TEST (2015) 24:776–795
DOI 10.1007/s11749-015-0434-6

ORIGINAL PAPER

Bootstrap-based model selection criteria for beta regressions

Fábio M. Bayer[1] · **Francisco Cribari-Neto**[2]

Received: 2 June 2014 / Accepted: 3 March 2015 / Published online: 15 March 2015

Abstract This paper addresses the issue of model selection in the beta regression model focused on small samples. The Akaike information criterion (AIC) is a model selection criterion widely used in practical applications. The AIC is an estimator of the expected log-likelihood value, and measures the discrepancy between the true model and the estimated model. In small samples, the AIC is biased and tends to select overparameterized models. To circumvent that problem, we propose two new selection criteria, namely: the bootstrapped likelihood quasi-CV and its 632QCV variant. We use Monte Carlo simulation to compare the finite sample performances of the two proposed criteria to those of the AIC and its variations that use the bootstrapped log-likelihood in the class of varying dispersion beta regressions. The numerical evidence shows that the proposed model selection criteria perform well in small samples. We also present and discuss and empirical application.

Keywords AIC · Beta regression · Bootstrap · Cross-validation · Model selection · Varying dispersion

Mathematics Subject Classification 62J99 · 62F07 · 62F40 · 94A17

✉ Fábio M. Bayer
bayer@ufsm.br

[1] Departamento de Estatística and LACESM, Universidade Federal de Santa Maria, Santa Maria, Brazil

[2] Departamento de Estatística, Universidade Federal de Pernambuco, Recife, Brazil

1 Introduction

In regression analysis, practitioners are usually interested in selecting the model that yields the best fit from a broad class of candidate models. Thus, model selection is of paramount importance in regression analysis. Model selection is usually based on model selection criteria or information criteria. The Akaike information criterion (AIC) (Akaike 1973) is the most well-known and commonly used model selection criterion. Several alternative criteria have been developed in the literature, such as the SIC (Schwarz 1978), HQ (Hannan and Quinn 1979) and AICc (Hurvich and Tsai 1989).

The AIC was proposed for estimating (minus two times) the expected log-likelihood. Using Taylor series expansion and the asymptotic normality of the maximum likelihood estimator Akaike showed that the maximized log-likelihood function is a positively biased estimator for the expected log-likelihood. After computing such bias, the author derived the AIC as an asymptotically approximated correction for the expected log-likelihood. In small samples, however, the AIC is biased and tends to select models that are overparameterized (Hurvich and Tsai 1989).

Several variants of the AIC have been proposed in the literature. The first correction of the AIC, the AICc, was proposed in Sugiura (1978) for linear regression models. Later, Hurvich and Tsai (1989) expanded the applicability of the AICc to cover nonlinear regression and autoregressive models. They showed that the AICc is asymptotically equivalent to the AIC, but usually delivers more accurate model selection in finite samples. Analytical corrections to the AIC, such as AICc, can be nonetheless difficult to obtain in some classes of models (Shibata 1997). The analytical difficulties stem from distributional and asymptotic results, as well as from certain restrictive assumptions. To circumvent analytical difficulties and to obtain more accurate corrections in small samples, bootstrap (Efron 1979) variants of the AIC were considered in the literature. They have been introduced and explored in different classes of models. See, for instance, Cavanaugh and Shumway (1997), Ishiguro and Sakamoto (1991), Ishiguro et al. (1997), Seghouane (2010), Shang and Cavanaugh (2008) and Shibata (1997), who introduced the criteria known as WIC, AICb, EIC, among other denominations. Such bootstrap extensions typically outperform the AIC in finite samples. In addition, as noted by Shibata (1997), they can be easily computed.

Both the AIC and its bootstrap variants aim at estimating the expected log-likelihood using a bias correction for the maximized log-likelihood. In this paper, we follow the approach introduced by Pan (1999) and propose an estimator for the expected log-likelihood that does not require a bias adjustment term. In particular, nonparametric bootstrap and cross-validation (CV) are jointly used in a criterion called bootstrapped likelihood CV (BCV). Using the parametric bootstrap and a quasi-CV method, we define a new AIC variant. It uses the bootstrapped likelihood quasi-CV (BQCV). We also propose a slight modification known as 632QCV.

Model selection criteria based on the bootstrapped log-likelihood have been explored and successfully applied to autoregressive models (Ishiguro et al. 1997), state-space models (Bengtsson and Cavanaugh 2006; Cavanaugh and Shumway 1997), mixed models (Shang and Cavanaugh 2008), linear regression models (Pan 1999; Seghouane 2010) and logistic and Cox regression models (Pan 1999). In this paper,

we investigate model selection via bootstrap log-likelihood in the class of beta regression models. Such models were introduced by Ferrari and Cribari-Neto (2004) and are tailored for modeling responses that assume values in the standard unit interval, $(0, 1)$, such as rates and proportions. We consider the class of varying dispersion beta regressions, as described in Simas et al. (2010), Ferrari and Pinheiro (2011) and Cribari-Neto and Souza (2012). It generalizes the fixed dispersion beta regression model proposed by Ferrari and Cribari-Neto (2004). The model has two submodels, one for the mean and another one for the dispersion.

The chief goal of our paper is twofold. First, we propose new model selection criteria for beta regressions and then we numerically investigate their finite sample performances in small samples. We also provide simulation results on alternative model selection strategies. The numerical evidence shows that the criteria we propose typically yield reliable model selection in the class of beta regression models. Even though our focus lies in beta regression modeling, the two model selection criteria we propose can be used in other classes of regression models.

This paper is organized as follows. In the next section, we introduce the AIC and its bootstrap extensions. We also propose two new model selection criteria. Section 3 introduces the class of beta regression models. In Sect. 4, we present Monte Carlo simulation results on model selection in fixed and varying beta regression models. An empirical application is presented and discussed in Sect. 5. Finally, some concluding remarks are offered in Sect. 6.

2 Akaike information criterion and bootstrap variations

The distance measure between two densities can be measured using the Kullback–Leibler (KL) information (Kullback 1968), also known as entropy or discrepancy (Cavanaugh 1997). The KL information can be used to select an estimated model which is closest to the true model. The AIC was derived by Akaike (1973) by minimizing the KL information. In what follows, we shall follow Bengtsson and Cavanaugh (2006) to formalize the notion of selecting a model from a class of candidate models.

Suppose the n-dimensional vector Y is sampled from an unknown density $f(Y|\theta_{k_0})$, where θ_{k_0} is a k_0-vector of parameters. The respective parametric family of densities is denoted by $\mathcal{F}(k_i) = \left\{f(Y|\theta_{k_i})|\theta_{k_i} \in \Theta_{k_i}\right\}$, where Θ_{k_i} is the k_i-dimensional parametric space. Let $\hat{\theta}_{k_i}$ be the maximum likelihood estimate of θ_{k_i}. It is obtained by maximizing $f(Y|\theta_{k_i})$ in Θ_{k_i}, i.e., $f(Y|\hat{\theta}_{k_i})$ is the maximized likelihood function.

Using the AIC, it is possible to select the model that best approximates $f(Y|\theta_{k_0})$ from the class of families $\mathcal{F} = \{\mathcal{F}(k_1), \mathcal{F}(k_2), \ldots, \mathcal{F}(k_L)\}$. For notation simplicity, we will not consider different families in the class $\mathcal{F}$ which have the same dimension. We say that $f(Y|\hat{\theta}_k)$ is correctly specified if $f(Y|\theta_{k_0}) \in \mathcal{F}(k)$, where $\mathcal{F}(k)$ is the smallest dimensional family that contains $f(Y|\theta_{k_0})$. We say that $f(Y|\hat{\theta}_k)$ is overspecified if $f(Y|\theta_{k_0}) \in \mathcal{F}(k)$, but families of smaller dimension also contain $f(Y|\theta_{k_0})$. On the other hand, $f(Y|\hat{\theta}_k)$ is underspecified if $f(Y|\theta_{k_0}) \notin \mathcal{F}(k)$.

The KL measure can be used to determine which fitted model (i.e., which model in the collection $f(Y|\hat{\theta}_{k_1}), f(Y|\hat{\theta}_{k_2}), \ldots, f(Y|\hat{\theta}_{k_L})$) is closest to $f(Y|\theta_{k_0})$. The KL distance between the true model $f(Y|\theta_{k_0})$ and the candidate model $f(Y|\theta_k)$ is given by

$$d(\theta_{k_0}, \theta_k) = \mathrm{E}_0\left[\log\left\{\frac{f(Y|\theta_{k_0})}{f(Y|\theta_k)}\right\}\right],$$

where $\mathrm{E}_0(\cdot)$ denotes expectation under $f(Y|\theta_{k_0})$. Let

$$\delta(\theta_{k_0}, \theta_k) = \mathrm{E}_0\{-2\log f(Y|\theta_k)\}. \tag{1}$$

It is possible to show that $2d(\theta_{k_0}, \theta_k) = \delta(\theta_{k_0}, \theta_k) - \delta(\theta_{k_0}, \theta_{k_0})$. Since $\delta(\theta_{k_0}, \theta_{k_0})$ does not depend on θ_k minimizing $2d(\theta_{k_0}, \theta_k)$ or $d(\theta_{k_0}, \theta_k)$ is equivalent to minimizing the discrepancy $\delta(\theta_{k_0}, \theta_k)$. Therefore, the model $f(Y|\theta_k)$ that minimizes minus two times the expected log-likelihood, $\delta(\theta_{k_0}, \theta_k)$, is the closest model to the true model according to the Kullback–Leibler information.

Notice that,

$$\delta(\theta_{k_0}, \hat{\theta}_k) = \mathrm{E}_0\{-2\log f(Y|\theta_k)\}\,|_{\theta_k=\hat{\theta}_k}$$

measures the distance between the true model and the estimated candidate model. However, it is not possible to evaluate $\delta(\theta_{k_0}, \hat{\theta}_k)$, since it requires knowledge of density $f(Y|\theta_{k_0})$. Akaike (1973) used $-2\log f(Y|\hat{\theta}_k)$ as an estimator for $\delta(\theta_{k_0}, \hat{\theta}_k)$. Its bias

$$B = \mathrm{E}_0\left\{-2\log f(Y|\hat{\theta}_k) - \delta(\theta_{k_0}, \hat{\theta}_k)\right\} \tag{2}$$

can be asymptotically approximated by $-2k$, where k is the dimension of θ_k.

Thus, the expected value of Akaike's criterion,

$$\mathrm{AIC} = -2\log f(Y|\hat{\theta}_k) + 2k,$$

is asymptotically equal to the expected value of $\delta(\theta_{k_0}, \hat{\theta}_k)$, which is given by

$$\Delta(\theta_{k_0}, k) = \mathrm{E}_0\left\{\delta(\theta_{k_0}, \hat{\theta}_k)\right\}.$$

Notice that, $-2\log f(Y|\hat{\theta}_k)$ is a biased estimator of minus two times the expected log-likelihood and the penalizing term of the AIC, $2k$, is an adjustment term for the bias given in (2).

Since the AIC is based on a large sample approximation, it may perform poorly in small samples (Bengtsson and Cavanaugh 2006). Several variants of the AIC were developed aiming at delivering more accurate model selection in small samples. Sugiura (1978) developed the AICc, which in class of linear regression models is an unbiased estimator of $\Delta(\theta_{k_0}, k)$, that is, $\mathrm{E}_0\{\mathrm{AICc}\} = \Delta(\theta_{k_0}, k)$. Based on the results obtained by Sugiura (1978), Hurvich and Tsai (1989) extended the use of the AICc to cover nonlinear regression and for autoregressive models. The authors showed that the AICc is asymptotically equivalent to the AIC, i.e., $\mathrm{E}_0(\mathrm{AICc}) + o(1) = \Delta(\theta_{k_0}, k)$, and typically outperforms the AIC in small samples.

According to Cavanaugh (1997), the advantage of AICc over the AIC is that the former estimates the expected discrepancy more accurately than the latter. On the other hand, a clear advantage of the AIC over the AICc is that the AIC is universally

applicable, regardless of the class of models, whereas the AICc derivation is model dependent.

2.1 Bootstrap extensions of AIC

Bootstrap extensions of AIC (EIC) are criteria that use bootstrap estimators for the bias term B given in (2). They typically include a bias estimate which is more accurate than $-2k$ in small samples, thus leading to more reliable model selection. In what follows, we shall use five different bootstrap estimators, B_i $(i = 1, \ldots, 5)$ for B. The bias estimator B_i defines five bootstrap extensions of AIC which we denote by EICi, $i = 1, \ldots, 5$. The bootstrap variants of the AIC that we shall use for model selection in the class of beta regressions have the following form:

$$\text{EIC}i = -2\log f(Y|\hat{\theta}_k) + B_i, \quad i = 1, \ldots, 5.$$

Let Y^* be a bootstrap sample (generated either parametrically or nonparametrically) and let E_* denote the expected value with respect to distribution of Y^*. Consider W bootstrap samples $Y^*(i)$ and the corresponding estimates of $\hat{\theta}_k$: $\left\{\hat{\theta}_k^*(i)\right\}$, $i = 1, 2, \ldots, W$. Here, each estimate $\hat{\theta}_k^*(i)$ is the value of θ_k that maximizes the likelihood function $f(Y^*(i)|\theta_k)$.

Ishiguro et al. (1997) proposed a bootstrap extension of the AIC known as the EIC. It is a particular case of the WIC Ishiguro and Sakamoto (1991) obtained considering independent and identically distributed (i.i.d.) observations. We shall refer to such a criterion as EIC1. It estimates the bias in (2) as

$$B_1 = \text{E}_*\left\{2\log f(Y^*|\hat{\theta}_k^*) - 2\log f(Y|\hat{\theta}_k^*)\right\}.$$

A different bootstrap-based criterion was proposed in Cavanaugh and Shumway (1997) for the selection of state-space models; we shall refer to it as EIC2. The criterion estimates the bias in (2) as

$$B_2 = 2\text{E}_*\left\{2\log f(Y|\hat{\theta}_k) - 2\log f(Y|\hat{\theta}_k^*)\right\}.$$

We note that EIC1 and EIC2 are called AICb1 and AICb2, respectively, in Shang and Cavanaugh (2008) in the context of mixed models selection based on the parametric bootstrap.

Shibata (1997) showed that B_1 and B_2 are asymptotically equivalent and proposed the following three bootstrap estimators of (2):

$$B_3 = 2\text{E}_*\left\{2\log f(Y^*|\hat{\theta}_k^*) - 2\log f(Y^*|\hat{\theta}_k)\right\},$$

$$B_4 = 2\text{E}_*\left\{2\log f(Y^*|\hat{\theta}_k) - 2\log f(Y|\hat{\theta}_k^*)\right\},$$

$$B_5 = 2\text{E}_*\left\{2\log f(Y^*|\hat{\theta}_k^*) - 2\log f(Y|\hat{\theta}_k)\right\}.$$

We shall refer to the corresponding criteria as EIC3, EIC4 and EIC5.

Seghouane (2010) proposed corrected versions of the AIC for the linear regression model as asymptotic approximations to EIC1, EIC2, EIC3, EIC4 and EIC5 obtained using the parametric bootstrap.

2.2 Bootstrapped likelihood and cross-validation

The model selection criteria described so far aim at estimating the expected log-likelihood using a bias correction for the maximized log-likelihood function. Pan (1999), however, tried to obtain an estimator for the expected log-likelihood that does not require a bias adjustment. It uses cross-validation (CV) and bootstrap.

CV is widely used for estimating the error rate of prediction models (Efron 1983; Efron and Tibshirani 1997). In the context of model selection, according to Davies et al. (2005), the first CV-based criterion was the PRESS (Allen 1974). Bootstrap-based model selection was introduced by Efron (1986). Breiman and Spector (1992) and Hjorth (1994) discuss the use of CV and bootstrap in model selection.

According to Efron (1983) and Efron and Tibshirani (1997), CV typically reduces bias, but leads to variance inflation. Such variability can be reduced using the bootstrap method. In the context of model selection of models, Pan (1999) introduced a method that combines nonparametric bootstrap and CV: the bootstrapped likelihood CV (BCV). BCV yields an estimator of (1) that does not entail bias correction. For a sample Y of size n, the BCV is defined by

$$\mathrm{BCV} = \mathrm{E}_* \left\{ -2 \log f(Y^- | \hat{\theta}_k^*) \frac{n}{m^*} \right\},$$

where Y^* is the bootstrap sample generated nonparametrically, $Y^- = Y - Y^*$, that is, $Y = Y^- \cup Y^*$ and $Y^- \cap Y^* = \emptyset$, and $m^* > 0$ is the number of elements of Y^-. Thus, no observation of Y is used twice: each observation either belongs to Y^* or to Y^-.

Following Efron (1983), Pan (1999) argues that the BCV can overestimate (1) and, on the other hand, $-2 \log f(Y|\hat{\theta}_k)$ may underestimate it. Thus, following the 632+ rule Efron and Tibshirani (1997), Pan (1999) introduces the 632CV criterion as

$$632\mathrm{CV} = 0.368 \left\{ -2 \log f(Y|\hat{\theta}_k) \right\} + 0.632\mathrm{BCV}.$$

2.3 Proposed bootstrapped likelihood quasi-CV

We shall now introduce two new model selection criteria of models that incorporate corrections for small samples. Like the BCV, these criteria provide direct estimators for the expected log-likelihood.

Let F be the distribution function of the observed sample $Y = (y_1, \ldots, y_n)$ and let $\hat{F}$ be the estimated distribution function, i.e., $\hat{F}$ is the distribution function F evaluated at the estimative $\hat{\theta}$. We define

$$Y_p^* = (y_1^*, y_2^*, \ldots, y_n^*) \sim \hat{F} \quad \text{estimation sample (or training sample)},$$
$$Y = (y_1, y_2, \ldots, y_n) \sim F \quad \text{validation sample}.$$

Suppose, we have W pseudo-samples Y_p^* obtained from $\hat{F}$ and let $\{\hat{\theta}_k^{p*}(i),\ i = 1, 2, \ldots, W\}$ denote the set of W bootstrap replications of $\hat{\theta}_k$. We define the bootstrapped likelihood quasi-CV (BQCV) criterion as follows:

$$\text{BQCV} = \text{E}_{p*}\left\{-2\log f(Y|\hat{\theta}_k^{p*})\right\},$$

where E_{p*} is the expected value with respect to the distribution of Y_p^*.

It follows from the strong law of large numbers that

$$\frac{1}{W}\sum_{i=1}^{W}\left\{-2\log f(Y|\hat{\theta}_k^{p*}(i))\right\} \xrightarrow[W\to\infty]{a.s.} \text{E}_{p*}\left\{-2\log f(Y|\hat{\theta}_k^{p*})\right\},$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence.

The computation of BQCV can be performed as follows:

1. Estimate θ using the sample $Y = (y_1, \ldots, y_n)$;
2. Generate W pseudo-samples Y_p^* from $\hat{F}$;
3. For each $Y_p^*(i)$, $i = 1, \ldots, W$, compute $\hat{\theta}_k^{p*}(i)$ and $-2\log f(Y|\hat{\theta}_k^{p*}(i))$;
4. Using the W replications of $-2\log f(Y|\hat{\theta}_k^{p*})$ compute

$$\text{BQCV} = \frac{1}{W}\sum_{i=1}^{W}\left\{-2\log f(Y|\hat{\theta}_k^{p*}(i))\right\}.$$

Based on pilot simulations, we recommend using $W = 200$.

The algorithm outlined above is not a genuine cross-validation scheme, hence the name quasi-CV. It is not a genuine cross-validation scheme because it does not partition the sample Y, but instead it treats the samples Y_p^* and Y as partitions of the same data set. In each bootstrap replication, we use a procedure which is similar to the twofold CV. Here, the training sample is the pseudo-sample of the parametric bootstrap scheme, Y_p^*, and the validation sample is the observed sample, Y.

Following the approach used by Pan (1999) for obtaining the 632CV, we propose another model selection criterion, which we call 632QCV. It is a variant of the BQCV and is given by

$$632\text{QCV} = 0.368\left\{-2\log f(Y|\hat{\theta}_k)\right\} + 0.632\text{BQCV}.$$

3 The beta regression model

Many studies in different fields examine how a set of covariates is related to a response variable that assumes values in continuous interval, $(0, 1)$, such as rates and proportions; see, e.g., Brehm and Gates (1993); Hancox et al. (2010); Kieschnick and

McCullough (2003); Ferrari and Cribari-Neto (2004); Smithson and Verkuilen (2006); Zucco (2008); Verhaelen et al. (2013), and Whiteman et al. (2014). Such modeling can be done using the class of beta regression models, which was introduced by Ferrari and Cribari-Neto (2004). It assumes that the response variable (y) follows the beta law. The beta distribution is quite flexible since its density can assume a number of different shapes depending on the parameter values. The beta density can be indexed by mean (μ) and dispersion (σ) parameters when written as

$$f(y|\mu,\sigma)=\frac{\Gamma\left(\frac{1-\sigma^2}{\sigma^2}\right)}{\Gamma\left(\mu\left(\frac{1-\sigma^2}{\sigma^2}\right)\right)\Gamma\left((1-\mu)\left(\frac{1-\sigma^2}{\sigma^2}\right)\right)}y^{\mu\left(\frac{1-\sigma^2}{\sigma^2}\right)-1}(1-y)^{(1-\mu)\left(\frac{1-\sigma^2}{\sigma^2}\right)-1}, \tag{3}$$

where $0 < y < 1$, $0 < \mu < 1$, $0 < \sigma < 1$, $\Gamma(\cdot)$ is the gamma function and $V(\mu) = \mu(1-\mu)$ is the variance function. The mean and the variance of y are, respectively, by $\mathrm{E}(y) = \mu$ and $\mathrm{var}(y) = V(\mu)\sigma^2$.

Let $Y = (y_1, \ldots, y_n)$ be a vector of independent random variables, where y_t, $t = 1, \ldots, n$, has density (3) with mean μ_t and unknown dispersion σ_t. The varying dispersion beta regression model can be written as

$$g(\mu_t) = \sum_{i=1}^{r} x_{ti}\beta_i = \eta_t, \tag{4}$$

$$h(\sigma_t) = \sum_{i=1}^{s} z_{ti}\gamma_i = \nu_t, \tag{5}$$

where $\beta = (\beta_1, \ldots, \beta_r)^\top$ and $\gamma = (\gamma_1, \ldots, \gamma_s)^\top$ are vectors of unknown parameters and $x_t = (x_{t1}, \ldots, x_{tr})^\top$ and $z_t = (z_{t1}, \ldots, z_{ts})^\top$ are observations on r and s independent variables, $r + s = k < n$. In what follows, we denote the matrix of regressors used in the mean submodel by X, i.e., X is the $n \times r$ matrix whose tth line is x_t. Likewise, Z is the matrix of regressors used in the dispersion submodel. When intercepts are included in the mean and dispersion submodels, $x_{t1} = z_{t1} = 1$, for $t = 1, \ldots, n$. In addition, $g(\cdot)$ and $h(\cdot)$ are strictly monotonic and twice differentiable link functions with domain in $(0, 1)$ and image in $\mathbb{R}$. In the parameterization we use, the same link functions can be used in the mean and dispersion submodels. Commonly used link functions are logit, probit, log–log, complementary log–log and Cauchy. A detailed discussion of link functions can be found in McCullagh and Nelder (1989) and Koenker and Yoon (2009). Finally, we note that the constant dispersion beta regression model is obtained by setting $s = 1$, $z_{t1} = 1$ and $h(\cdot)$ is the identity function.

Joint estimation of β and γ can be performed by maximum likelihood. Let $\theta_k = (\beta_1, \ldots, \beta_r, \gamma_1, \ldots, \gamma_s)^\top$ and let be Y an n-vector of independent beta random variables. The log-likelihood function is

$$\log f(Y|\theta_k) = \sum_{t=1}^{n} \log f(y_t|\mu_t, \sigma_t),$$

where

$$\log f(y_t|\mu_t,\sigma_t) = \log\Gamma\left(\frac{1-\sigma_t^2}{\sigma_t^2}\right) - \log\Gamma\left(\mu_t\left(\frac{1-\sigma_t^2}{\sigma_t^2}\right)\right) - \log\Gamma\left((1-\mu_t)\left(\frac{1-\sigma_t^2}{\sigma_t^2}\right)\right)$$
$$+\left[\mu_t\left(\frac{1-\sigma_t^2}{\sigma_t^2}\right)-1\right]\log y_t + \left[(1-\mu_t)\left(\frac{1-\sigma_t^2}{\sigma_t^2}\right)-1\right]\log(1-y_t).$$

The score function is obtained by differentiating the log-likelihood function with respect to the unknown parameters. Closed-form expressions for the score function and Fisher's information matrix are given in Appendix A.

Let $U_\beta(\beta,\gamma)$ and $U_\gamma(\beta,\gamma)$ be the score functions for β and γ, respectively. The maximum likelihood estimators are obtained by solving

$$\begin{cases} U_\beta(\beta,\gamma) = 0, \\ U_\gamma(\beta,\gamma) = 0. \end{cases}$$

The solution to such a system of equations does not have a closed form. Hence, maximum likelihood estimates are usually obtained by numerically maximizing the log-likelihood function.

A global goodness-of-fit measure can be obtained by transforming the likelihood ratio as Nagelkerke (1991)

$$R^2_{LR} = 1 - \left(\frac{L_{\text{null}}}{L_{\text{fit}}}\right)^{2/n},$$

where L_{null} is the maximized likelihood function of the model without regressors and L_{fit} is the maximized likelihood function of the fitted regression model. An alternative measure is the square of the correlation coefficient between $g(y)$ and $\widehat{\eta} = X\widehat{\beta}$, where $\widehat{\beta}$ denotes the maximum likelihood estimator of β. Such a measure, which we denote by R^2_{FC}, was proposed by Ferrari and Cribari-Neto (2004) for constant dispersion beta regressions.

4 Numerical evaluation

In this section, we investigate the performances of the AIC and its bootstrap variations in small samples when used in the selection of beta regression models. All simulations were performed using the Ox matrix programming language (Doornik 2007). All log-likelihood maximizations were numerically carried out using the quasi-Newton nonlinear optimization algorithm known as BFGS with analytic first derivatives.[1]

We consider beta regression models with mean submodel as given in (4) and dispersion submodel as given in (5). We used 1000 Monte Carlo replications and, for each sample, $W = 200$ bootstrapped log-likelihoods were computed. We experimented with larger values of W but noticed that they only yielded negligible improvements

[1] For details on the BFGS algorithm, see Press et al. (1992).

in the model selection criteria performances. For the bootstrap extensions of AIC, we investigated the use of the parametric bootstrap, EICi_p, as well as the use of the nonparametric bootstrap, EICi_{np}. We also considered alternative model selection criteria in the Monte Carlo simulations: AICc (Hurvich and Tsai 1989), SIC (Schwarz 1978), SICc (McQuarrie 1999), HQ (Hannan and Quinn 1979) and HQc (McQuarrie and Tsai 1998).[2] The covariates values were obtained as random $\mathcal{U}(0, 1)$ draws; they were kept constant throughout the experiment. The logit link function was used in both submodels.

Performance evaluation of the different criteria is done as in Hannan and Quinn (1979), Hurvich and Tsai (1989), Shao (1996), McQuarrie et al. (1997), McQuarrie and Tsai (1998), Pan (1999), Shi and Tsai (2002), Davies et al. (2005), Shang and Cavanaugh (2008), Hu and Shao (2008), Liang and Zou (2008), and Seghouane (2010). For each criterion, we present the frequency of correct order selection ($=k_0$), as well as the frequencies of underspecified ($<k_0$) and overspecified ($>k_0$) selected models.

The following data generating processes were used:

$$\text{logit}(\mu_t) = -1.5 + x_{t2} + x_{t3}, \quad \text{logit}(\sigma_t) = -0.7 - 0.6x_{t2} - 0.6x_{t3}, \tag{6}$$

$$\text{logit}(\mu_t) = 1 - 0.75x_{t2} - 0.25x_{t3}, \quad \text{logit}(\sigma_t) = -0.7 - 0.5x_{t2} - 0.3x_{t3}, \tag{7}$$

$$\text{logit}(\mu_t) = -1.5 + x_{t2} + x_{t3}, \quad \text{logit}(\sigma_t) = -1.1 - 1.1x_{t2} - 1.1x_{t3}, \tag{8}$$

$$\text{logit}(\mu_t) = 1 - 0.75x_{t2} - 0.25x_{t3}, \quad \text{logit}(\sigma_t) = -1.45 - 1x_{t2} - 0.5x_{t3}. \tag{9}$$

The first two models, (6) and (7), entail large dispersion whereas the remaining two models, (8) and (9), have small dispersion. Considering the parameters values, we note that the regression models in (6) and (8) are easily identifiable whereas the models in (7) and (9) are weakly identifiable. In the weak identifiability scenario, variations in the covariates have different impacts on the mean response. The terminology "easily identified models" is used here in the same sense as in McQuarrie and Tsai (1998), Caby (2000) and Frazer et al. (2009). We emphasize that such a concept of model identifiability differs from the usual concept which relates to the model parameters uniqueness (Paulino and Pereira 1994; Rothenberg 1971). The numerical results for models with large and small dispersion are similar and, for that reason, we only present results for models with small dispersion, (8) and (9).

In all cases, the correct model order dimension is $k_0 = 6$: there are three parameters in the mean submodel and three parameters in the regression structure for the dispersion. The sample sizes are $n = 25, 30, 40, 50$ and five candidate covariates are considered for both submodels. The candidate models are sequentially nested for the mean submodel, that is, the candidate model with r parameters in the mean regression structure consists of the submodel with the $1, 2, \ldots, r$ first parameters. The dispersion submodels are also sequentially nested. Thus, for each value of r we vary s from 1 to 6, totaling $6 \times 6 = 36$ candidate models.

Since the true model belongs to the set of candidate models, the evaluation of the different selection criteria is done by counting the number of times that each criterion selects the correct model order (k_0, r_0 or s_0). Three different approaches

[2] The use of these criteria in beta regression models is done in an ad hoc manner.

Table 1 Frequencies of correct and incorrect order selection from 1000 independent replications; mean and dispersion regressors jointly selected in an easily identified model [Model (8)]

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	195	100	705	229	169	602	181	273	546	156	345	499
AICc	618	167	215	532	233	235	329	373	298	242	464	294
SIC	476	122	402	557	190	253	518	312	170	439	423	138
SICc	883	72	45	864	99	37	718	237	45	607	337	56
HQ	274	121	605	325	191	484	288	333	379	253	436	311
HQc	734	128	138	671	192	137	507	349	144	386	457	157
BQCV	861	107	32	640	267	93	309	**466**	225	187	**506**	307
632QCV	678	**234**	88	387	**371**	242	151	420	429	80	362	558
$EIC1_p$	964	28	8	950	28	22	893	77	30	886	73	41
$EIC2_p$	980	19	1	920	63	17	722	249	29	515	419	66
$EIC3_p$	856	129	15	521	368	111	267	422	311	203	429	368
$EIC4_p$	215	6	779	314	0	686	424	7	569	704	11	285
$EIC5_p$	93	9	898	97	11	892	505	53	442	821	103	76
$EIC1_{np}$	991	9	0	955	36	9	799	183	18	486	425	89
$EIC2_{np}$	997	3	0	985	11	4	921	76	3	675	300	25
$EIC3_{np}$	463	133	404	438	273	289	275	399	326	174	433	393
$EIC4_{np}$	998	2	0	981	15	4	894	99	7	674	293	33
$EIC5_{np}$	281	78	641	379	243	378	229	355	416	151	365	484
BCV	999	1	0	993	6	1	948	50	2	795	193	12
632CV	997	3	0	978	17	5	890	104	6	649	308	43

were considered. First, we used the different model selection criteria to jointly select the mean and dispersion regressors; the results are given in Tables 5 and 2. Afterwards, for a correctly specified dispersion submodel, we used the model selection criteria to select the regressors in the mean submodel; the results are given in Tables 1 and 4. Finally, for a correctly specified mean submodel, we performed model selection on the dispersion submodel; the results are presented in Tables 5 and 6. In all tables, the best results are highlighted.

The figures in Table 1 show that the proposed criteria yield reliable joint selection of mean and dispersion regressors in easily identifiable models. We note that for $n = 25$ and $n = 30$, 632QCV was the best performing criterion. For $n = 40$ and $n = 50$, BQCV was the best performer. Among the extensions (EIC's) of the AIC, the criterion that stands out is the EIC3 in their two versions, both with parametric and with nonparametric bootstrap. In this scenario, the AICc stands out when compared to alternative criteria that do not make use of bootstrapped log-likelihood. It is noteworthy the poor performance of the BCV, 632CV and EIC's criteria. When the sample size increases, the performances of the nonparametric EIC's improve, becoming similar. The same does not hold, however, for the parametric EIC's: $EIC1_p$ and $EIC4_p$ perform poorly in all sample sizes.

Table 2 Frequencies of correct and incorrect order selection from 1000 independent replications; mean and dispersion regressors jointly selected in a weakly identified model [Model (9)]

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	317	69	614	439	100	461	517	136	347	484	157	359
AICc	778	75	147	748	118	134	736	112	152	676	150	174
SIC	662	56	282	778	78	144	861	65	74	889	67	44
SICc	963	18	19	957	33	10	966	25	9	957	30	13
HQ	424	71	505	572	101	327	685	105	210	694	133	173
HQc	867	58	75	856	86	58	867	74	59	849	95	56
BQCV	866	82	52	791	145	64	699	161	140	544	**229**	227
632QCV	726	**158**	116	573	**237**	190	452	**218**	330	313	225	462
EIC1_p	960	30	10	931	50	19	901	61	38	827	113	60
EIC2_p	984	14	2	966	26	8	949	33	18	887	80	33
EIC3_p	885	88	27	721	213	66	657	154	189	587	180	233
EIC4_p	326	7	667	484	8	508	641	15	344	804	44	152
EIC5_p	11	0	989	29	0	971	228	9	763	630	128	242
EIC1_{np}	994	6	0	977	17	6	951	42	7	856	117	27
EIC2_{np}	1000	0	0	995	5	0	977	23	0	930	62	8
EIC3_{np}	593	91	316	672	160	168	636	172	192	582	169	249
EIC4_{np}	999	1	0	994	4	2	968	32	0	912	76	12
EIC5_{np}	362	34	604	588	148	264	572	208	220	496	204	300
BCV	1000	0	0	1000	0	0	991	9	0	969	28	3
632CV	999	1	0	994	6	0	975	25	0	911	76	13

Under weak identifiability, the good performances of the BQCV and 632QCV criteria become even more evident; see Table 2. The 632QCV criterion was the best performer for $n = 25, 30, 40$. For $n = 50$, BQCV outperformed the competition. It is noteworthy that for $n = 25, 30$, the 632QCV criterion outperformed all nonbootstrap-based criteria by at least 200 %. The EIC3 performs well relative to the other bootstrap extensions when regressors are jointly selected for both submodels in a weakly identifiable model. We also note the weak performances of the BCV and 632CV criteria. The AICc clearly outperforms the AIC. For instance, the AIC selected an overspecified model in 614 replications whereas that happened only 147 times when the AICc was used.

We shall now focus on selecting regressors for the mean submodel. Here, the dispersion submodel is correctly specified and the interest lies in identifying which covariates must be included in the mean submodel. The results for a weakly identifiable model are displayed in Table 3. They again show the good finite sample performances of our two model selection criteria. For $n = 25, 30$, the 632QCV criterion was the best performer. For $n = 40$, the best performer was BQCV, and for $n = 50$, the EIC2_p criterion outperformed the competition. Once again, the best performing AIC extension was EIC3 and the BCV and 632CV criteria performed poorly. The figures in Table 3

Table 3 Frequencies of correct and incorrect order selection from 1000 independent replications; mean regressors selected in an easily identified model [Model (8)]

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$
AIC	120	360	520	98	426	476	70	531	399	44	617	339
AICc	326	519	155	209	589	202	114	654	232	68	739	193
SIC	278	461	261	233	538	229	192	652	156	144	754	102
SICc	559	390	51	459	486	55	335	613	52	239	714	47
HQ	160	402	438	136	496	368	105	607	288	75	722	203
HQc	383	501	116	291	579	130	190	677	133	121	769	110
BQCV	487	510	3	279	699	22	115	**776**	109	56	801	143
632QCV	316	**668**	16	168	**779**	53	65	705	230	27	673	300
EIC1_p	932	68	0	904	94	2	869	110	21	827	157	16
EIC2_p	790	210	0	560	438	2	297	680	23	147	**823**	30
EIC3_p	543	456	1	279	694	27	112	705	183	58	737	205
EIC4_p	404	4	592	408	5	587	818	8	174	959	17	24
EIC5_p	313	2	685	722	4	274	968	9	23	968	13	19
EIC1_{np}	970	30	0	927	72	1	603	384	13	300	634	66
EIC2_{np}	974	26	0	958	41	1	725	268	7	471	495	34
EIC3_{np}	325	502	173	222	624	154	126	677	197	76	722	202
EIC4_{np}	974	26	0	956	43	1	723	269	8	463	497	40
EIC5_{np}	324	426	250	248	558	194	162	584	254	83	616	301
BCV	976	24	0	965	35	0	783	210	7	582	396	22
632CV	975	25	0	953	47	0	689	302	9	419	540	41

also show that the AICc and the HQc are the best performers among the criteria that do not use bootstrapped log-likelihood.

Table 4 contains the frequencies of correct model selection for the mean submodel when the model is weakly identifiable. The criteria that stands out are the same of the previous settings. For $n = 25, 30, 40$ ($n = 50$), 632QCV (BQCV) was the best performer. The EIC5_p criterion tends to select models that are overspecified in small samples; see also Table 2.

In our third and final approach, the mean submodel is correctly specified and the interest lies in selecting covariates for the dispersion submodel. The results are presented in Table 5. They show that the 632QCV criterion performs well when the model is easily identifiable; indeed, it was the best performer in all sample sizes. The 632QCV criterion was the only bootstrap AIC variant that outperformed all nonbootstrap-based criteria when $n = 25$. For the remaining sample sizes, only BQCV and EIC3_p outperformed the criteria that do not employ bootstrapped log-likelihood. Table 6 presents results for a weakly identifiable model. This was the only scenario in which 632QCV was not the best performing model selection criterion for $n = 25, 30$; it still performs well, nonetheless. For larger sample sizes, $n = 40, 50$, the proposed criterion was

Table 4 Frequencies of correct and incorrect order selection from 1000 independent replications; mean regressors selected in a weakly identified model [Model (9)]

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$<r_0$	$=r_0$	$>r_0$	$<r_0$	$=r_0$	$>r_0$	$<r_0$	$=r_0$	$>r_0$	$<r_0$	$=r_0$	$>r_0$
AIC	369	173	458	421	202	377	437	256	307	453	269	278
AICc	711	171	118	702	199	99	614	250	136	598	259	143
SIC	626	153	221	729	164	107	722	196	82	772	180	48
SICc	892	86	22	893	93	14	850	131	19	851	135	14
HQ	449	173	378	535	204	261	572	246	182	611	246	143
HQc	783	145	72	791	152	57	724	204	72	727	211	62
BQCV	846	153	1	779	211	10	600	330	70	546	**325**	129
632QCV	743	**247**	10	651	**307**	42	451	**376**	173	378	323	299
EIC1_p	941	58	1	908	90	2	836	139	25	796	168	36
EIC2_p	958	42	0	918	82	0	808	182	10	765	221	14
EIC3_p	856	142	2	737	243	20	582	297	121	551	288	161
EIC4_p	583	17	400	625	21	354	848	60	92	899	75	26
EIC5_p	52	1	947	238	4	758	764	89	147	796	101	103
EIC1_{np}	982	17	1	985	15	0	905	93	2	827	166	7
EIC2_{np}	983	16	1	988	12	0	945	54	1	875	123	2
EIC3_{np}	696	175	129	717	218	65	618	278	104	576	290	134
EIC4_{np}	983	16	1	990	10	0	940	59	1	873	124	3
EIC5_{np}	652	149	199	690	219	91	579	300	121	549	292	159
BCV	985	14	1	992	8	0	958	41	1	907	91	2
632CV	985	14	1	989	11	0	934	65	1	863	133	4

the best performer. For $n = 25$ ($n = 50$), model selection based on the HQ (EIC3_p) criterion was the most accurate.

The simulation results presented above lead to important conclusions on beta regression model selection. Such conclusions can be summarized as follows:

- The model selection criteria proposed in this paper generally work very well and lead to accurate model selection. The 632QCV criterion performed better as the sample size was small and the BQCV performed better in larger samples.
- Among the criteria that do not use the bootstrapped log-likelihood, the AICc and the HQc criteria were the best performers. The AICc stood out when the sample size was small and the HQc performed better in larger samples.
- Among the AIC extensions (EIC's), the EIC3 was the criterion that delivered most accurate model selection. Its nonparametric bootstrap implementation (EIC3_{np}) displayed the best performances in small samples and EIC3_p performed best in larger sample sizes.
- The finite sample performances of the different information criteria are considerably superior when such criteria are used to select regressors for the mean submodel rather than to pursue dispersion submodel selection; compare the results in Tables 3 and Table 5, and also the results in Tables 4 and 6.

Table 5 Frequencies of correct and incorrect order selection from 1000 independent replications; mean and dispersion regressors jointly selected in an easily identified model [Model (8)]

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$
AIC	328	230	442	295	313	392	254	401	345	246	466	288
AICc	632	252	116	523	342	135	396	452	152	339	499	162
SIC	567	221	212	553	303	144	519	394	87	535	409	56
SICc	838	143	19	785	183	32	679	297	24	650	331	19
HQ	398	247	355	378	332	290	361	430	209	357	480	163
HQc	705	223	72	620	301	79	508	421	71	467	465	68
BQCV	882	118	0	783	216	1	471	493	36	348	571	81
632QCV	734	**266**	0	574	**413**	13	304	**582**	114	214	**572**	214
$EIC1_p$	974	26	0	941	59	0	810	174	16	716	227	57
$EIC2_p$	994	6	0	969	31	0	800	196	4	664	327	9
$EIC3_p$	842	158	0	625	371	4	348	511	141	299	517	184
$EIC4_p$	617	21	362	622	15	363	701	55	244	809	116	75
$EIC5_p$	155	11	834	313	37	650	541	140	319	535	174	291
$EIC1_{np}$	1000	0	0	995	5	0	870	127	3	736	246	18
$EIC2_{np}$	1000	0	0	999	1	0	952	47	1	883	114	3
$EIC3_{np}$	564	236	200	509	355	136	358	468	174	290	480	230
$EIC4_{np}$	1000	0	0	1000	0	0	947	52	1	872	126	2
$EIC5_{np}$	533	183	284	514	303	183	382	388	230	318	388	294
BCV	1000	0	0	1000	0	0	962	38	0	922	76	2
632CV	1000	0	0	998	2	0	935	62	3	852	145	3

– The criteria that employ bootstrapped log-likelihood for beta regression model selection clearly outperform the competitors.

We emphasize that the two model selection criteria we propose can be used in other classes of regression models based on likelihood inferences, such as generalized linear models (McCullagh and Nelder 1989) and count data models (Winkelmann 2008). Numerical evaluation of their finite sample performances in different contexts will be done in future research.

5 Application

We use the data given in Griffiths et al. (1993) (Griffiths et al. 1993, Table 15.4) on food expenditure, income and number of people in 38 households of a major city in the United States. These data were modeled by Ferrari and Cribari-Neto (2004), who used a constant dispersion beta regression. We performed model selection using the two-step model selection scheme proposed in Bayer and Cribari-Neto (2015) coupled with the BQCV and 632QCV criteria proposed in this paper. In this scheme, the dispersion is taken to be constant and the mean submodel covariates are selected; next, using the selected mean submodel, model selection is carried out in the dispersion submodel.

Table 6 Frequencies of correct and incorrect order selection from 1000 independent replications; dispersion regressors selected in a weakly identified model [Model (9)]

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$
AIC	455	110	435	487	156	357	520	201	279	531	214	255
AICc	812	98	90	762	135	103	689	182	129	669	206	125
SIC	738	89	173	782	109	109	807	122	71	818	147	35
SICc	947	38	15	939	50	11	913	69	18	900	87	13
HQ	537	**114**	349	606	152	242	658	186	156	685	192	123
HQc	876	70	54	848	104	48	809	131	60	787	164	49
BQCV	975	25	0	928	70	2	800	175	25	690	243	67
632QCV	911	88	1	832	161	7	619	**290**	91	530	**293**	177
$EIC1_p$	991	9	0	967	33	0	900	93	7	833	139	28
$EIC2_p$	999	1	0	997	3	0	950	49	1	911	83	6
$EIC3_p$	943	57	0	812	**180**	8	665	237	98	610	239	151
$EIC4_p$	729	18	253	778	6	216	829	26	145	900	78	22
$EIC5_p$	45	2	953	181	16	803	559	147	294	566	188	246
$EIC1_{np}$	1000	0	0	1000	0	0	967	33	0	892	101	7
$EIC2_{np}$	1000	0	0	1000	0	0	985	15	0	950	47	3
$EIC3_{np}$	723	110	167	742	151	107	636	224	140	606	228	166
$EIC4_{np}$	1000	0	0	1000	0	0	987	13	0	945	51	4
$EIC5_{np}$	643	88	269	711	132	157	595	231	174	546	262	192
BCV	1000	0	0	1000	0	0	994	6	0	971	27	2
632CV	1000	0	0	1000	0	0	980	20	0	933	62	5

As shown in Bayer and Cribari-Neto (2015), this selection scheme tends typically outperforms the joint selection of regressors for the mean and dispersion submodels at a much lower computational cost. An implementation of such a model selection procedure in `R` language (R Core Team 2014) with the proposed BQCV and 632QCV criteria and two-step scheme is available at http://www.ufsm.br/bayer/auto-beta-reg.zip. The file contains computer code for model selection in beta regressions and also the dataset used in this empirical application.

Following Ferrari and Cribari-Neto (2004), we model the proportion of food expenditure (y) as a function of income (x_2) and of the number of people (x_3) in each household. We use the logit link function for the mean and dispersion submodels. The following covariates are also considered for inclusion in both submodels: the interaction between income and the number of people ($x_4 = x_2 \times x_3$), $x_5 = x_2^2$ and $x_6 = x_3^2$.

Assuming constant dispersion, the selected mean submodel, both by BQCV and by 632QCV, uses x_3 and x_4 as covariates. Assuming that this is the correct submodel for mean, we now select the regressors to be included in the dispersion submodel. The dispersion submodel selected by the BQCV and 632QCV criteria only includes one

Table 7 Parameter estimates of the selected varying dispersion beta regression model; data on food expenditure

Parameter	Estimate	Std. error	z stat	p value
Submodel for μ				
β_1 (Constant)	−1.3040	0.1103	−11.826	0.0000
β_3 (Number of people)	0.2890	0.0754	3.835	0.0005
β_4 (Interaction)	−0.0031	0.0011	−2.975	0.0054
Submodel for σ				
γ_1 (Constant)	−2.4825	0.3720	−6.673	0.0000
γ_3 (Number of people)	0.2011	0.1118	1.798	0.0813
$R^2_{FC} = 0.4586$				
$R^2_{LR} = 0.5448$				

covariate, namely: x_3. The parameter estimates of the selected model are presented in the Table 7.

We note that the parameter estimates show that there is a positive relation between the mean response and the number of people in each household, as well as a negative relationship with the interaction variable (x_4). There is also a positive relationship between the number of people in each household and the response dispersion. The varying dispersion beta regression model we selected and fitted has a pseudo-R^2 considerably larger than that of the constant dispersion model used by Ferrari and Cribari-Neto (2004): $R^2_{ML} = 0.5448$ versus $R^2_{ML} = 0.4088$.

6 Conclusions

In this paper, we considered the issue of beta regression model selection in small samples. We proposed two new model selection criteria for the class of varying dispersion beta regression models. The new criteria were obtained as bootstrap variations of the AIC and provide direct estimators for the expected log-likelihood. The proposed criteria are based on the bootstrap method and on a procedure called quasi-CV. They are then called bootstrapped likelihood quasi-CV (BQCV) and 632QCV. In addition to the proposed criteria, we investigated other criteria corrected for small samples. We did an extensive literature review and identified different bootstrap variations of the AIC that have been proposed for other classes of models. The finite sample performances of the proposed criteria relative to alternative model selection schemes were numerically evaluated in the context of varying dispersion beta regression modeling. The Monte Carlo evidence we presented favors the criteria we proposed: they typically lead to more accurate model selection than alternative criteria. We thus suggest the use of BQCV and 632QCV for beta regression model selection. An empirical application was also presented and discussed.

Acknowledgments We gratefully acknowledge partial financial support from CAPES, CNPq, and FAPERGS. We also thank two referees for comments and suggestions.

Appendix A: score function and information matrix of the beta regression model with varying dispersion

This appendix presents the score function and Fisher's information matrix for the varying dispersion beta regression model described in Sect. 3.

The score function is obtained by differentiating the log-likelihood function with respect to the unknown parameters. The score function of $\log f(Y|\theta_k)$ with respect to β is given by

$$U_\beta(\beta,\gamma)! = X^\top \Phi\, T(y^* - \mu^*),$$

where $\Phi = \text{diag}\left\{\frac{1-\sigma_1^2}{\sigma_1^2},\ldots,\frac{1-\sigma_n^2}{\sigma_n^2}\right\}$, $T = \text{diag}\left\{\frac{1}{g'(\mu_1)},\ldots,\frac{1}{g'(\mu_n)}\right\}$, $y^* = (y_1^*,\ldots,y_n^*)^\top$, $\mu^* = (\mu_1^*,\ldots,\mu_n^*)^\top$, $y_t^* = \log\left(\frac{y_t}{1-y_t}\right)$, $\mu_t^* = \psi\left(\mu_t\left(\frac{1-\sigma_t^2}{\sigma_t^2}\right)\right) - \psi\left((1-\mu_t)\left(\frac{1-\sigma_t^2}{\sigma_t^2}\right)\right)$ and $\psi(\cdot)$ is the digamma function, i.e., $\psi(u) = \frac{\partial \log\Gamma(u)}{\partial u}$, for $u > 0$. The score function of $\log f(Y|\theta_k)$ with respect to γ is

$$U_\gamma(\beta,\gamma) = Z^\top H a,$$

where $H = \text{diag}\left\{\frac{1}{h'(\sigma_1)},\ldots,\frac{1}{h'(\sigma_n)}\right\}$ and $a = (a_1,\ldots,a_n)^\top$, the tth element of a being $a_t = -\frac{2}{\sigma_t^3}\left\{\mu_t\left(y_t^* - \mu_t^*\right) + \log(1-y_t) - \psi\left((1-\mu_t)(1-\sigma_t^2)/\sigma_t^2\right) + \psi\left((1-\sigma_t^2)/\sigma_t^2\right)\right\}$.

Fisher's information matrix for β and γ is given by

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

where $K_{(\beta,\beta)} = X^\top \Phi W X$, $K_{(\beta,\gamma)} = (K_{(\gamma,\beta)})^\top = X^\top C T H Z$ and $K_{(\gamma,\gamma)} = Z^\top D Z$. Also, we have $W = \text{diag}\{w_1,\ldots,w_n\}$, $C = \text{diag}\{c_1,\ldots,c_n\}$ and $D = \text{diag}\{d_1,\ldots,d_n\}$, where

$$w_t = \frac{(1-\sigma_t^2)}{\sigma_t^2}\left[\psi'\left(\frac{\mu_t(1-\sigma_t^2)}{\sigma_t^2}\right) + \psi'\left(\frac{(1-\mu_t)(1-\sigma_t^2)}{\sigma_t^2}\right)\right]\frac{1}{[g'(\mu_t)]^2},$$

$$c_t = \frac{(2-2\sigma_t^2)}{\sigma_t^5}\left[\mu_t\psi'\left(\frac{\mu_t(1-\sigma_t^2)}{\sigma_t^2}\right) - (1-\mu_t)\psi'\left(\frac{(1-\mu_t)(1-\sigma_t^2)}{\sigma_t^2}\right)\right], \text{ and}$$

$$d_t = \frac{4}{\sigma_t^6}\left[\mu_t^2\psi'\left(\frac{\mu_t(1-\sigma_t^2)}{\sigma_t^2}\right) - (1-\mu_t)^2\psi'\left(\frac{(1-\mu_t)(1-\sigma_t^2)}{\sigma_t^2}\right) - \psi'\left(\frac{(1-\sigma_t^2)}{\sigma_t^2}\right)\right] \times \frac{1}{[h'(\sigma_t)]^2}.$$

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) Proceedings of the second international symposium on information theory, pp 267–281

Allen D (1974) The relationship between variable selection and data augmentation and a method for prediction. Technometrics 16:125–127

Bayer FM, Cribari-Neto F (2015) Model selection criteria in beta regression with varying dispersion. Commun Stat Simul Comp. doi:10.1080/03610918.2014.977918

Bengtsson T, Cavanaugh J (2006) An improved Akaike information criterion for state-space model selection. Comput Stat Data Anal 50(10):2635–2654

Brehm J, Gates S (1993) Donut shops and speed traps: evaluating models of supervision on police behavior. Am J Polit Sci 37(2):555–581

Breiman L, Spector P (1992) Submodel selection and evaluation in regression: the X-random case. Int Stati Rev 60:291–319

Caby E (2000) Review: regression and time series model selection. Technometrics 42(2):214–216

Cavanaugh J (1997) Unifying the derivations for the Akaike and corrected Akaike information criteria. Statist Probab Lett 33(2):201–208

Cavanaugh JE, Shumway RH (1997) A bootstrap variant of AIC for state-space model selection. Stat Sin 7:473–496

Cribari-Neto F, Souza T (2012) Testing inference in variable dispersion beta regressions. J Statist Comput Simul 82(12)

Davies S, Neath A, Cavanaugh J (2005) Cross validation model selection criteria for linear regression based on the Kullback–Leibler discrepancy. Stat Methodol 2(4):249–266

Doornik J (2007) An object-oriented matrix language Ox 5. Timberlake Consultants Press, London. http://www.doornik.com/

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Stat 7(1):1–26

Efron B (1983) Estimating the error rate of a prediction rule: improvement on cross-validation. J Am Stat Assoc 78(382):316–331

Efron B (1986) How biased is the apparent error rate of a prediction rule? J Am Stat Assoc 81(393):461–470

Efron B, Tibshirani R (1997) Improvements on cross-validation: the 632+ bootstrap method. J Am Stat Assoc 92(438):548–560

Ferrari SLP, Cribari-Neto F (2004) Beta regression for modelling rates and proportions. J Appl Stat 31(7):799–815

Ferrari SLP, Pinheiro EC (2011) Improved likelihood inference in beta regression. J Stat Comput Simul 81(4):431–443

Frazer LN, Genz AS, Fletcher CH (2009) Toward parsimony in shoreline change prediction (i): basis function methods. J Coastal Res 25(2):366–379

Griffiths WE, Hill RC, Judge GG (1993) Learning and practicing econometrics. Wiley, New York

Hancox D, Hoskin CJ, Wilson RS (2010) Evening up the score: sexual selection favours both alternatives in the colour-polymorphic ornate rainbowfish. Anim Behav 80(5):845–851

Hannan EJ, Quinn BG (1979) The determination of the order of an autoregression. J Roy Stat Soc Ser B 41(2):190–195

Hjorth JSU (1994) Computer intensive statistical methods: validation, model selection and Bootstrap. Chapman and Hall

Hu B, Shao J (2008) Generalized linear model selection using R^2. J Stat Plan Inf 138(12):3705–3712

Hurvich CM, Tsai CL (1989) Regression and time series model selection in small samples. Biometrika 76(2):297–307

Ishiguro M, Sakamoto Y (1991) WIC: an estimation-free information criterion., Research memorandumInstitute of Statistical Mathematics, Tokyo

Ishiguro M, Sakamoto Y, Kitagawa G (1997) Bootstrapping log likelihood and EIC, an extension of AIC. Ann Inst Stat Math 49(3):411–434

Kieschnick R, McCullough BD (2003) Regression analysis of variates observed on (0, 1): percentages, proportions and fractions. Stat Modell 3(3):193–213

Koenker R, Yoon J (2009) Parametric links for binary choice models: a fisherian-bayesian colloquy. J Econ 152(2):120–130

Kullback S (1968) Information theory and statistics. Dover

Liang H, Zou G (2008) Improved aic selection strategy for survival analysis. Comput Stat Data Anal 52:2538–2548
McCullagh P, Nelder J (1989) Generalized linear models, 2nd edn. Chapman and Hall
McQuarrie A, Shumway R, Tsai CL (1997) The model selection criterion AICu. Statist Probab Lett 34(3):285–292
McQuarrie A, Tsai CL (1998) Regression and time series model selection. World Scientific, Singapure
McQuarrie A (1999) A small-sample correction for the Schwarz SIC model selection criterion. Statist Probab Lett 44(1):79–86
Nagelkerke NJD (1991) A note on a general definition of the coefficient of determination. Biometrika 78(3):691–692
Pan W (1999) Bootstrapping likelihood for model selection with small samples. J Comput Graph Stat 8(4):687–698
Paulino CDM, Pereira CAB (1994) On identifiability of parametric statistical models. J Ital Stat Soc 3(1):125–151
Press W, Teukolsky S, Vetterling W, Flannery B (1992) Numerical recipes in C: the art of scientific computing, 2nd edn. Cambridge University Press
R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/
Rothenberg TJ (1971) Identification in parametric models. Econometrica 39(3):577–591
Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464
Seghouane AK (2010) Asymptotic bootstrap corrections of AIC for linear regression models. Signal Process 90:217–224
Shang J, Cavanaugh J (2008) Bootstrap variants of the Akaike information criterion for mixed model selection. Comput Stat Data Anal 52(4):2004–2021
Shao J (1996) Bootstrap model selection. J Am Stat Assoc 91(434):655–665
Shi P, Tsai CL (2002) Regression model selection: a residual likelihood approach. J Roy Stat Soc Ser B 64(2):237–252
Shibata R (1997) Bootstrap estimate of Kullback–Leibler information for model selection. Stat Sin 7:375–394
Simas AB, Barreto-Souza W, Rocha AV (2010) Improved estimators for a general class of beta regression models. Comput Stat Data Anal 54(2):348–366
Smithson M, Verkuilen J (2006) A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol Methods 11(1):54–71
Sugiura N (1978) Further analysts of the data by Akaike's information criterion and the finite corrections—further analysts of the data by Akaike's. Commun Stat Theor M 7(1):13–26
Verhaelen K, Bouwknegt M, Carratalà A, Lodder-Verschoor F, Diez-Valcarce M, Rodríguez-Lázaro D, de Roda Husman AM, Rutjes SA (2013) Virus transfer proportions between gloved fingertips, soft berries, and lettuce, and associated health risks. Int J Food Microbiol 166(3):419–425
Whiteman A, Young DE, He X, Chen TC, Wagenaar RC, Stern C, Schon K (2014) Interaction between serum BDNF and aerobic fitness predicts recognition memory in healthy young adults. Behav Brain Res 259(1):302–312
Winkelmann R (2008) Econometric analysis of count data, 5th edn. Springer, p 320
Zucco C (2008) The president's "new" constituency: Lula and the pragmatic vote in Brazil's 2006 presidential elections. J Lat Am Stud 40(1):29–49