$$\begin{aligned}\frac{\partial^2\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -\exp(-\boldsymbol{z}_i^\top\boldsymbol{\delta})\boldsymbol{x}_i\boldsymbol{x}_i^\top,\\ \frac{\partial^2\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\delta}^\top} &= -\exp(-\boldsymbol{z}_i^\top\boldsymbol{\delta})(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})\boldsymbol{x}_i\boldsymbol{z}_i^\top,\\ \frac{\partial^2\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}^\top} &= -\frac{1}{2}\exp(-\boldsymbol{z}_i^\top\boldsymbol{\delta})(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2\boldsymbol{z}_i\boldsymbol{z}_i^\top.\end{aligned}$$

We will now obtain expressions for $A_n(\boldsymbol{\theta};\boldsymbol{y})$ and $B_n(\boldsymbol{\theta};\boldsymbol{y})$ using the first- and second-order derivatives of the total log-likelihood function. Let $w_i = \exp(-\boldsymbol{z}_i^\top\boldsymbol{\delta})$, $e_i = y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}$, $W = \operatorname{diag}(w_1,\ldots,w_n)$ and $E = \operatorname{diag}(e_1,\ldots,e_n)$. Additionally, let $\boldsymbol{u} = (1,\ldots,1)^\top$ be the n-dimensional vector of ones. We can now write the total log-likelihood first-order derivatives as

$$\frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}} = X^\top W(\boldsymbol{y} - X\boldsymbol{\beta}) \quad \text{and} \quad \frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\delta}} = \frac{1}{2}Z^\top(WE^2 - I_n)\boldsymbol{u}.$$

It is possible to write $B_n(\boldsymbol{\theta};\boldsymbol{y}) = n^{-1}\sum_{i=1}^{n} B_i(\boldsymbol{\theta};y_i)$ as

$$B_n(\boldsymbol{\theta};\boldsymbol{y}) = \frac{1}{n}\begin{bmatrix} X^\top W^2 E^2 X & \frac{1}{2}X^\top WE(WE^2 - I_n)Z \\ \\ \frac{1}{2}Z^\top WE(WE^2 - I_n)X & \frac{1}{4}Z^\top (WE^2 - I_n)^2 Z \end{bmatrix}.$$

It follows that $B_n(\boldsymbol{\theta};\boldsymbol{y}) = n^{-1}M^\top M$, where M is the matrix of order $n \times (p+q)$ given by $M = \begin{bmatrix} WEX & \frac{1}{2}(WE^2 - I_n)Z \end{bmatrix}$.

Finally, $A_n(\boldsymbol{\theta};\boldsymbol{y}) = n^{-1}\sum_{i=1}^{n} A_i(\boldsymbol{\theta};y_i)$ can be expressed as

$$A_n(\boldsymbol{\theta};\boldsymbol{y}) = -\frac{1}{n}\begin{bmatrix} X^\top WX & X^\top WEZ \\ \\ Z^\top WEX & \frac{1}{2}Z^\top WE^2 Z \end{bmatrix}.$$

APPENDIX E – SAMPLE MATRICES FOR ATYPICAL CASES DETECTION IN BETA REGRESSIONS

We will now provide simple expressions for $A_n(\boldsymbol{\theta};\boldsymbol{y})$ and $B_n(\boldsymbol{\theta};\boldsymbol{y})$ in varying precision beta regressions. Let $g_1(\mu_i)=\boldsymbol{x}_i^\top\boldsymbol{\beta}=\eta_{1i}$ and $g_2(\phi_i)=\boldsymbol{z}_i^\top\boldsymbol{\delta}=\eta_{2i}$, where $\boldsymbol{x}_i$, $\boldsymbol{z}_i$, X and Z are as in the previous appendix. The i-th log-likelihood function for $Y_1,\dots,Y_n$ with observed values $y_1,\dots,y_n$ is

$$\ell_i(\boldsymbol{\theta};y_i)=\log\Gamma(\phi_i)-\log\Gamma(\mu_i\phi_i)-\log\Gamma((1-\mu_i)\phi_i)+(\mu_i\phi_i-1)y_i^*+(\phi_i-2)y_i^\dagger,$$

where $y_i^*=\log(y_i/(1-y_i))$ and $y_i^\dagger=\log(1-y_i)$. Let $Y_i^*=\log(Y_i/(1-Y_i))$ and $Y_i^\dagger=\log(1-Y_i)$. It can be easily verified that $\mathbb{E}(Y_i^*)=\psi(\mu_i\phi_i)-\psi((1-\mu_i)\phi_i)$, $\mathbb{E}(Y_i^\dagger)=\psi((1-\mu_i)\phi_i)-\psi(\phi_i)$, where ψ is the digamma function. We will denote $\mathbb{E}(Y_i^*)$ and $\mathbb{E}(Y_i^\dagger)$ by μ_i^* and $\mu_i^\dagger$, respectively. The derivatives of $\ell_i(\boldsymbol{\theta};y_i)$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are

$$\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}}=\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\mu_i}\frac{d\mu_i}{d\eta_{1i}}\frac{\partial\eta_{1i}}{\partial\boldsymbol{\beta}}=\phi_i(y_i^*-\mu_i^*)\frac{1}{g_1'(\mu_i)}\boldsymbol{x}_i,$$
$$\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\delta}}=\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\phi_i}\frac{d\phi_i}{d\eta_{2i}}\frac{\partial\eta_{2i}}{\partial\boldsymbol{\delta}}=[\mu_i(y_i^*-\mu_i^*)+(y_i^\dagger-\mu_i^\dagger)]\frac{1}{g_2'(\phi_i)}\boldsymbol{z}_i.$$

Let $\dot{t}_i=1/g_1'(\mu_i)$ and $\dot{h}_i=1/g_2'(\phi_i)$. The blocks of $B_i(\boldsymbol{\theta};y_i)$ are

$$\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}}\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}^\top}=-\phi_i^2(y_i^*-\mu_i)^2\dot{t}_i^2\boldsymbol{x}_i\boldsymbol{x}_i^\top,$$
$$\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}}\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\delta}^\top}=\phi_i(y_i^*-\mu_i^*)[\mu_i(y_i^*-\mu_i^*)+(y_i^\dagger-\mu_i^\dagger)]\dot{t}_i\dot{h}_i\boldsymbol{x}_i\boldsymbol{z}_i^\top,$$
$$\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\delta}}\frac{\partial\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\delta}^\top}=[\mu_i(y_i^*-\mu_i^*)+(y_i^\dagger-\mu_i^\dagger)]^2\dot{h}_i^2\boldsymbol{z}_i\boldsymbol{z}_i^\top.$$

Similarly, the blocks of $A_i(\boldsymbol{\theta};y_i)$ are

$$\frac{\partial^2\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}=-\phi_i\dot{q}_i\boldsymbol{x}_i\boldsymbol{x}_i^\top,\ \frac{\partial^2\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\delta}^\top}=-\dot{f}_i\dot{t}_i\dot{h}_i\boldsymbol{x}_i\boldsymbol{z}_i^\top,\ \frac{\partial^2\ell_i(\boldsymbol{\theta};y_i)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}^\top}=-\dot{\nu}_i\boldsymbol{z}_i\boldsymbol{z}_i^\top,$$

where the quantities $\dot{q}_i$, $\dot{f}_i$ and $\dot{\nu}_i$ are the corresponding undotted quantities in Appendix B of Ferrari, Espinheira and Cribari-Neto (2011) .

We will now express $A_n(\boldsymbol{\theta};\boldsymbol{y})$ and $B_n(\boldsymbol{\theta};\boldsymbol{y})$ using the first- and second-order derivatives of the total log-likelihood function. Let $\dot{e}_i=y_i^*-\mu_i^*$, $\dot{w}_i=\mu_i(y_i^*-\mu_i^*)+(y_i^\dagger-\mu_i^\dagger)$, $\dot{\boldsymbol{w}}=(w_1,\dots,w_n)^\top$, $\dot{E}=\operatorname{diag}(\dot{e}_1,\dots,\dot{e}_n)$ and $\dot{W}=\operatorname{diag}(\dot{w}_1,\dots,\dot{w}_n)$. Also, let $\dot{T}=\operatorname{diag}(\dot{t}_1,\dots,\dot{t}_n)$, $\dot{H}=\operatorname{diag}(\dot{h}_1,\dots,\dot{h}_n)$, $\dot{Q}=\operatorname{diag}(\dot{q}_1,\dots,\dot{q}_n)$, $\dot{F}=\operatorname{diag}(\dot{f}_1,\dots,\dot{f}_n)$, $\dot{\mathcal{V}}=\operatorname{diag}(\dot{\nu}_1,\dots,\dot{\nu}_n)$, $\Phi=\operatorname{diag}(\phi_1,\dots,\phi_n)$, and $\boldsymbol{\mu}^*=(\mu_1^*,\dots,\mu_n^*)^\top$. We can now write the total log-likelihood first-order derivatives as

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = X^\top \Phi \dot{T}(\boldsymbol{y} - \boldsymbol{\mu}^*) \quad \text{and} \quad \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\delta}} = Z^\top \dot{H} \dot{\boldsymbol{w}}.$$

It is possible to write $B_n(\boldsymbol{\theta};\boldsymbol{y}) = n^{-1}\sum_{i=1}^n B_i(\boldsymbol{\theta};y_i)$ as

$$B_n(\boldsymbol{\theta};\boldsymbol{y}) = \frac{1}{n}\begin{bmatrix} X^\top \Phi^2 \dot{T}^2 \dot{E}^2 X & X^\top \Phi \dot{T} \dot{E} \dot{W} \dot{H} Z \\ \\ Z^\top \Phi \dot{T} \dot{E} \dot{W} \dot{H} X & Z^\top \dot{H}^2 \dot{W}^2 Z \end{bmatrix}.$$

As in Appendix D, we can write the matrix $B_n(\boldsymbol{\theta};\boldsymbol{y})$ as $B_n(\boldsymbol{\theta};\boldsymbol{y}) = n^{-1}\dot{M}^\top \dot{M}$, where $\dot{M} = \begin{bmatrix} X^\top \Phi \dot{T} \dot{E} & Z^\top \dot{H} \dot{W} \end{bmatrix}$, a matrix of order $n \times (p+q)$. Finally, the matrix $A_n(\boldsymbol{\theta};\boldsymbol{y}) = n^{-1}\sum_{i=1}^n A_i(\boldsymbol{\theta};y_i)$ can be expressed as

$$A_n(\boldsymbol{\theta};\boldsymbol{y}) = -\frac{1}{n}\begin{bmatrix} X^\top \Phi \dot{Q} X & X^\top \dot{F} \dot{T} \dot{H} Z \\ \\ Z^\top \dot{F} \dot{T} \dot{H} X & Z^\top \dot{\mathcal{V}} Z \end{bmatrix}.$$