Prediction Measures in Nonlinear Beta Regression Models

Patrícia Leone Espinheira¹ Luana C. Meireles da Silva¹ Alisson de Oliveira Silva¹ Raydonal Ospina¹,²

¹ Departamento de Estatística, Universidade Federal de Pernambuco, Cidade Universitária, Recife/PE, 50740–540, Brazil.
² E-mail: raydonal@de.ufpe.br

Abstract
Nonlinear models are frequently applied to determine the optimal supply natural gas to a given residential unit based on economical and technical factors, or used to fit biochemical and pharmaceutical assay nonlinear data. In this article we propose PRESS statistics and prediction coefficients for a class of nonlinear beta regression models, namely \( P^2 \) statistics. We aim at using both prediction coefficients and goodness-of-fit measures as a scheme of model select criteria. In this sense, we introduce for beta regression models under nonlinearity the use of the model selection criteria based on robust pseudo-\( R^2 \) statistics. Monte Carlo simulation results on the finite sample behavior of both prediction-based model selection criteria \( P^2 \) and the pseudo-\( R^2 \) statistics are provided. Three applications for real data are presented. The linear application relates to the distribution of natural gas for home usage in São Paulo, Brazil. Faced with the economic risk of too overestimate or to underestimate the distribution of gas has been necessary to construct prediction limits and to select the best predicted and fitted model to construct best prediction limits it is the aim of the first application. Additionally, the two nonlinear applications presented also highlight the importance of considering both goodness-of-predictive and goodness-of-fit of the competitive models.

Keywords: Nonlinear beta regression; PRESS; prediction coefficient; pseudo-\( R^2 \), power prediction.

1 Introduction

Ferrari and Cribari-Neto (2004) introduced a regression model in which the response is beta-distributed, its mean being related to a linear predictor through a link function. The linear predictor includes independent variables and regression parameters. Their model also includes a precision parameter whose reciprocal can be viewed as a dispersion measure. In the standard formulation of the beta regression model it is assumed that the precision is constant across observations. However, in many practical situations this assumption does not hold. Smithson and Verkuilen (2006) consider a beta regression specification in which dispersion is not constant, but is a function of covariates and unknown parameters and (Simas et al., 2010) introduces the class of nonlinear beta regression models. Parameter estimation is carried out by maximum likelihood (ML) and standard asymptotic hypothesis testing can be easily
performed. Practitioners can use the betareg package, which is available for the R statistical software (http://www.r-project.org), for fitting beta regressions. Cribari-Neto and Zeileis (2010) provide an overview of varying dispersion beta regression modeling using the betareg package. Diagnostic tools and improve ML estimation were accomplished in Espinheira et al. (2008a,b); Ospina et al. (2006); Chien (2011) and others. Inference for beta regression also have been developed in a Bayesian context (Figuero-Zúñiga et al. 2013; Brascum et al. (2007) and Cepeda-Cuervo and Gamerman (2005).)

Recently Espinheira et al. (2014) build and evaluated bootstrap-based prediction intervals for the class of beta regression models with varying dispersion. However, a prior approach it is necessary, namely: the selection of the model with the best predictive ability, regardless of the goodness-of-fit. Indeed, the model selection is a crucial step in data analysis, since all inferential performance is based on the selected model. Bayer and Cribari-Neto (2017) evaluated the performance of different model selection criteria in samples of finite size in a beta regression model, such as Akaike Information Criterion (AIC) (Akaike, 1973), Schwarz Bayesian Criterion (SBC) (Schwarz, 1978) and various approaches based on pseudo-$R^2$ such as the coefficient of determination adjusted $R^2_F$C proposed by Ferrari and Cribari-Neto (2004) and the version based on log-likelihood functions, namely by $R^2_{LR}$. Indeed, the authors proposed two new model selection criteria and a fast two step model selection scheme considering both mean and dispersion submodels, for beta regression models with varying dispersion.

However, these methods do not offer any insight about the quality of the predictive values in agreement with the findings of Spiess and Neumeyer (2010) for nonlinear models. In this context, Allen (1974), proposed the PRESS (Predictive Residual Sum of Squares) criterion, that can be used as a measure of the predictive power of a model. The PRESS statistic is independent from the goodness-of-fit of the model, since, that its calculation is made by leaving out the observations that the model is trying to predict (Palmer and O’Connell, 2009). The PRESS statistics can be viewed as a sum of squares of external residuals (Bartoli, 2009). Thus, similarly of the approach of $R^2$, Mediavilla et al. (2008) proposed a coefficient of prediction based on PRESS namely $P^2$. The $P^2$ statistic can be used to select models from a predictive perspective adding important information about the predictive ability of the model in various scenarios.

Our chief goal in this paper is to propose versions of the PRESS statistics and the coefficients of prediction $P^2$ associated, for the linear and nonlinear beta regression models. As a second contribution, in especial to beta regression under nonlinearity, we evaluate the behavior of $R^2_{LR}$ (Bayer and Cribari-Neto, 2017) and $R^2_{FC}$ (Ferrari and Cribari-Neto, 2004) measures both when the model is correctly specified and when under model misspecification. The results of the simulations showed as the prediction coefficients can be useful in detecting misspecifications, or indicate difficulties on to estimate beta regression models when the data are close to the boundaries of the standard unit interval. Finally, the real data applications are the last and important contribution. Here we provide guidance for researchers in choosing and interpreting the measures proposed. In fact, based on these applications we can shown how it is important to consider both coefficients of prediction and coefficients of determination to build models more useful to describe the data.

2 The $P^2$ statistic measure

Consider the linear model, $Y = X \beta + \varepsilon$ where $Y$ is a vector $n \times 1$ of responses, $X$ is a known matrix of covariates of dimension $n \times p$ of full rank, $\beta$ is the parameter vector of dimension $p \times 1$ and $\varepsilon$ is a vector $n \times 1$ of errors. We have the least squares estimators: $\hat{\beta} = (X^\top X)^{-1}X^\top y$, the residual $e_t = y_t - x_t^\top \hat{\beta}$ and the predicted value $\hat{y}_t = x_t^\top \hat{\beta}$, where $x_t = (x_{t1}, \ldots, x_{tp})$, and $t = 1, \ldots, n$. Let $\hat{\beta}_{(t)}$ be the least squares estimate of $\beta$ without the $t$th observation and $\hat{y}_{(t)} = x_t^\top \hat{\beta}_{(t)}$ be the predicted value of the case deleted.
such that \( e_{(t)} = y_t - \hat{y}_{(t)} \) is the prediction error or external residual. Thus, for multiple regression, the classic statistic
\[
\text{PRESS} = \sum_{t=1}^{n} e_{(t)}^2 = \sum_{t=1}^{n} (y_t - \hat{y}_{(t)})^2,
\]
which can be rewritten as \( \text{PRESS} = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2/(1 - h_{tt})^2 \), where \( h_{tt} \) is the \( t \)th diagonal element of the projection matrix \( X(X^\top X)^{-1}X^\top \).

Now, let \( y_1, \ldots, y_n \) be independent random variables such that each \( y_t \), for \( t = 1, \ldots, n \), is beta distributed, beta-distributed, denoted by \( y_t \sim B(\mu_t, \phi_t) \), i.e., each \( y_t \) has density function given by
\[
f(y_t; \mu_t, \phi_t) = \frac{\Gamma(\phi_t)}{\Gamma(\mu_t)\Gamma(1 - \mu_t)} y_t^{\mu_t - 1}(1 - y_t)^{(1 - \mu_t)\phi_t - 1}, \quad 0 < y_t < 1,
\]
where \( 0 < \mu_t < 1 \) and \( \phi_t > 0 \). Here, \( \text{E}(y_t) = \mu_t \) and \( \text{Var}(y_t) = V(\mu_t)/(1 + \phi_t) \), where \( V(\mu_t) = \mu_t(1 - \mu_t) \). Simas et al. (2010) proposed the class of nonlinear beta regression models in which the mean of \( y_t \) and the precision parameter can be written as
\[
g(\mu_t) = \eta_{1t} = f_1(x_t^\top; \beta) \quad \text{and} \quad h(\phi_t) = \eta_{2t} = f_2(z_t^\top, \gamma),
\]
where \( \beta = (\beta_1, \ldots, \beta_k)^\top \) and \( \gamma = (\gamma_1, \ldots, \gamma_q)^\top \) are, respectively, \( k \times 1 \) and \( q \times 1 \) vectors of unknown parameters \( (\beta \in \mathbb{R}^k; \gamma \in \mathbb{R}^q) \), \( \eta_{1t} \) and \( \eta_{2t} \) are the nonlinear predictors, \( x_t = (x_{t1}, \ldots, x_{tk}) \) and \( z_t = (z_{t1}, \ldots, z_{tq}) \) are vectors of covariates (i.e., vectors of independent variables), \( t = 1, \ldots, n \), \( k_1 \leq k \), \( q_1 \leq q \) and \( k + q < n \). Both \( g(\cdot) \) and \( h(\cdot) \) are strictly monotonic and twice differentiable link functions. Furthermore, \( f_1(\cdot), i = 1, 2, \) are differentiable and continuous functions, such that the matrices \( J_1 = \partial \eta_{1t}/\partial \beta \) and \( J_2 = \partial \eta_{2t}/\partial \gamma \) have full rank (their ranks are equal to \( k \) and \( q \), respectively). The parameters that index the model can be estimated by maximum likelihood (ML). In the Appendix, we present the log-likelihood function, the score vector and Fisher’s information matrix for the nonlinear beta regression model.

In the nonlinear beta regression model, the ML estimator \( \hat{\beta} \) can be viewed as the least squares estimator of \( \beta \) (see Appendix) obtained by regressing
\[
\hat{y} = \hat{\Phi}^{1/2} \hat{W}^{1/2} u_1 \quad \text{on} \quad \hat{J}_1 = \hat{\Phi}^{1/2} \hat{W}^{1/2} J_1,
\]
with \( \Phi = \text{diag}(\phi_1, \ldots, \phi_n) \), \( J_1 = \partial \eta_{1t}/\partial \beta \). Here, matrix \( W \) and \( u_1 \) are given in (15)–(17) in the Appendix. Thus, the prediction error is \( \hat{y}_t - \hat{\gamma}_{(t)} = \hat{\phi}_t^{1/2} \hat{w}_{t1}^{1/2} u_{1t} - \hat{\phi}_t^{1/2} \hat{w}_{t1}^{1/2} J_{1t} \hat{\beta}_{(t)} \), in which \( J_{1t}^\top \) is the \( t \)th row of the \( J_1 \) matrix. Using the ideas proposed by (Pregibon, 1981) we have that \( \hat{\beta}_{(t)} = \hat{\beta} - \left( \hat{J}_1^\top \hat{\Phi} \hat{W} J_1 \right)^{-1} J_{1t} \hat{\phi}_t^{1/2} \hat{w}_{t1}^{1/2} r_{t}^{\beta}/(1 - h_{tt}^*), \) where \( r_t^\beta \) is the weighted 1 residual (Espinheira et al., 2008a) defined as
\[
r_{t}^{\beta} = \frac{y_{t}^{*} - \hat{\mu}_t^{*}}{\sqrt{\hat{\nu}_t}},
\]
where, \( y_{t}^{*} = \log\{y_t/(1 - y_t)\} \), \( \mu_t^* = \psi(\mu, \phi_t) - \psi((1 - \mu_t)\phi_t) \) and \( \nu_t \) is given in 15 in the Appendix. Hence, we can write \( \hat{y}_t - \hat{\gamma}_{(t)} = r_{t}^{\beta}/(1 - h_{tt}^*) \), where \( h_{tt}^* \) is the \( t \)th diagonal element of projection matrix
\[
H^* = (\hat{\Phi} \hat{W})^{1/2} J_1 (J_1^\top \hat{\Phi} \hat{W} J_1)^{-1} J_{1t}^\top (\hat{\Phi} \hat{W})^{1/2}.
\]
Finally, for the nonlinear beta regressions models the classic PRESS statistic based on (1) is given by

\[
PRESS = \sum_{t=1}^{n} (\hat{y}_t - \bar{y}(t))^2 = \sum_{t=1}^{n} \left( \frac{r_{t}^{\beta}}{1 - h_{tt}} \right)^2. 
\] (6)

Note that the \( t^{th} \) observation in (6) is not used in fitting the regression model to predict \( y_t \), then both the external predicted values \( \hat{y}(t) \) and the external residuals \( e(t) \) are independent of \( y_t \). This fact enables the PRESS statistic to be a true assessment of the prediction capabilities of the regression model regardless of the overall model fit quality. Additionally, when the predictors in (3) are linear functions of the parameters, i.e., \( g(\mu_t) = x_t^{\top} \beta \) and \( h(\phi_t) = z_t^{\top} \gamma \), the expression in (6) also represent the PRESS statistic for a class of linear beta regression models with \( p = k + q \) unknown regression parameters.

Considering the same approach to construct the determination coefficient \( R^2 \) for linear models, we can think in a prediction coefficient based on PRESS, namely

\[
P^2 = 1 - \frac{PRESS}{SST_{(t)}},
\] (7)

where \( SST_{(t)} = \sum_{t=1}^{n}(y_t - \bar{y}(t))^2 \) and \( \bar{y}(t) \) is the arithmetic average of the \( y(t) \), \( t = 1, \ldots, n \). It can be shown that \( SST_{(t)} = (n/(n - p))^2SST \), wherein \( p \) is the number of model parameters and \( SST \) is the Total Sum of Squares for the full data. For a class of beta regression models with varying dispersion, \( SST = \sum_{t=1}^{n}(\hat{y}_t - \bar{y})^2 \), \( \bar{y} \) is the arithmetic average of the \( y_{(t)} \). It is noteworthy that the measures \( R^2 \) and \( P^2 \) are distinct, since that the \( R^2 \) propose to measure the model fit quality and the \( P^2 \) measure the predictive power of the model.

Cook and Weisberg (1982) suggest other versions of PRESS statistic based on different choices of residuals. Thus, we present another version of PRESS statistic and \( P^2 \) measure by considering the combined residual proposed by Espinheira et al. (2017). In this way,

\[
PRESS_{\beta\gamma} = \sum_{t=1}^{n} \left( \frac{r_{t}^{\beta\gamma}}{1 - h_{tt}} \right)^2 \quad \text{and} \quad P^{2}_{\beta\gamma} = 1 - \frac{PRESS_{\beta\gamma}}{SST_{(t)}},
\] (8)

respectively, where

\[
r_{t}^{\beta\gamma} = \frac{(y^*_t - \bar{u}^*_t) + \bar{a}_t}{\sqrt{\zeta_t}}, \quad a_t = \mu_t(y^*_t - \mu^*_t) + \log(1 - y_t) - \psi(1 - \mu_t)\phi_t + \psi(\phi_t) \\
and \quad \zeta_t = (1 + \mu_t)^2 \psi'(\mu_t\phi_t) + \mu_t^2 \psi'((1 - \mu_t)\phi_t) - \psi'(\phi_t).
\] (9)

Note that, \( P^2 \) and \( P^2_{\beta\gamma} \), given in (7) and (8), respectively, are not positive quantifiers. In fact, the \( PRESS/SST_{(t)} \) is a positive quantity, thus the \( P^2 \) and the \( P^2_{\beta\gamma} \) are measures that take values in \((-\infty; 1]\). The closer to one the better is the predictive power of the model.

In order to check the model goodness-of-fit with linear or nonlinear predictors for a class of beta regression, we evaluate the \( R^2_{FC} \) defined as the square of the sample coefficient of correlation between \( g(y) \) and \( \bar{y} \) (Ferrari and Cribari-Neto, 2004), and its penalized version based on Bayer and Cribari-Neto (2017) given by \( R^2_{FC} = 1 - (1 - R^2_{FC})(n - 1)/(n - (k_1 + q_1)) \), where \( k_1 \) and \( q_1 \) are, respectively, the number of covariates of the mean submodel and dispersion submodel.

We also evaluate two version of pseudo-\( R^2 \) based on likelihood ratio. The first one proposed by Nagelkerke (1991): \( R^2_{L,R} = 1 - (L_{null}/L_{fit})^{2/n} \), where \( L_{null} \) is the maximum likelihood achievable (satu-
rated model) and \( L_{\text{fit}} \) is the achieved by the model under investigation. The second one is a proposal of Bayer and Cribari-Neto (2017) that takes account the inclusion of covariates both in the mean submodel and in the precision submodel, given by:

\[
R^2_{LR_c} = 1 - (1 - R^2_{LR}) \left( \frac{n-1}{n-(1+\alpha)k_1-(1-\alpha)q_1} \right)^\delta,
\]

where \( \alpha \in [0,1] \) and \( \delta > 0 \). Based on simulation results obtained in Bayer and Cribari-Neto (2017) we choose in this work the values \( \alpha = 0.4 \) and \( \delta = 1 \). Therefore, penalized versions of \( P^2 \) and \( P^2_{\beta\gamma} \) are respectively now given by: \( P^2_c = 1 - (1-P^2)(n-1)/(n-(k_1+q_1)) \) and \( P^2_{\beta\gamma_c} = 1 - (1-P^2_{\beta\gamma})(n-1)/(n-(k_1+q_1)) \).

3 Simulation

In this section we simulate several different data generating processes to evaluate the performance of the predictive measures. The Monte Carlo experiments were carried out using both fixed and varying dispersion beta regressions as data generating processes. All results are based on 10,000 Monte Carlo replications.

**Linear models:** Table 1 shows the mean values of the predictive statistics obtained by simulation of fixed dispersion beta regression model that involves a systematic regression component for the mean given by

\[
\log \left( \frac{\mu_t}{1-\mu_t} \right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5}, \quad t = 1, \ldots, n,
\]

(10)

The covariate values were independently obtained as random draws of the following distributions: \( X_{ti} \sim U(0,1), \ i = 2, \ldots, 5 \) and were kept fixed throughout the experiment. The precisions, the sample sizes and the range of mean response are, respectively, \( \phi = (20, 50, 148, 400), n = (40, 80, 120, 400), \mu \in (0.005, 0.12), \mu \in (0.90, 0.99) \) and \( \mu \in (0.20, 0.88) \). Under the model specification given in (10) we investigate the performances of the statistics by omitting covariates. In this case, we considered the Scenarios 1, 2 and 3, in which are omitted, three, two and one covariate, respectively. In a fourth scenario the estimated model is correctly specified (true model).

The results in Table 1 show that the mean values of all statistics increase as covariates are included in the model and the value of \( \phi \) increases. On the other hand, as the size of the sample increases, the misspecification of the model is evidenced by lower values of the statistics (Scenarios 1, 2 and 3). It shall be noted that the values of all statistics are considerably larger when \( \mu \in (0.20, 0.88) \). Additionally, its values approaching one when the estimated model is closest to the true model. For instance, in Scenario 4 for \( n = 40, \phi = 150 \) the values of \( P^2 \) and \( R^2_{LR} \) are, respectively, 0.936 and 0.947.

The behavior of the statistics for finite samples changes substantially when \( \mu \in (0.90; 0.99) \). It is noteworthy the reduction of its values, revealing the difficulty in fitting the model and make prediction when \( \mu \approx 1 \) (The log-likelihood of the model tends to no longer limited). Indeed, in this range of \( \mu \) is more difficult to make prediction than to fit the model. For example, in Scenario 1, when three covariates are omitted from the model, \( n = 40 \) and \( \phi = 150 \) the \( P^2 \) value equals to 0.071, whereas the \( R^2_{LR} \) value is 0.243. Similar results were obtained for \( n = 80, 120 \). Even under true specification (Scenario 4) the model predictive power is more affected than the model quality of fit by the fact of \( \mu \approx 1 \). For instance, when \( n = 120 \) and \( \phi = 50 \) we have \( P^2_{\beta\gamma} = 0.046 \) and \( R^2_{LR} = 0.565 \). The same difficulty in obtaining
predictions and in fitting the regression model occurs when \( \mu \in (0.005, 0.12) \). Once again the greatest difficulty lies on the predictive sense.

Figure 1 present the boxplots of the 10,000 replications of the statistics: \( P^2_{\beta \gamma}, P^2_{\beta \gamma_c}, R^2_{LR}, R^2_{LRc}, R^2_{FC} \) and \( R^2_{FCc} \) when the model is correctly specified (scenario 4), \( n = 40 \) and \( \phi = 150 \). In all boxplots the “side point” represents the mean value of the replications of the statistics. In the panel (a) we present the boxplots when \( \mu \approx 0 \). In the panel (b) we present the boxplots when \( \mu \) is scattered on the standard unit interval and in the panel (c) we present the boxplots for \( \mu \approx 0 \). This figure shows that the means and the medians of all statistics are close. We also can notice based on the Figure 1 that both prediction power and goodness-of-fit of the model are affected when \( \mu \) is close to the boundaries of the standard unit interval. However, it is noteworthy the great difficult to make prediction. Additionally, is possible to notice that the versions of \( R^2 \) displays similar behavior. In it follows we shall investigate the empirical distributions behaviour of the statistics proposed.

![Correct specification](image)

**Figure 1:** Model estimated correctly: \( g(\mu_t) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5}; \mu \in (0.20, 0.88); \beta = (-1.9, 1.2, 1.0, 1.1, 1.3)^\top; \mu \in (0.90, 0.99); \beta = (1.8, 1.2, 1.1, 0.9)^\top; \mu \in (0.005, 0.12); \beta = (-1.5, -1.2, -1.0, -1.1, -1.3)^\top. \)

In Figures 2 and 3 we consider \( \mu \in (0.20, 0.88) \). In Figure 2 the model is estimated correctly, \( n = 40 \) and \( \phi = (20, 50, 150, 400) \). We notice that the prediction power distortions increase as the precision parameter increases, as expected. In Figure 3 we consider a misspecification problem (three omitted covariates). For illustration, we consider only \( \phi = 50 \) and \( n = 40, 80, 120, 400 \). It is important to notice that here the performance of the prediction measures does not deteriorate when the sample size is increased. Based on these figures we can reach some conclusions. First, the model precision affects both its predict power and goodness-of-fit. Second, for this range of \( \mu \) the performance of the statistics are similar revealing the correct specification of the model (Figure 2). Third, when three covariates are omitted, with the increasing of sample size the replications values of the statistics tend being concentrated at small values due to the misspecification problem (Figure 3).

In what follows, we shall report simulation results on the finite-sample performance of the statistics when the dispersion modeling is neglected. To that end, the true data generating process considers varying dispersion, but a fixed dispersion beta regression is estimated. We also used different covariates in the mean and precision submodels. The samples sizes are \( n = 40, 80, 120 \). We generated 20 values for each covariate and replicated them to get covariate values for the three sample sizes (once, twice and three times, respectively).
**Table 1:** Values of the statistics. True model versus misspecification models (omitted covariates (Scenarios 1, 2 and 3)).

| Estimated model | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 |
|-----------------|------------|------------|------------|------------|
| $\phi \rightarrow$ | 20 50 150 | 20 50 150 | 20 50 150 | 20 50 150 |
| $\beta \in (0.20, 0.88)$ | $\beta = (-1.9, 1.2, 1.0, 1.1, 1.3)^T$. | $\beta = (-1.9, 1.2, 1.0, 1.1, 1.3)^T$. | $\beta = (-1.9, 1.2, 1.0, 1.1, 1.3)^T$. | $\beta = (-1.9, 1.2, 1.0, 1.1, 1.3)^T$. |
| 40 | | | | |
| $P_{\beta_{LR}}$ | 0.270 0.391 0.385 0.420 0.473 0.512 0.518 0.620 0.671 0.700 0.860 0.942 | 0.506 0.561 0.624 0.643 0.681 0.821 0.925 | 0.450 0.557 0.617 0.649 0.694 0.825 0.927 | \(P_{\beta_{LR}}\) |
| $R_{\beta_{LR}}$ | 0.270 0.391 0.385 0.420 0.473 0.512 0.518 0.620 0.671 0.700 0.860 0.942 | 0.506 0.561 0.624 0.643 0.681 0.821 0.925 | 0.450 0.557 0.617 0.649 0.694 0.825 0.927 | \(R_{\beta_{LR}}\) |
| 80 | | | | |
| $P_{\beta_{LR}}$ | 0.282 0.339 0.370 0.377 0.455 0.498 0.518 0.566 0.600 0.666 0.833 0.939 | 0.512 0.590 0.650 0.666 0.666 0.666 0.833 0.939 | 0.450 0.557 0.617 0.649 0.694 0.825 0.927 | \(P_{\beta_{LR}}\) |
| $R_{\beta_{LR}}$ | 0.282 0.339 0.370 0.377 0.455 0.498 0.518 0.566 0.600 0.666 0.833 0.939 | 0.512 0.590 0.650 0.666 0.666 0.666 0.833 0.939 | 0.450 0.557 0.617 0.649 0.694 0.825 0.927 | \(R_{\beta_{LR}}\) |
| 120 | | | | |
| $P_{\beta_{LR}}$ | 0.279 0.346 0.374 0.360 0.439 0.483 0.469 0.578 0.638 0.666 0.833 0.928 | 0.506 0.584 0.643 0.666 0.666 0.666 0.833 0.928 | 0.450 0.557 0.617 0.649 0.694 0.825 0.927 | \(P_{\beta_{LR}}\) |
| $R_{\beta_{LR}}$ | 0.279 0.346 0.374 0.360 0.439 0.483 0.469 0.578 0.638 0.666 0.833 0.928 | 0.506 0.584 0.643 0.666 0.666 0.666 0.833 0.928 | 0.450 0.557 0.617 0.649 0.694 0.825 0.927 | \(R_{\beta_{LR}}\) |

$\mu \in (0.90, 0.99)$; $\beta = (1.8, 1.2, 1.1, 0.9, 0.7)^T$.  

| Estimated model | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 |
|-----------------|------------|------------|------------|------------|
| $\phi \rightarrow$ | 20 50 150 | 20 50 150 | 20 50 150 | 20 50 150 |
| $\beta \in (0.85, 0.12)$ | $\beta = (-1.5, -1.2, -1.0, -1.1, -1.3)^T$. | $\beta = (-1.5, -1.2, -1.0, -1.1, -1.3)^T$. | $\beta = (-1.5, -1.2, -1.0, -1.1, -1.3)^T$. | $\beta = (-1.5, -1.2, -1.0, -1.1, -1.3)^T$. |
| 40 | | | | |
| $P_{\beta_{LR}}$ | 0.270 0.352 0.352 0.429 0.472 0.456 0.564 0.625 0.694 0.825 0.927 | 0.506 0.602 0.656 0.694 0.847 0.936 | 0.450 0.557 0.617 0.649 0.825 0.927 | \(P_{\beta_{LR}}\) |
| $R_{\beta_{LR}}$ | 0.270 0.352 0.352 0.429 0.472 0.456 0.564 0.625 0.694 0.825 0.927 | 0.506 0.602 0.656 0.694 0.847 0.936 | 0.450 0.557 0.617 0.649 0.825 0.927 | \(R_{\beta_{LR}}\) |
| 80 | | | | |
| $P_{\beta_{LR}}$ | 0.282 0.339 0.375 0.361 0.447 0.494 0.463 0.593 0.655 0.686 0.851 0.939 | 0.506 0.584 0.643 0.666 0.666 0.666 0.833 0.939 | 0.450 0.557 0.617 0.649 0.694 0.825 0.927 | \(P_{\beta_{LR}}\) |
| $R_{\beta_{LR}}$ | 0.282 0.339 0.375 0.361 0.447 0.494 0.463 0.593 0.655 0.686 0.851 0.939 | 0.506 0.584 0.643 0.666 0.666 0.666 0.833 0.939 | 0.450 0.557 0.617 0.649 0.694 0.825 0.927 | \(R_{\beta_{LR}}\) |
| 120 | | | | |
| $P_{\beta_{LR}}$ | 0.279 0.345 0.374 0.360 0.439 0.483 0.469 0.578 0.639 0.666 0.833 0.928 | 0.506 0.584 0.643 0.666 0.666 0.666 0.833 0.928 | 0.450 0.557 0.617 0.649 0.694 0.825 0.927 | \(P_{\beta_{LR}}\) |
| $R_{\beta_{LR}}$ | 0.279 0.345 0.374 0.360 0.439 0.483 0.469 0.578 0.639 0.666 0.833 0.928 | 0.506 0.584 0.643 0.666 0.666 0.666 0.833 0.928 | 0.450 0.557 0.617 0.649 0.694 0.825 0.927 | \(R_{\beta_{LR}}\) |
contain the values of the statistics. We notice that for each scenario the prediction power 
would remain constant as the sample size changes. The numerical results were obtained using the 
following beta regression model $g(\mu_t) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5}$. $\mu \in (0.20, 0.88)$; $\beta = (-1.9, 1.2, 1.0, 1.1, 1.3)\top$.

This was done so that the intensity degree of nonconstant dispersion

$$
\lambda = \frac{\phi_{\max}}{\phi_{\min}} = \max_{t=1,\ldots,n} \{\phi_t\} \min_{t=1,\ldots,n} \{\phi_t\},
$$

would remain constant as the sample size changes. The numerical results were obtained using the 
following beta regression model $g(\mu_t) = \log(\mu_t/(1-\mu_t)) = \beta_1 + \beta_i x_{ti}$, and $\log(\phi_t) = \gamma_1 + \gamma_i z_{ti}$, $x_{ti} \sim U(0,1)$, $z_{ti} \sim U(-0.5, 0.5)$, $i = 2,3,4,5$, and $t = 1,\ldots,n$ under different choices of parameters (Scenarios): **Scenario 5**: $\beta = (-1.3, 3.2)\top$, $\mu \in (0.22, 0.87)$, $[\gamma = (3.5, 3.0)\top; \lambda \approx 20]$, $[\gamma = (3.5, 4.0)\top; \lambda \approx 50]$ and $[\gamma = (3.5, 5.0)\top; \lambda \approx 150]$. **Scenario 6**: $\beta = (-1.9, 1.2, 1.6, 2.0)\top$, $\mu \in (0.24, 0.88)$, $[\gamma = (2.4, 1.2, -1.7, 1.0)\top; \lambda \approx 20]$, $[\gamma = (2.9, 2.0, -1.7, 2.0)\top; \lambda \approx 50]$ and $[\gamma = (2.9, 2.0, -1.7, 2.8)\top; \lambda \approx 100]$. **Scenarios 7 and 8 (Full models)**: $\beta = (-1.9, 1.2, 1.0, 1.1, 1.3)\top$, $\mu \in (0.20, 0.88)$, $[\gamma = (3.2, 2.5, -1.1, 1.9, 2.2)\top; \lambda \approx 20]$, $[\gamma = (3.2, 2.5, -1.1, 1.9, 3.2)\top; \lambda \approx 50]$, and $[\gamma = (3.2, 2.5, 1.1, 1.9, 4.0)\top; \lambda \approx 100]$. All results were obtained using 10,000 replics Monte Carlo replications.

Table 2 contain the values of the statistics. We notice that for each scenario the prediction power measure not present high distortion when we increase intensity degree of nonconstant dispersion. However, in the case of misspecification the statistics display smaller values in comparison with Scenario 8 (True specification), in which as greater is $\lambda$ greater are the values of the statistics, as expected. Other important impression lies in the fact that the values of the $R^2_{FC}$ are considerably smaller than the values of the others statistics, in special when $\lambda$ increases.

That is a strong evidence that the $R^2_{FC}$ does not perform well under nonconstant dispersion models.
Table 2: Values of the statistics. Misspecified models, $\phi$ fixed: Scenarios 5, 6 and 7 versus Scenario 8 (correct specification).

| Scenario 5 | Scenario 6 | Scenario 7 | Scenario 8 |
|------------|------------|------------|------------|
| $g(\mu_1) = \beta_1 + \beta_2 x_{12} + \beta_3 x_{13}$ | $g(\mu_1) = \beta_1 + \beta_2 x_{12} + \beta_3 x_{13} + \beta_4 x_{14}$ | $g(\mu_1) = \beta_1 + \beta_2 x_{12} + \beta_3 x_{13} + \beta_4 x_{14} + \beta_5 x_{15}$ | $g(\mu_1) = \beta_1 + \beta_2 x_{12} + \beta_3 x_{13} + \beta_4 x_{14} + \beta_5 x_{15}$ |
| $h(\phi_1) = \gamma_1 + \gamma_2 x_{12} + \gamma_3 x_{13}$ | $h(\phi_1) = \gamma_1 + \gamma_2 x_{12} + \gamma_3 x_{13} + \gamma_4 x_{14}$ | $h(\phi_1) = \gamma_1 + \gamma_2 x_{12} + \gamma_3 x_{13} + \gamma_4 x_{14} + \gamma_5 x_{15}$ | $h(\phi_1) = \gamma_1 + \gamma_2 x_{12} + \gamma_3 x_{13} + \gamma_4 x_{14} + \gamma_5 x_{15}$ |

$\lambda \rightarrow P^2_{0.20,0.88}, \phi = 50$

$P^2_{0.20,0.88}$

$0.759$ $0.718$ $0.674$ $0.545$ $0.563$ $0.523$ $0.638$ $0.624$ $0.529$ $0.885$ $0.906$ $0.914$

$P^2_{0.20,0.88}$

$0.739$ $0.695$ $0.647$ $0.493$ $0.515$ $0.469$ $0.585$ $0.569$ $0.460$ $0.851$ $0.878$ $0.888$

$P^2_{0.20,0.88}$

$0.758$ $0.716$ $0.671$ $0.546$ $0.567$ $0.528$ $0.637$ $0.624$ $0.530$ $0.885$ $0.906$ $0.913$

$P^2_{0.20,0.88}$

$0.738$ $0.693$ $0.643$ $0.494$ $0.517$ $0.474$ $0.584$ $0.568$ $0.460$ $0.851$ $0.878$ $0.888$

$R^2_{LR}$

$0.782$ $0.743$ $0.700$ $0.580$ $0.611$ $0.577$ $0.670$ $0.653$ $0.554$ $0.796$ $0.816$ $0.840$

$R^2_{LR}$

$0.764$ $0.722$ $0.675$ $0.532$ $0.567$ $0.529$ $0.622$ $0.602$ $0.488$ $0.735$ $0.761$ $0.792$

In fact, under nonconstant dispersion models the better performances are of the $P^2$ statistics, both in identifying wrong and correct specifications.

In fact, under nonconstant dispersion models the better performances are of the $P^2$ statistics, both in identifying wrong and correct specifications.
other measures. The comparison among the best measures indicate that the median of $P^2$ and $R^2_{LR}$ performance are significantly better than the measures based on pseudo-$R^2$ and besides reveal some asymmetry of the statistics when the intensity degree of nonconstant dispersion levels increasing. These findings hold in all observation scenarios.

Figure 4: True model: $g(\mu_t) = \beta_1 + \beta_2 x_{t2}$, $h(\phi_t) = \gamma_1 + \gamma_2 z_t$.

**Nonlinear models:** In it follows we shall present Monte Carlo experiments for the class of nonlinear beta regression models. To that end we shall use the starting values scheme for the estimation by maximum likelihood proposed by Espinheira et al. (2017). The numerical results were obtained using the following beta regression model as data generating processes:

$$\log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + x_{t2}^{\beta_2} + \beta_3 \log(x_{t3} - \beta_4) + \frac{x_{t3}}{\beta_5}, \quad t = 1, \ldots, n$$

$x_{t2} \sim U(1, 2), x_{t3} \sim U(4.5, 34.5)$ and $\phi$ were kept fixed throughout the experiment. The precisions and the sample sizes are $\phi = (20, 50, 150, 400)$, $n = (20, 40, 60, 200, 400)$. Here, the vector of the parameters of the submodel of mean is $\beta = (1.0, 1.9, -2.0, 3.4, 7.2)^{\top}$ that produce approximately a range of values for the mean given by $\mu \in (0.36, 0.98)$. To evaluate the performances of statistics on account of nonlinearity negligence we consider the following model specification: $\log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}$. All results are based on 10,000 Monte Carlo replications an for each replication, we generated the response values as $y_t \sim B(\mu_t, \phi_t)$, $t = 1, \ldots, n$.

Table 3 contains numerical results for the fixed dispersion beta regression model as data generating processes. Here, we compared the performances of the statistics both under incorrect specifications and under correct specification of the nonlinear beta regression model. The results presented in this table
reveal that the \( P^2 \) and \( P^2_{\beta\gamma} \) statistics outperform the \( R^2 \) statistics in identifying more emphatically the misspecification. We must emphasize that the response mean is scattered on the standard unit interval.

Thus, we should not have problems to make prediction and the smaller values of \( P^2 \) statistics in comparison with the values of the \( R^2 \) statistics is due to the better performance of the statistics based on residuals in identifying misspecification problems. For example, fixing the precision on \( \phi = 400 \), for \( n = 20 \), we have values of \( P^2, P^2_{\beta\gamma}, R^2_{LR} \) and \( R^2_{FC} \) equal to 0.576, 0.601, 0.700, 0.637, respectively. For \( n = 40 \) and \( n = 60 \) the values of the statistics are 0.568, 0.593, 0.698, 0.634 and 0.562, 0.588, 0.698, 0.633, respectively. We can also notice that the values of the penalized versions of the statistics tend to be greater as the sample size increasing, what it makes sense.

Figure 5 summarizes the predictive power measure performance with boxplots over the Monte Carlo replicas. The boxplots clearly show the statistical significance of the performance differences between the measures. The outperformance of the \( P^2 \) and \( P^2_{\beta\gamma} \) statistics in identifying misspecification is more clear when we analyzed the plot. When the sample size increases, the distributions of the statistics based on residuals tend been concentrated in small values. For the other hand, the distributions of the \( R^2 \) statistics tend been concentrated at the same values, considerably greater than the values of the \( P^2 \) and \( P^2_{\beta\gamma} \) statistics.

Figure 6 summarizes the empirical distribution behavior of predictive power measure when \( n = 60 \). The graphs show that the median performance of \( R^2_{LR} \) is significantly worse than the performance of the \( P^2 \) and \( P^2_{\beta\gamma} \) measures. However, under true specification the statistics perform equally well and as the precision of the model increases the values of the statistics tend being concentrated close to one. Also, we notice that the performance comparison among different levels of \( \phi \) shows a systematic increase of the power performance.

Table 3: Values of the statistics. True model: \( g(\mu_t) = \beta_1 + x_{t2}^{\beta_2} + \beta_3 \log(x_{t3} - \beta_4) + x_{t3}^{\beta_5}, x_{t2} \sim U(1, 2), x_{t3} \sim U(4.5, 34.5), \beta = (1.0, 1.9, -2.0, 3.4, 7.2)^{\top}, \mu \in (0.36, 0.98), t = 1, \ldots, n, \phi \) fixed. Misspecification: \( g(\mu_t) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} \) (omitted nonlinearity).

| Estimated Model | With misspecification: \( g(\mu_t) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} \) | Correctly |
|-----------------|-------------------------------------------------|-----------|
| \( \phi \to \)  | 20      50     150     400                  | 20  50    150     400                | 20  50    150     400                |
| \( P^2 \)       | 0.485  0.535  0.564  0.576                     | 0.438  0.508  0.550  0.568            | 0.420  0.496  0.543  0.562          |
| \( P^2_{\beta\gamma} \) | 0.388  0.448  0.483  0.497                     | 0.391  0.467  0.513  0.532            | 0.388  0.469  0.518  0.539          |
| \( R^2_{LR} \)  | 0.502  0.556  0.588  0.601                     | 0.456  0.531  0.575  0.593            | 0.439  0.520  0.568  0.588          |
| \( R^2_{LRc} \) | 0.409  0.473  0.511  0.526                     | 0.411  0.492  0.539  0.559            | 0.409  0.494  0.545  0.566          |
| \( R^2_{FC} \)  | 0.578  0.647  0.684  0.700                     | 0.563  0.639  0.681  0.698            | 0.557  0.636  0.680  0.698          |
| \( R^2_{FCc} \) | 0.499  0.581  0.625  0.643                     | 0.526  0.608  0.654  0.673            | 0.533  0.616  0.662  0.681          |
| \( R^2_{FCc} \) | 0.486  0.574  0.619  0.637                     | 0.448  0.556  0.612  0.634            | 0.437  0.550  0.609  0.633          |
| \( R^2_{FCc} \) | 0.389  0.494  0.548  0.569                     | 0.402  0.519  0.580  0.604            | 0.407  0.526  0.588  0.613          |

**Nonlinearity on dispersion model:** The last simulations consider two nonlinear submodels both to mean and dispersion, namely:

\[
\log \left( \frac{\mu_t}{1 - \mu_t} \right) = \beta_1 + x_{t2}^{\beta_2} \quad \text{and} \quad \log (\phi_t) = \gamma_1 + z_{t2}^{\gamma_2}.
\]
Figure 5: Misspecification: omitted nonlinearity. Estimated model: $g(\mu_t) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}$. True model: $g(\mu_t) = \beta_1 + x_{t2}^{\beta_2} + \beta_3 \log(x_{t3} - \beta_4) + \frac{x_{t3}}{\beta_5}$, $x_{t2} \sim U(1,2)$, $x_{t3} \sim U(4.5,34.5)$, $\beta = (1.0, 1.9, -2.0, 3.4, 7.2)^T$, $t = 1, \ldots, n$, $\phi = 150$, $\mu \in (0.36, 0.98)$.

Figure 6: Model correctly specified. True model: $g(\mu_t) = \beta_1 + x_{t2}^{\beta_2} + \beta_3 \log(x_{t3} - \beta_4) + \frac{x_{t3}}{\beta_5}$, $x_{t2} \sim U(1,2)$, $x_{t3} \sim U(4.5,34.5)$, $\beta = (1.0, 1.9, -2.0, 3.4, 7.2)^T$, $t = 1, \ldots, n$, $\phi = 150$, $\mu \in (0.36, 0.98)$.

We fixed: $n = 400$, $\beta = (-1.1, 1.7)^T$, $x_t \sim U(0.3, 1.3)$; $(\mu \in (0.28, 0.61))$, $z_t \sim U(0.5, 1.5)$ and we varying $\gamma$ such that $\gamma = (2.6, 3.0)^T$; $\lambda \approx 25$, $\gamma = (1.6, 3.1)^T$; $\lambda \approx 29$, $\gamma = (0.9, 3.2)^T$; $\lambda \approx 35$ and $\gamma = (-0.3, 3.9)^T$; $\lambda \approx 100$, $t = 1, \ldots, n$. In Figure 7 we present the boxplots of the $P^2$, $P_{2\gamma}^2$ and $R_{LR}^2$ under negligence of nonlinearity, that is the estimated model is $\log \left( \frac{\mu_t}{1-\mu_t} \right) = \beta_1 + \beta_2 x_t$ and $\log(\phi_t) = \ldots$
Based on this figure we notice that once again the statistics based on residuals outperform the $R^2$ statistics since that the values of the $P^2$ and $P^2_{\beta\gamma}$ are considerably smaller than the values of the $R^2_{LR}$ statistic, in especial when the nonconstant dispersion is more severe (when $\lambda$ increases).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.png}
\caption{Misspecified model: $g(\mu_t) = \beta_1 + \beta_2 x_t$, $h(\phi_t) = \gamma_1 + \gamma_2 z_t$. True model: $g(\mu_t) = \beta_1 + x_t^{\beta_2}$, $h(\phi_t) = \gamma_1 + z_t^{\gamma_2}$, $x_t \sim U(0.3, 1.3)$, $z_t \sim U(0.5, 1.5)$, $\beta = (-1.1, 1.7)^\top$, $t = 1, \ldots, n$, $\mu \in (0.28, 0.61)$, $n = 400$.
\end{figure}

Figure 8 summarizes the predictive power measure performance with boxplots over the Monte Carlo replics. Here, we evaluated the empirical distribution of the $R^2_{FC}$ statistics under nonconstant dispersion for two models estimated correctly. The plots reveals evidences that the pseudo-$R^2$ measures are not a good statistics for model selection when the dispersion varying along the observations. It is clear that $P^2$ and $P^2_{\beta\gamma}$ measures become more powerful as the $\lambda$ increases.

## 4 Applications

In what follows we shall present an application based on real data.

**Application I:** The application relates to the distribution of natural gas for home usage (e.g., in water heaters, ovens and stoves) in São Paulo, Brazil. Such a distribution is based on two factors: the simultaneity factor ($F$) and the total nominal power of appliances that use natural gas, computed power $Q_{max}$. Using these factors one obtains an indicator of gas release in a given tubulation section, namely: $Q_p = F \times Q_{max}$. The simultaneity factor assumes values in $(0, 1)$, and can be interpreted as the probability of simultaneous appliances usage. Thus, based on $F$ the company that supplies the gas decides how much gas to supply to a given residential unit.
Here we present two important informations. First, we notice that by the beta regression model with link log-log for the mean submodel and link log for the dispersion submodel, the maximum likelihood parameter estimates are $\hat{\beta}_1 = -0.63$, $\hat{\beta}_2 = -0.31$, $\hat{\gamma}_1 = 3.81$ and $\hat{\gamma}_2 = 0.77$. Furthermore, the estimative of intensity of nonconstant dispersion is $\hat{\lambda} = 21.16$ (see (16)), such that $\hat{\phi}_{\text{max}} = 242.39$ and $\hat{\phi}_{\text{min}} = 11.45$. Selected among the candidates the best model in a predictive perspective, we still can use the PRESS statistic to identifying which observations are more difficult to predict. In this sense, we plot the individual components of $\text{PRESS}_{\beta\gamma}$ versus the observations index and we added a horizontal line at $3 \sum_{t=1}^n \text{PRESS}_{\beta\gamma_t}/n$ and singled out points that considerably exceeded this threshold.

The Table 4 displays two important informations. First, we notice that by the $R^2$ measures the models equally fits well. Second, the $P^2$ and $P^2_{\beta\gamma}$ measures lead to the same conclusions, selecting the beta regression model with link log-log for the mean submodel and link log for the dispersion submodel, as the best model to make prediction to the data on simultaneity factor. The maximum likelihood parameter estimates are $\hat{\beta}_1 = -0.63$, $\hat{\beta}_2 = -0.31$, $\hat{\gamma}_1 = 3.81$ and $\hat{\gamma}_2 = 0.77$. Furthermore, the estimative of intensity of nonconstant dispersion is $\hat{\lambda} = 21.16$ (see (16)), such that $\hat{\phi}_{\text{max}} = 242.39$ and $\hat{\phi}_{\text{min}} = 11.45$. Selected among the candidates the best model in a predictive perspective, we still can use the PRESS statistic to identifying which observations are more difficult to predict. In this sense, we plot the individual components of $\text{PRESS}_{\beta\gamma}$ versus the observations index and we added a horizontal line at $3 \sum_{t=1}^n \text{PRESS}_{\beta\gamma_t}/n$ and singled out points that considerably exceeded this threshold.

![Figure 8: True model: $g(\mu_t) = \beta_1 + x_t^{\beta_2}$, $h(\phi_t) = \gamma_1 + z_t^{\gamma_2}$, $x_t \sim U(0.3, 1.3)$, $z_t = x_t$, $\beta = (-1.1, 1.7)^T$, $\gamma = (2.3, 5.3)^T$, $n = 400$.](image-url)
Table 4: Values of the statistics from the candidate models. Data on simultaneity factor

| Candidate models                                                                 | submodel | Mean log(\(\mu_t/(1-\mu_t)\)) | Dispersion – – log(\(\phi_t\)) = log(\(\phi_t\)) = |
|----------------------------------------------------------------------------------|----------|---------------------------------|---------------------------------------------------|
|                                                                                  | \(\beta_1 + \beta_2 x_{11}\) | \(\beta_1 + \beta_2 x_{11}\) | \(\gamma_1 + \gamma_2 x_{11}\) | \(\gamma_1 + \gamma_2 x_{11}\) |
| **Covariates**                                                                  |          | 0.66                            | 0.70                                              | 0.68                                      | 0.68                                      | 0.88                                      |
| **Dispersion**                                                                  |          | 0.64                            | 0.42                                              | 0.70                                      | 0.68                                      | 0.70                                      | 0.74                                      | 0.74                                      |
| **PRESS**                                                                       |          | 0.65                            | 0.62                                              | 0.70                                      | 0.70                                      | 0.70                                      | 0.70                                      | 0.70                                      |
| **PRESS_c**                                                                     |          | 0.64                            | 0.39                                              | 0.68                                      | 0.68                                      | 0.68                                      | 0.68                                      | 0.87                                      |
| **PRESS_{\beta_1\gamma}**                                                       |          | 0.64                            | 0.42                                              | 0.70                                      | 0.70                                      | 0.70                                      | 0.70                                      | 0.70                                      |
| **PRESS_{\beta_2\gamma}**                                                       |          | 0.72                            | 0.70                                              | 0.74                                      | 0.74                                      | 0.74                                      | 0.74                                      | 0.74                                      |
| **PRESS_{LR}**                                                                  |          | 0.69                            | 0.65                                              | 0.70                                      | 0.70                                      | 0.70                                      | 0.70                                      | 0.70                                      |
| **PRESS_{LR_c}**                                                                |          | 0.69                            | 0.72                                              | 0.70                                      | 0.70                                      | 0.70                                      | 0.70                                      | 0.70                                      |
| **PRESS_{FC}**                                                                  |          | 0.67                            | 0.70                                              | 0.67                                      | 0.67                                      | 0.67                                      | 0.67                                      | 0.67                                      |
| **PRESS_{FC_c}**                                                                |          | 0.67                            | 0.70                                              | 0.67                                      | 0.67                                      | 0.67                                      | 0.67                                      | 0.67                                      |

Figure 9 shows that the cases 3, 11, 33 and 33 arise as the observations with more predictive difficulty and are worthy of further investigation.

**Application II:** The second application considers a nonlinear beta regression used to modeling the proportion of killed grasshopper \(y\) at an assays on a grasshopper Melanopus sanguinipes with the insecticide carbofuran and the synergist piperonyl butoxide. This model was proposed by Espinheira et al. (2017) after careful building the scheme of starting values to iterative process of maximum likelihood estimation and after a meticulous residual analysis. Our aim here is applies both predictive power statistics and goodness-of-fit statistics to confirm or not the choice of the model made by residual analysis. The covariates are the dose of the insecticide \(x_1\) and the dose of the synergist \(x_2\). The data can be found in McCullagh and Nelder (1989, p. 385). Additionally, \(y \in [0.04, 0.84]\), \(\mu_y = 0.4501\) and the median of the response is equal to 0.4967.
The model selected with its estimates and respective p-values is present in Espinheira et al. (2017) and is given by 
\[ \log(\mu_t / (1 - \mu_t)) = \beta_1 + \beta_2 \log(x_{t1} - \beta_3) + \beta_4 \frac{x_{t2}}{\sqrt{x_{t2} + \beta_5}} \] 
and \[ \sqrt{\phi_t} = \gamma_1 + \gamma_2 x_{t1} + \gamma_3 x_{t2}, \quad t = 1, \ldots, 15. \]
Now, by using residual analysis as diagnostic tools several linear beta regression models are compared with nonlinear model. After competition the nonlinear model: 
\[ \log(\mu_t / (1 - \mu_t)) = \beta_1 + \beta_2 \log(x_{t1} + 1.0) + \beta_3 x_{t2} \sqrt{\phi_t} = \gamma_1 + \gamma_2 x_{t1} + \gamma_3 x_{t2}, \quad t = 1, \ldots, 15 \] was selected. The estimatives of parameters are \( \hat{\beta}_1 = -4.25; \hat{\beta}_2 = 1.79; \hat{\beta}_3 = 0.04; \hat{\gamma}_1 = 1.19 \) and \( \hat{\gamma}_2 = 0.19 \). The values of the \( P^2 \) and \( R^2 \) measures for the two candidate models are present Table 5. Based on this table we can note that the nonlinear model outperforms the linear model under all statistics, that is, the nonlinear model has both better predictive power and better goodness-of-fit.

With aim in identifying observations for which to make prediction can be a hard task we plot values of PRESS statistic versus indices of the observations. In Figure 10 it is noteworthy how the case 14 is strongly singled out. In fact, this case was also singled out in plots of residual analysis made by Espinheira et al. (2017). However, the observation 14 is not an influential case, in sense of to affect inferential results. Besides, the choose model was capable to estimated well this case.

Thus, we confirm by the model selection measures that the beta nonlinear model proposed by Espinheira et al. (2017) is a suitable alternative to modeling of the data of insecticide carbocuran and the synergist piperonyl butoxide McCullagh and Nelder (1989, p. 385).

### Table 5: Values of the statistics from the candidate models. Data on insecticide.

| Candidate models | Linear Models | Nonlinear models |
|------------------|---------------|------------------|
| Mean \( \log(\mu_t / (1 - \mu_t)) = \beta_1 + \beta_2 \log(x_{t1} - \beta_3) \) | \( \log(\mu_t / (1 - \mu_t)) = \beta_1 + \beta_2 \log(x_{t1} + 1.0) + \beta_3 x_{t2} \sqrt{\phi_t} = \gamma_1 + \gamma_2 x_{t1} + \gamma_3 x_{t2}, \) | \( \gamma_1 + \gamma_2 x_{t1} + \gamma_3 x_{t2} + \gamma_4 (x_{t1} x_{t2}) \) |
| Dispersion | | |
| submodel \( \sqrt{\phi_t} = \gamma_1 + \gamma_2 x_{t1} \) | \( \sqrt{\phi_t} = \gamma_1 + \gamma_2 x_{t1} + \gamma_3 x_{t2} + \gamma_4 (x_{t1} x_{t2}) \) |
| \( P^2 \) | 0.89 | 0.99 |
| \( P_{c}^2 \) | 0.85 | 0.99 |
| \( P_{v}^2 \) | 0.89 | 0.99 |
| \( P_{\text{vcc}}^2 \) | 0.86 | 0.99 |
| \( R^2_{LR} \) | 0.83 | 0.99 |
| \( R_{\text{LRc}}^2 \) | 0.70 | 0.99 |
| \( R_{\text{FC}}^2 \) | 0.79 | 0.97 |
| \( R_{\text{FCc}}^2 \) | 0.71 | 0.94 |

**Application III:** In the latter application we will use the dataset about available chlorine fraction after weeks of manufacturing from an investigation performed at Proctor & Gamble. A certain product must have a fraction of available chlorine equal to 0.50 at the time of manufacturing. It is known that chlorine fraction of the product decays with time. Eight weeks after the production, before the product is consumed, in theory there is a decline to a level 0.49.

The theory related to the problem indicates that the available chlorine fraction \( y \) decays according to a nonlinear function of the number of weeks \( x \) after fabrication of the product and unknown parameters (Draper and Smith, 1981 p. 276), given by

\[ \eta_t = \beta_1 + (0.49 - \beta_1) \exp\{\beta_2(x_t - 8)\}. \]
The level 0.49 depends on several uncontrolled factors, as for example warehousing environments or handling facilities. Thus, the predictions based on theoretical model can be not reliable.

Cartons of the product were analyzed over a period aiming answer some questions like as: “When should warehouse material be scrapped?” or “When should store stocks be replaced?” According to knowledgeable chemists an equilibrium asymptotic level of available chlorine should be expected somewhere close to 0.30.

From predictor based on \((12)\) we can note that when \(x = 8\) the nonlinear model provides a true level for the available chlorine fraction (no error), wherein \(\eta = 0.49\). We consider a new logit nonlinear beta regression model. We replaced the deterministic value 0.49 by an additional parameter at the predictor. Thus, the new nonlinear predictor for mean submodel is given by
\[
\eta_t = \beta_1 + (\beta_3 - \beta_1)\exp\{ -\beta_2(x_t - 8) \},
\]
\(t = 1, \ldots, 42\). Here the available chlorine fraction ranged from 0.38 to 0.49, being the mean and median are approximately equal 0.42. We investigated some competitive models. Table 6 the results of final candidates models and its statistics. The findings reveals that the model log(\(\mu_t/(1 - \mu_t)\)) = \(\beta_1 + (\beta_3 - \beta_1)\exp\{ \beta_2(x_{t1} - 8) \} \) and log(\(\phi_t\)) = \(\gamma_1 + \gamma_2\log(x_{t1}) + \exp\{ \gamma_3(x_{t1} - 8) \}\) is the best performer in sense that it displays the higher statistics values. To estimate this model was necessary to build a starting values procedure for log-likelihood maximization as proposed by Espinheira et al. (2017). Since that we have more parameters than covariates firstly we used the theoretical information about the asymptotic level and found a initial guess to \(\beta_1\) equal to 0.30. Thus, based on equation we took some values to \(y\) and \(x_1\) and found a initial guess to \(\beta_2, \beta_2^{(0)} = 0.02\). Then we carried out the scheme of starting values to be used in nonlinear beta regression maximum likelihood estimation (Espinheira et al., 2017). The parameters estimatives are \(\hat{\beta}_1 = -0.45963\) \(\hat{\beta}_2 = -0.04166\) \(\hat{\beta}_3 = 0.09479\) \(\hat{\gamma}_1 = 13.18335\) \(\hat{\gamma}_2 = -0.05413\) \(\hat{\gamma}_3 = 2.63158\). It is importance emphasize that the \(\beta_2\) estimative conduce to a level of chlorine fraction equal to 0.4896 \(\approx 0.49\) that for this dataset confirm the theory that there is a decline to a level 0.49.

5 Conclusion

In this paper we develop the \(P^2\) and \(P^2_{\beta_3}\) measures based on two versions of PRESS statistics for the class of beta regression models. The \(P^2\) coefficient consider the PRESS statistic based on ordinary residual
Table 6: Values of the statistics from the candidate models. Data on chlorine fraction

| Candidate models | \(\eta_{1t}\) | \(\eta_{2t}\) | \(P^2\) | \(P^2_c\) | \(P^2_{\beta_\gamma}\) | \(P^2_{\beta_\gamma c}\) | \(R^2_{LR}\) | \(R^2_{FC}\) |
|------------------|----------------|----------------|--------|--------|--------|--------|--------|--------|
| log(\(\mu_t / (1 - \mu_t)\)) = \(\beta_1 + (\beta_3 - \beta_1)\exp(\beta_2(x_{1t} - 8))\) | \(-\log(-\log(\mu_t)) = \beta_1 + (\beta_3 - \beta_1)\exp(\beta_2(x_{1t} - 8))\) | \(-\log(-\log(\mu_t)) = \beta_1 + (\beta_3 - \beta_1)\exp(\beta_2(x_{1t} - 8))\) | \(\log(\phi_t) = \gamma_1 + \gamma_2 log(x_{1t}) + \exp(\gamma_3(x_{1t} - 8))\) | \(\log(\phi_t) = \gamma_1 + \gamma_2 log(x_{1t}) + \exp(\gamma_3(x_{1t} - 8))\) | \(0.88\) | \(0.87\) | \(0.88\) | \(0.87\) |
Appendix

Fisher’s scoring iterative algorithm: In what follows we shall present the score function and Fisher’s information for \( \beta \) and \( \gamma \) in nonlinear beta regression models (Simas et al., 2010). The log-likelihood function for model (2) is given by \( \ell(\beta, \gamma) = \sum_{t=1}^{n} \ell_t(\mu_t, \phi_t) \), and \( \ell_t(\mu_t, \phi_t) = \log \Gamma(\phi_t) - \log \Gamma(\mu_t \phi_t) - \log \Gamma((1 - \mu_t) \phi_t) + (\mu_t \phi_t - 1) \log y_t + \{(1 - \mu_t) \phi_t - 1\} \log (1 - y_t) \). The score function for \( \beta \) is

\[
U_\beta(\beta, \gamma) = J^T_1 \Phi T(y^* - \mu^*),
\]

where \( J_1 = \partial \eta / \partial \beta \) (an \( n \times k \) matrix), \( \Phi = \text{diag}\{\phi_1, \ldots, \phi_n\} \), the \( t \)th elements of \( y^* \) and \( \mu^* \) being given in (7). Also, \( T = \text{diag}\{1/g'(\mu_1), \ldots, 1/g'(\mu_n)\} \). The score function for \( \gamma \) can be written as \( U_\gamma(\beta, \gamma) = J^T_2 Ha \), where \( J_2 = \partial \eta_2 / \partial \gamma \) (an \( n \times q \) matrix), \( a_t \) is given in (9) and \( H = \text{diag}\{1/h'(\phi_1), \ldots, 1/h'(\phi_n)\} \). The components of Fisher’s information matrix are

\[
K_{\beta\beta} = J^T_1 \Phi W J^T_1, \quad K_{\beta\gamma} = J^T_1 CT H J^T_2 \quad \text{and} \quad K_{\gamma\gamma} = J^T_2 DJ^T_2. \tag{14}
\]

Here, \( W = \text{diag}\{w_1, \ldots, w_n\} \), where

\[
w_t = \phi_t v_t [1/(g'(\mu_t))^2] \quad \text{and} \quad v_t = \{\psi'(\mu_t \phi_t) + \psi'((1 - \mu_t) \phi_t)\}. \tag{15}
\]

Also, \( C = \text{diag}\{c_1, \ldots, c_n\}; \ c_t = \phi_t \{\psi'(\mu_t \phi_t) \mu_t - \psi'((1 - \mu_t) \phi_t)(1 - \mu_t)\}; \ d_t = \xi_t/(h'(\mu_t))^2 \) and \( \xi_t = \{\psi'(\mu_t \phi_t) \mu_t + \psi'((1 - \mu_t) \phi_t)(1 - \mu_t)^2 - \psi'(\phi_t)\} \). To propose PRESS statistics for a beta regression we shall based on Fisher iterative maximum likelihood scheme and weighted least square regressions. Fisher’s scoring iterative scheme used for estimating \( \beta \), both to linear and nonlinear regression model, can be written as

\[
\beta^{(m+1)} = \beta^{(m)} + (K_{\beta\beta}^{(m)})^{-1} U_\beta^{(m)}(\beta). \tag{16}
\]

Where \( m = 0, 1, 2, \ldots \) are the iterations which are carried out until convergence. The convergence happens when the difference \( |\beta^{(m+1)} - \beta^{(m)}| \) is less than a small, previously specified constant.

From (13), (14) and (16) it follows that the \( m \)th scoring iteration for \( \beta \), in the class of linear and nonlinear regression model, can be written as \( \beta^{(m+1)} = \beta^{(m)} + (J^T_1 \Phi W^{(m)} J_1)^{-1} J^T_1 \Phi W^{(m)} T^{(m)} (y^* - \mu^*(m)) \), where the \( t \)th elements of the vectors \( y^* \) and \( \mu^* \) are given in (5). It is possible rewrite this equation in terms of weighted least squares estimator as \( \beta^{(m+1)} = (J^T_1 \Phi W^{(m)} J_1)^{-1} \Phi W^{(m)} u^{(m)}(\mu^*(m)) \). Here, \( u^{(m)}(\mu) = J_1 \beta^{(m)} + W^{-1} T^{(m)} (y^* - \mu^*(m)). \) Upon convergence,

\[
\hat{\beta} = (J^T_1 \hat{\Phi} W J_1)^{-1} \hat{\Phi} J^T_1 \hat{W} u_1 \quad \text{where} \quad u_1 = J_1 \hat{\beta} + \hat{W}^{-1} \hat{T}(y^* - \hat{\mu}^*). \tag{17}
\]

Here, \( \hat{W}, \hat{T}, \hat{H} \) and \( \hat{D} \) are the matrices \( W, T, H \) and \( D \), respectively, evaluated at the maximum likelihood estimates. We note that \( \hat{\beta} \) in (17) can be viewed as the least squares estimates of \( \beta \) obtained by regressing \( \hat{\Phi}_{1/2} \hat{W}_{1/2} u_1 \) on \( \hat{\Phi}_{1/2} \hat{W}_{1/2} J_1 \).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Second International Symposium on Information Theory (Tsukhakador, 1971), pp. 267–281. Budapest: Akadémiai Kiadó.

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

Bartoli, A. (2009). On computing the prediction sum of squares statistic in linear least squares problems with multiple parameter or measurement sets. International Journal of Computer Vision 85(2), 133–142.

Bayer, F. M. and F. Cribari-Neto (2017). Model selection criteria in beta regression with varying dispersion. Communications in Statistics, Simulation and Computation 46, 720–746.

Brascum, A. J., E. O. Johnson, and M. C. Thurmond (2007). Bayesian beta regression: applications to household expenditures and genetic distances between foot-and-mouth disease viruses. Australian and New Zealand Journal of Statistics 49(3), 287–301.

Cepeda-Cuervo, E. and D. Gamerman (2005). Bayesian methodology for modeling parameters in the two parameter exponential family. Estadística 57, 93–105.
Chien, L.-C. (2011). Diagnostic plots in beta-regression models. *Journal of Applied Statistics* 38(8), 1607–1622.

Cook, R. D. and S. Weisberg (1982). *Residuals and Influence in Regression*. Chapman and Hall.

Cribari-Neto, F. and A. Zeileis (2010). Beta regression in R. *Journal of Statistical Software* 34(2), 1–24.

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

Espinheira, P. L., S. Ferrari, and F. Cribari-Neto (2014). Bootstrap prediction intervals in beta regressions. *Computational Statistics* 29(5), 1263–1277.

Espinheira, P. L., S. L. P. Ferrari, and F. Cribari-Neto (2008b). Influence diagnostics in beta regression. *Computational Statistics and Data Analysis* 52, 4417–4431.

Espinheira, P. L., E. G. Santos, and F. Cribari-Neto (2017). On nonlinear beta regression residuals. *Biometrical Journal* n/a(n/a), n/a–n/a.

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

Figuero-Zúñiga, J. I., R. B. Arellano-Valle, and S. L. Ferrari (2013). Mixed beta regression: A bayesian perspective. *Computational Statistics and Data Analysis* 61, 137–147.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (2 ed.). London: Chapman and Hall.

Mediavilla, F., L. F, and V. A. Shah (2008). A comparison of the coefficient of predictive power, the coefficient of determination and aic for linear regression. In K. JE (Ed.), *Decision Sciences Institute, Atlanta*, pp. 1261–1266.

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

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

Palmer, P. B. and D. G. O’Connell (2009, sep). Regression analysis for prediction: Understanding the process. *Cardiopulmonary Physical Therapy Journal* 20(3), 23–26.

Pregibon, D. (1981, 07). Logistic regression diagnostics. *The Annals of Statistics* 9(4), 705–724.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6(2), 461–464.

Simas, A. B., W. Barreto-Souza, and A. V. Rocha (2010). Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis* 54(2), 348–366.

Smithson, M. and J. Verkuilen (2006). A Better Lemon Squeezer? Maximum-Likelihood Regression With Beta-Distributed Dependent Variables. *Psychological Methods* 11(1), 54–71.

Spiess, A.-N. and N. Neumeyer (2010). An evaluation of $r^2$ as an inadequate measure for nonlinear models in pharmacological and biochemical research: a monte carlo approach. *BMC Pharmacology* 10(1), 6.

Zerbinatti, L. (2008). Predição de fator de simultaneidade através de modelos de regressão para proporções contínuas. Msc thesis, University of São Paulo.