

# Metabolisable energy prediction in energy feedstuffs and evaluation of the stepwise validation procedure using bootstrapping<sup>1</sup>

Predição da energia metabolizável de ingredientes energéticos e avaliação da validação do *stepwise* utilizando *bootstrap*

Newton Tavares Escocard de Oliveira<sup>2\*</sup>, Paulo Cesar Pozza<sup>3</sup>, Leandro Dalcin Castilha<sup>3</sup>, Tiago Junior Pasqueti<sup>4</sup>, and Carolina Natali Langer<sup>2</sup>

**ABSTRACT** - The use of predicted values of apparent metabolisable energy (AME), obtained from regression equations, can be useful for both research institutions and nutrition industries. However, there is a need to validate independent samples to ensure that the predicted equation for AME is reliable. In this study, data was collected in order to estimate the prediction equations of corn, sorghum and wheat bran for pig feed, based on the chemical composition, in addition to evaluating the validity of the stepwise selection procedure regressive method of non-parametric bootstrap resampling. Data from metabolism trials in pigs and the chemical composition of feedstuffs was collected from both Brazilian and international literature, expressed as dry matter. After the residue analysis, five models of multiple linear regression were adjusted to randomly generate 1000 bootstrap samples of equal size from the database via meta-analysis. The five estimated models were adjusted for all bootstrapped samples using the stepwise method. The highest percentage significance for regressor (PSR) value was observed for digestible energy (100%) in the AME<sub>1</sub> model, and gross energy (95.7%) in the AME<sub>2</sub> model, indicating high correlation of the regressive model with AME. The regressors selected for AME<sub>4</sub> and AME<sub>5</sub> resulted in a PSR of greater than 50%, and were validated for estimating the AME of pig feed. However, the percentage of joint occurrence of regressor models showed low reliability, with values between 2.6% (AME<sub>2</sub>) and 23.4% (AME<sub>4</sub>), suggesting that the stepwise procedure was invalid.

**Key words:** Chemical composition. Corn. Meta-analysis. Pigs. Regression models.

**RESUMO** - A utilização de valores preditos de energia metabolizável aparente (EMA), obtidos a partir de equações de regressão, pode ser útil para instituições de pesquisa e indústrias de nutrição. Contudo, há a necessidade de validação em amostras independentes para assegurar que a equação de predição da EMA seja confiável. Neste estudo, foram coletados dados para estimar equações de predição da EMA do milho, sorgo e farelo de trigo para suínos em função de composição química, além de avaliar a validade do procedimento *stepwise* de seleção de regressoras pelo método de reamostragem *bootstrap* não paramétrico. Dados de ensaios de metabolismo com suínos e de composição química dos alimentos foram coletados na literatura científica nacional e internacional e expressos na matéria seca. Após a análise de resíduos, foram ajustados cinco modelos de regressão linear múltipla e geradas aleatoriamente 1000 amostras *bootstrap* de mesmo tamanho a partir do banco de dados formado via meta-análise. Em todas as amostras *bootstrap* foram ajustados os cinco modelos estimados, utilizando-se do *stepwise*. Os maiores percentuais de significância por regressora (PSR) ocorreram para a energia digestível (100%) no modelo EMA<sub>1</sub>, e para a energia bruta (95,7%) no modelo EMA<sub>2</sub>, indicando alta correlação das regressoras com a EMA. As regressoras selecionadas nos EMA<sub>4</sub> e EMA<sub>5</sub> apresentaram PSR maior que 50% e são válidas para estimar a EMA de alimentos energéticos para suínos. Entretanto, os percentuais de ocorrência conjunta de regressoras dos modelos apresentaram baixa confiabilidade, com valores entre 2,6% (EMA<sub>2</sub>) e 23,4% (EMA<sub>4</sub>), invalidando o procedimento *stepwise*.

**Palavras-chave:** Composição química. Meta-análise. Milho. Modelos de regressão. Suínos.

DOI: 10.5935/1806-6690.20190016

\*Author for correspondence

Received for publication 02/09/2016; approved on 05/02/2018

<sup>1</sup>Parte da Dissertação de Mestrado em Zootecnia do último autor, apresentada na Universidade Estadual do Oeste do Paraná/UNIOESTE

<sup>2</sup>Curso de Zootecnia, Centro de Ciências Agrárias, Universidade Estadual do Oeste do Paraná/UNIOESTE, Rua, Pernambuco 1777, Marechal Cândido Rondon-PR, Brasil, 85.960-000, newtonescocard@hotmail.com (ORCID ID 0000-0003-3347-9399), carolina\_langer@hotmail.com

<sup>3</sup>Departamento de Zootecnia, Universidade Estadual de Maringá/UEM, Av. Colombo 5790, Jardim Universitário, Maringá-PR, Brasil, 87.020-900, pcpozza@uem.br, leandrocastilha@hotmail.com

<sup>4</sup>Curso de Zootecnia, Universidade Estadual do Mato Grosso do Sul/UEMS, Unidade de Aquidauana, Rodovia Graziela Maciel Barroso, Zona Rural, Aquidauana-MS, Brasil, 79.200-000, pasquettizoo@yahoo.com.br

## INTRODUCTION

Knowledge of the apparent metabolisable energy (AME) of feedstuffs is essential for formulating balanced rations (SAKOMURA; ROSTAGNO, 2016), as nutrient requirements are expressed in terms of energy levels of feed, which affect feed intake and pig performance (ROSTAGNO *et al.*, 2011).

The direct determination of AME values for pig feed is time consuming, labour intensive and costly, as it involves performing metabolic tests. A quick alternative, which is also practical and economical, is the use of regression equations to predict AME values. These equations use the chemical composition of feedstuffs, which are routinely obtained in laboratories (PELIZZERI *et al.*, 2013; POZZA *et al.*, 2008), and can be used as a tool to adjust AME values for variability in chemical composition data. Furthermore, the values obtained are more suitable than the use of values of feed composition tables (SAKOMURA; ROSTAGNO, 2016).

However, the presence of outliers, due to experimental errors or errors in the determination of chemical composition of feed, combined with model adjustment strategies, can change the parameter estimates of the equations from the ordinary least squares method, resulting in variation of the predicted AME values for pig feed. This can compromise the validity of the model, related to the stability and reasonableness of the regression coefficients, and therefore, the usefulness of the model for giving accurate predictions for new data samples (CASTILHO *et al.*, 2015; OREDEIN; OLATAYO; LOYINMI, 2011).

The strategy for adjusting the regression models of AME is based on the chemical composition of feed energy. These results are obtained for different experimental conditions, allowing a greater range of AME values (LOVATTO *et al.*, 2007; MARIANO *et al.*, 2012; NASCIMENTO *et al.*, 2009, 2011), and therefore, enabling determination of the most representative adjustments to fit regression models. These models can be tested using sample data taken from different studies in the literature.

Some researchers have adjusted the equations used for predicting the AME of pig feed for the chemical composition (CASTILHA *et al.*, 2011; MORGAN *et al.*, 1987; NOBLET; PEREZ, 1993; PELIZZERI *et al.*, 2013). A great variety of regressor sets for predictive AME models have been reported in the literature. Some studies have evaluated the validity of regressor selection procedures on independent data samples in order to verify the predictive ability of the models (CASTILHO *et al.*, 2015; OREDEIN; OLATAYO; LOYINMI, 2011).

The non-parametric bootstrap method can be used to simulate real situations, in addition to validating the regressor selection procedures (SCALON; FREIRE; CUNHA, 1998). In this method, several samples are randomly taken from a representative sample of the original population, with replacement of each sample, allowing the frequency of significance of the regressor and the joint frequency of occurrence of the regressor to be obtained relative to the total sample.

The objective of the present study was to adjust the prediction equations for the AME of corn, sorghum and wheat bran from data on the chemical composition and energy for pigs, available in the Brazilian and international scientific literature, in addition to evaluating the validity of the stepwise selection procedure of regressive models that are adjustable using the non-parametric bootstrap method.

## MATERIALS AND METHODS

Data collected from metabolism tests conducted on pigs weighing between 7.0 and 75.0 kg was obtained in the 20 Brazilian and 22 international scientific literatures published in the period from 1965 to 2011, aimed at forming a database containing the chemical composition and energy of corn, sorghum and wheat bran.

After the literature search and tabulation of the data, detailed screening was performed to exclude any conflicting data that could result in biased estimates of parameters. After the exclusion of incomplete and conflicting data, 142 records remained, of which 69 were Brazilian and 73 international studies. Within the Brazilian data, 46 records detailed the chemical and energy composition of corn, 11 for sorghum and 12 for wheat bran. In the international literature there were 41, 17 and 15 studies that provided data for corn, sorghum and wheat bran, respectively.

All chemical composition and energy values recorded for the natural matter were converted to dry matter, in order to standardise and subsequently predict the AME values.

The influential observations were initially assessed to determine whether they met the assumptions of normality, homogeneity and linearity of residuals using multiple linear regression models, after which the standardised residual student (RStudent) analysis was performed. The criteria for the identification of outliers was based on the normal distribution curve. RStudent values higher than three standard deviations from the mean were considered to be influential, and consequently removed from the database.

The prediction of AME values, based on the reported chemical and energy composition of feedstuffs, was performed by the adjustment of multiple linear regression models (COELHO-BARROS *et al.*, 2008). The methods used were the ordinary least squares and stepwise procedure regressor selection methods, including dummy variables (binary values) for the “source” (Brazilian or international research) and “feedstuffs” (corn, sorghum or wheat bran) regressors.

The use of the “source” and “feedstuffs” indicator variables allowed for the models to be classified according to the selected dummy variables. For the “source” dummy (SD) variables, the coding for the Brazilian studies occurred when the variable received a zero (SD = 0), while the coding SD = 1 identified international studies. For the “feedstuffs” dummy (FD) variables, corn was encoded as two classes of variables (FD<sub>1</sub> and FD<sub>2</sub>), which assumed that the value was zero. The effect of sorghum was expressed when FD<sub>1</sub> = 0 and FD<sub>2</sub> = 1, and the effect of wheat bran was expressed when FD<sub>1</sub> = 1 and FD<sub>2</sub> = 0.

A total of five models of multiple linear regression were fitted to the AME data. The complete model (AME<sub>1</sub>) was represented by the additive effects of the intercept, the simple effects of the apparent digestible energy (ADE), gross energy (GE), crude fibre (CF), ether extract (EE), crude protein (CP), ash, source dummy (SD), feedstuffs dummy 1 (FD<sub>1</sub>) and feedstuffs dummy 2 (FD<sub>2</sub>), in addition to interactions of the dummies with the regressor chemical and energy composition.

The second fitted model (AME<sub>2</sub>) represented the complete model without the inclusion of ADE and interactions between ADE and classification variables. The third model (AME<sub>3</sub>) was fitted for the complete model without the inclusion of ADE or GE, or their interactions with the dummy variables. The fourth model (AME<sub>4</sub>) was fitted from the complete model after removing ADE, GE and CF, and their interactions with dummies. The fifth and final model (AME<sub>5</sub>) corresponded to the complete model without ADE, GE, CF and EE, and without their interactions with the indicator variables.

The significance of each parameter was assessed by a partial t-test to determine whether the null hypothesis ( $\beta_1 = 0$ ) and the occurrence of multicollinearity between the regressors was verified by the observation of inflation factor variance associated with each regressor. The goodness-of-fit of the regression models to the AME data was evaluated by determining the coefficient of determination (R<sup>2</sup>). The precision of the estimates was assessed from their respective standard deviations.

Using the original sample obtained from the literature, we randomly generated 1000 bootstrap samples of the same size, obtained by resampling and replacement

of each sample (COELHO-BARROS *et al.*, 2008), in order to evaluate the stepwise procedure regressor selection based on their respective selection frequencies.

For all bootstrapped samples, the five models described were fitted, using the stepwise procedure and the ordinary least squares method for estimating the parameters. The percentage of significance for regressor (PSR) and the percentage of joint occurrence of regressor (PJOR) models were verified for the 1000 bootstrap samples.

The bootstrap estimates of j<sup>th</sup> regression coefficient (b<sub>j</sub><sup>\*</sup>) were arranged in ascending order, after which the confidence intervals of the parameters were estimated by the percentile bootstrap method. In this method, 500 different bootstrap samples of the same size, obtained by resampling and replacement of each sample, were used to assess the significance of the bias (b<sub>j</sub> - b<sub>j</sub><sup>\*</sup>) of each coefficient and the significance of the regression coefficients.

A significance level of 0.05 was used for all statistical tests. All statistical analyses were performed using R software (R CORE TEAM, 2013).

## RESULTS AND DISCUSSION

In order to suggest high reliability of the estimates, the coefficients of determination (R<sup>2</sup>) obtained needed to be high and show that the estimated equations explained more than 90% of the variation in the AME data as a function of the chemical composition of corn, sorghum and wheat bran, as obtained from the Brazilian and international literature data (Table 1).

Among the estimation equations, it was observed a higher R<sup>2</sup> value when the ADE was included as a predictor variable in the model (Table 1). This suggests that the ADE explained most of the variation in AME, due to its high correlation with the response variable ( $r_{xy} = 0.9969$ ) compared to the other regressors. Noblet and Perez (1993) also observed higher R<sup>2</sup> values from equations that included the ADE as a predictor.

Despite having the highest R<sup>2</sup> value (99.67%) compared to the other equations (Table 1), the AME<sub>1</sub> model is not the most suitable in practice. This is because faeces collection is required to determine the ADE in metabolic experiments, making its use costly with little applicability (PELIZZERI *et al.*, 2013). However, the use of prediction models that include ADE as the regressor allow the energy loss in the faeces to be estimated, and therefore, do not require the collection and measurement of total excreted urine.

**Table 1-** Multiple linear regression models of apparent metabolisable energy (AME), expressed in kcal kg<sup>-1</sup>, based on the chemical and energy composition of corn, sorghum and wheat bran in pig feed, estimated from Brazilian and international data

Model <sup>1</sup>	Regressor	Parameter estimate	Standard deviation	R <sup>2</sup>
AME <sub>1</sub> (n = 96)	Intercept <sup>2</sup>	2.55	62.00	0.9967
	ADE	0.969	0.01	
	SD	36.78	5.53	
	FD <sub>1</sub> *CP	-4.22	0.87	
AME <sub>2</sub> (n = 106)	Intercept	2306.26	396.51	0.9239
	GE	0.39	0.09	
	CF	-123.95	4.50	
	EE	33.05	13.87	
	SD*ash	-49.63	8.29	
AME <sub>3</sub> (n = 117)	Intercept	3675.39	56.52	0.9106
	EE	50.02	14.50	
	SD*CF	-25.62	5.20	
	FD <sub>1</sub> *CF	-91.23	3.28	
AME <sub>4</sub> (n = 129)	Intercept	3824.44	60.05	0.9023
	ash	-105.29	18.84	
	EE	45.01	13.10	
	FD <sub>1</sub> *CP	-37.26	3.68	
AME <sub>5</sub> (n = 124)	Intercept	3982.99	28.40	0.9247
	ash	-79.97	17.17	
	FD <sub>1</sub> *CP	-44.78	3.52	
	FD <sub>2</sub> *ash	-43.42	16.27	

<sup>1</sup>ADE - Apparent digestible energy; CP - Crude protein; GE - Gross energy; CF - Crude fibre; EE - Ether extract; SD - Source dummy; FD<sub>1</sub> - Feedstuff dummy 1; FD<sub>2</sub> - Feedstuff dummy 2; <sup>2</sup>Probability = 0,9673

Morgan *et al.* (1987) estimated equations for predicting general feedstuff AME using the same regressors, however, the goodness-of-fit of the models to the data was much lower than that observed in our study ( $0.39 \leq R^2 \leq 0.43$ ). In addition to the R<sup>2</sup> value, it is also essential to determine the significant predictors of models, because equations that contain up to four chemical composition variables are easier, require less time and are more cost effective (POZZA *et al.*, 2008).

After replacement of the binary values in the AME<sub>1</sub>, AME<sub>2</sub>, AME<sub>3</sub>, AME<sub>4</sub> and AME<sub>5</sub> models, the daughter equations were derived. Except for the AME<sub>5</sub> model, the results showed that the corn and sorghum data can be combined into a single database, because the estimates for corn and sorghum, provided by the AME<sub>1</sub>, AME<sub>2</sub>, AME<sub>3</sub> and AME<sub>4</sub> models, were the same (Table 2).

Unlike corn and sorghum, the AME<sub>1</sub>, AME<sub>4</sub> and AME<sub>5</sub> estimation models for wheat bran showed that CP was an important independent variable for reducing

the estimated AME (Table 2). This can be observed from the average CP of corn, sorghum and wheat bran reported in the literature, where the protein content of wheat bran (CP = 15.62%) is much higher than for corn (CP = 7.88%) and sorghum (CP = 8.97%) (ROSTAGNO *et al.*, 2011). The significant interaction between FD<sub>1</sub> and CP in the AME<sub>1</sub>, AME<sub>4</sub> and AME<sub>5</sub> parent models (Table 1) suggests that the effect of CP was important in order to distinguish wheat bran from corn and sorghum.

In models that simultaneously showed the regressive CP and ADE, it was observed that CP reduced the estimated AME of the corn (CASTILLA *et al.*, 2011; LANGER *et al.*, 2013) and diets (NOBLET; PEREZ, 1993) for pigs. However, when ADE was not included in the models (NOBLET; PEREZ, 1993), the CP was positively correlated with AME. Morgan *et al.* (1987) observed a positive influence of CP in the estimation of AME for food when ADE was not included. When the ADE was included in the model, the CP estimates was negative as it needed to be adjusted for urinary nitrogen loss.

**Table 2** - Regression models of apparent metabolisable energy (AME), expressed in kcal kg<sup>-1</sup>, based on the chemical and energy composition of corn, sorghum and wheat bran in pig feed, expressed as dry matter

Feedstuff	Data source	Regression model <sup>1</sup>
Corn, sorghum	B	$AME_{1A} = 2.547 + 0.969ADE$
Corn, sorghum	I	$AME_{1B} = 39.33 + 0.969ADE$
Wheat bran	B	$AME_{1C} = 2.547 + 0.969ADE - 4.217CP$
Wheat bran	I	$AME_{1D} = 39.33 + 0.969ADE - 4.217CP$
Corn, sorghum, wheat bran	B	$AME_{2A} = 2306.26 + 0.39GE - 123.95CF + 33.05EE$
Corn, sorghum, wheat bran	I	$AME_{2B} = 2306.26 + 0.39GE - 123.95CF + 33.05EE - 49.63ash$
Corn, sorghum	B	$AME_{3A} = 3675.39 + 50.02EE$
Corn, sorghum	I	$AME_{3B} = 3675.39 + 50.02EE - 25.62CF$
Wheat bran	B	$AME_{3C} = 3675.39 + 50.02EE - 91.23CF$
Wheat bran	I	$AME_{3D} = 3675.39 + 50.02EE - 116.85CF$
Corn, sorghum	B, I	$AME_{4A} = 3824.44 - 105.29ash + 45.01EE$
Wheat bran	B, I	$AME_{4B} = 3824.44 - 105.29ash + 45.01EE - 37.26CP$
Corn	B, I	$AME_{5A} = 3982.99 - 79.97ash$
Sorghum	B, I	$AME_{5B} = 3982.99 - 123.39ash$
Wheat bran	B, I	$AME_{5C} = 3982.99 - 79.97ash - 44.78CP$

<sup>1</sup>ADE - Apparent digestible energy; CP - Crude protein; GE - Gross energy; CF - Crude fibre; EE - Ether extract; B - Brazilian data; I - International data

The estimated values of AME based on the international data were higher than that of the Brazilian data, regardless of the feedstuff source, as the estimates of the intercept were higher (Table 2). Differences in soil fertility, climate, management and genetics, among others, can affect the nutritional quality of feedstuff, and therefore, the use of nutrients by the pigs (NATIONAL RESEARCH COUNCIL, 2012).

The prediction of AME based on the chemical composition and dummies, without the inclusion of ADE (AME<sub>2</sub> model), showed significance for the GE, CF, EE and SD\*ash regressors (Table 1). Replacement of the binary values in the AME<sub>2</sub> model generated two equations, separated by feedstuff source, for the corn, sorghum and wheat bran from the Brazilian or international data (Table 2). These two equations showed no numerical difference in the regression coefficients associated with the GE, CF and EE for corn, sorghum and wheat bran. However, when the AME was estimated using international data, there was a negative effect of ash (Table 2). This indicated that the AME estimate of international foods would be less than the AME of Brazilian foods, when ADE is not included in the model.

The signs of these predictors were similar to the signs of regressors observed by Noblet and Perez (1993) in models including GE. These authors reported a positive correlation between GE and AME, and negative correlations for ash and fibre with AME, with a

$R^2 = 0.85$  obtained by the models. The use of the AME<sub>2</sub> model can be a viable alternative as there is no need to conduct experiments on animals. However, the GE analysis requires the use of a bomb calorimeter.

Replacement of binary values in the AME<sub>3</sub> model yielded two equations for corn and sorghum and two daughter equations for wheat bran, which were dependent on the feedstuff source data reported in the literature (Table 2). Again, we were able to combine the corn and sorghum data in a single database, as there were separate equations for the Brazilian and international data. This can be observed in Table 1, where the parent equation included the interaction between the SD and CF (SD\*CF).

The independent variable for chemical composition for differentiating the origin of corn and sorghum was CF (Table 2). The CF had a negative effect on AME for the international data, indicating that the AME estimate for corn and sorghum from international data was less than the estimated value from the Brazilian data. For wheat bran, the AME estimate for the international data was lower than that of the Brazilian data, due to the lower regression coefficient associated with the CF for international wheat bran data.

Similarly, CF was the component that differentiated wheat bran (WB) from the corn and sorghum cultivars, as the AME estimates obtained for WB would be lower than corn and sorghum cultivars, regardless of the source

(Brazilian or international) (Table 2). The differentiation of WB from the corn and sorghum cultivars by including the CF regressor may be due to differences in the amount of CF in these food sources. The average CF (%), as obtained from meta-analysis data, was 10.13% for WB and 2.52% for corn and sorghum. Noblet and Perez (1993) also reported a negative effect of fibre and a positive effect of ether extract on the prediction equations of AME diets for pigs.

For the AME<sub>4</sub> model, there was a distinction between WB and corn and sorghum, characterised by the interaction between FD<sub>1</sub> and CP in the parent equation (Table 1). Replacing the binary values resulted in two independent origin equations, one for corn and sorghum and another for WB (Table 2). This suggests that model adjustment with the inclusion of ash, EE and CP, without CF, was not sufficient to discriminate between the feedstuffs in the literature, due to the similarities in digestibility and metabolism of ash, EE and CP, or from difference in CF content between different ingredients.

Similar to that observed for the AME<sub>1</sub> model, CP was the nutrient that differentiated the corn and sorghum cultivars from WB in the AME<sub>4</sub> model, showing a negative effect on the AME of WB. This is possibly due to differences in the CP of WB compared to corn and sorghum cultivars. If the protein presents low quality, or if it is present in excess in the feed, there will be an increase in the nitrogen load of the animal, resulting in increased energy expenditure required for nitrogen excretion, therefore, reducing the amount of energy available to the animal (POZZA *et al.*, 2008).

There were no differences between the data sources for the corn and sorghum cultivars and WB with the adjusted AME<sub>4</sub> model (Table 2). Therefore, a single equation can be used for both the Brazilian and international data. For both it was observed the negative effect of ash together with the positive effect of EE on the AME. This is due to the saponification of ash with fats, resulting in reduced ADE and AME in the feed (POZZA *et al.*, 2008). These results are consistent with those obtained by Morgan *et al.* (1987), which showed the negative effect of ash on the AME of pig feed due to the diluent action of ash on the GE, reducing the organic matter content of feed.

For the AME<sub>5</sub> model, three independent equations of origin for the corn, sorghum and wheat bran (WB) were produced. Models adjusted for the corn and sorghum cultivars showed that ash was a significant independent variable. The estimated values of AME for the sorghum cultivars were lower than the AME values for the corn cultivars, due to a lower estimate of the regression coefficient. The CP exerted a negative effect on the

prediction of AME for WB, and was the regressor that differentiated the WB equation from the AME equations predicted for corn and sorghum (Table 2).

The AME<sub>3</sub>, AME<sub>4</sub> and AME<sub>5</sub> models presented few regressors for chemical composition (Tables 1 and 2). This is more applicable as the laboratory tests required are routinely performed, and there is no need to use bomb calorimeter or conduct experiments on animals, thereby reducing the cost of research and the execution time (PELIZZERI *et al.*, 2013; POZZA *et al.*, 2008).

The results obtained using the bootstrap resampling method (Table 3) showed that, after the model had been adjusted to the data for 1000 bootstrap samples, the ADE showed 100% PSR in the AME<sub>1</sub> model. This means that ADE was selected ( $p < 0.05$ ) in all 1000 bootstrap samples during the final stage of the stepwise procedure for selecting the regressor. Similarly, GE was highly significant in the 1000 bootstrap samples using the AME<sub>2</sub> model, with 95.7% selection. The high correlation between ADE and GE with AME highlights the importance of this regressor in its ability to explain the AME of pig feed.

In general, the values generated by the PSR models were deemed to be satisfactory (Table 3), as the selection frequency was above 50%, with the exception of the FD<sub>1</sub>\*CP interaction in the AME<sub>1</sub> model (PSR = 26.8%), SD\*ash and EE in the AME<sub>2</sub> model (PSR of 28.3 and 46.4%, respectively) and SD\*CF in the AME<sub>3</sub> model (PSR = 45.3%).

In the AME<sub>4</sub> and AME<sub>5</sub> models, the PSR of all regressors were greater than 50.0% (Table 3). Few reports are available in the animal science regarding the use of simulation procedures for the validation of covariate regression models. However, Scalón, Freire and Cunha (1998) recommended a minimum of 50% for the regressor selection frequency index in bootstrap samples to validate the predictive ability of multiple linear regression models with covariates. Their study was related to the health of newborns, and the parameters were estimated by the ordinary least squares method that appeared in the original model, selected by the stepwise procedure.

These results indicate that regressors can be used in the AME<sub>4</sub> and AME<sub>5</sub> models for the prediction of AME models that did not include ADE and GE, due to the high correlation with AME, suggesting that experiments are not required in order to predict the AME of pig feed.

The PJOR values observed in the 1000 bootstrap samples were low, ranging from 2.6% (AME<sub>2</sub>) to 23.4% (AME<sub>4</sub>) (Table 3), indicating low reliability of the predicted models for estimating the AME of corn, sorghum and wheat bran in pig feed. Reliability increases with higher PJOR values, and therefore, a higher chance of regressive

**Table 3** - Percentage of significance for regressor (PSR) and joint occurrence of regressor (PJOR) in 1000 bootstrap samples, including the average estimates, standard deviation and percentile confidence interval bootstrap (PCIB) in 500 different bootstrap samples

Model	Regressor <sup>1</sup>	PSR (%)	PJOR (%)	Average estimates	Standard deviation	PCIB(95%)
AME <sub>1</sub>	Intercept	100.0	9.4	33.46	78.38	[-102.50; 190.41]
	ADE	100.0		0.959	0.020	[0.918; 0.997]
	SD	63.6		45.33	6.00	[33.41; 57.01]
	FD <sub>1</sub> *CP	26.8		-4.666	1.095	[-6.92; -2.69]
AME <sub>2</sub>	Intercept	100.0	2.6	2251.99	383.46	[1466.05; 3057.57]
	GE	95.7		0.403	0.089	[0.224; 0.585]
	CF	64.6		124.71	5.58	[-135.43; -113.97]
	EE	46.4		32.77	12.73	[8.04; 57.94]
	SD*ash	28.3		-49.11	8.38	[-64.39; -32.71]
AME <sub>3</sub>	Intercept	100.0	6.4	3710.45	57.04	[3598.45; 3838.16]
	EE	63.0		37.53	13.27	[9.63; 63.86]
	SD*CF	45.3		-21.91	4.94	[-31.29; -12.25]
	FD <sub>1</sub> *CF	51.9		-91.44	4.16	[-99.44; -82.86]
AME <sub>4</sub>	Intercept	100.0	23.4	3828.83	64.46	[3694.70; 3946.86]
	ash	85.9		-106.58	27.37	[-163.54; -59.43]
	EE	70.7		44.64	12.54	[21.52; 70.02]
	FD <sub>1</sub> *CP	66.7		-37.08	5.25	[-46.88; -25.67]
AME <sub>5</sub>	Intercept	100.0	16.5	3983.70	38.65	[3909.31; 4058.70]
	ash	83.0		-80.77	26.80	[-129.79; -30.21]
	FD <sub>1</sub> *CP	66.9		-44.81	4.97	[-54.68; -35.97]
	FD <sub>2</sub> *ash	51.8		-44.16	17.84	[-81.73; -10.90]

<sup>1</sup>AME - Apparent metabolisable energy; ADE - Apparent digestible energy; CP - Crude protein; GE - Gross energy; CF - Crude fibre; EE - Ether extract; SD - Source dummy; FD<sub>1</sub> - Feedstuff dummy 1; FD<sub>2</sub> - Feedstuff dummy 2

set of selection in different databases, but originating from the main database obtained from the scientific literature.

Some criticisms have been made regarding the use of the stepwise procedure for regressor selection, due to bias of the coefficients and predictions, in addition to the instability of the model. Small changes in the data can have large impacts on the set of independent variables included in the model, the parameter estimates and its predictions (HESTERBERG *et al.*, 2008). Construction of the bootstrap sample requires the selection of random samples, which are the same size as the original sample, with replacement of each sample. Therefore, in a particular bootstrap sample, some data may be used more than once and others may be omitted (MONTGOMERY; PECK; VINNING, 2006). Changes to bootstrap sample may have affected the PJOR models in this study.

In setting the AME<sub>1</sub> model using the stepwise procedure, the intercept, ADE, SD and FD<sub>1</sub>\*CP appeared together 94 times in the 1000 bootstrap samples (Table 3). However, the number of possible models are set to include

up to five significant regressors, expressed as the sum of binominal coefficients for 27 regressors (n) and the number of successes (p) of between one and five, resulting in 101,583 possibilities. This indicates that although the reliability is low, the criteria for evaluating PJOR is subjective, given the high number of model adjustment possibilities. It is worth mentioning that there are no studies in the literature that have evaluated models based on the PJOR.

Similarly, for the AME<sub>2</sub> model adjusted by the stepwise procedure, the intercept, GE, CF, EE and SD\*ash interaction appeared together only 26 times in a total of 1000 bootstrap samples (Table 3). However, there were 44,551 potentially significant models in the final step of the stepwise process. Compared to the AME<sub>1</sub> model, the AME<sub>2</sub> model presented lower PJOR and a decreased number of possible model adjustments, indicating worse reliability.

For the AME<sub>3</sub> model, the PJOR was observed 64 out of 1000 times, from a total of 16,663 possible significant models, indicating lower reliability of this

model compared to  $AME_1$ . The  $AME_4$  model had an index of 234 occurrences in 1000 bootstrap samples (Table 3), however, the number of possible models from a total of 15 regressors and up to five successes (significant regressors at the end of the stepwise process) was 4,943 cases, which is much less than the previous models. The greatest PJOR can be expected for the  $AME_4$  model, based on the occurrence probabilities of the models.

For the adjusted  $AME_5$  model, there were 165 joint occurrences of regressors out of a total of 1000 bootstrap samples (Table 3). The number of possibilities for this adjusted model was 1023, the lowest of all the models, which indicates that the reliability of  $AME_5$  is lower than the  $AME_4$  model.

In the  $AME_1$  model, all parameters were significant except for the intercept. In the  $AME_2$ ,  $AME_3$ ,  $AME_4$  and  $AME_5$  models, all estimated parameters were found to be significant, as none of the estimated confidence intervals had a value of zero (Table 3). These results were similar to those obtained using the ordinary least squares method (Table 1), in which the multiple linear regression models for AME was adjusted based on the chemical and energy composition.

The parameters estimated using the ordinary least squares (OLS) method were among the minimum and maximum percent confidence interval bootstrap (PCIB) at 95% probability (Tables 1 and 3), indicating that bias was not present in the estimates of the bootstrap parameters, therefore, there was no need to use the confidence intervals to correct for bias.

The similarity between the parameter estimates and their standard deviations, obtained via the OLS and bootstrapping methods (Tables 1 and 3), showed that the residue analysis performed prior to the least squares adjustment resulted in the removal of outliers. This suggests that the chemical and energy composition data obtained from the literature and used in this study was adequate and consistent, as it met the assumptions of normality and homogeneity of errors for regression model adjustment.

## CONCLUSIONS

1. Based on the percentage of significance of the regressors, the regressive models  $AME_4 = 3824.440 - 105.294ash + 45.008EE - 37.257FD_1 * CP$  and  $AME_5 = 3982.994 - 79.970ash - 44.778FD_1 * CP - 43.416FD_2 * ash$  are valid for estimating the AME of corn, sorghum and wheat bran for use in pig diets;
2. For the five adjusted models, the stepwise procedure was not validated based on the percentage of joint occurrence of the regressors.

## ACKNOWLEDGEMENTS

The authors thank the Coordination for the Improvement of Higher Education Personnel (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES*) for the masters scholarship to C.N.L. and the State University of Parana West, UNIOESTE, for structural support.

## REFERENCES

- CASTILHA, L. D. *et al.* Modelos matemáticos para predição dos valores energéticos do milho para suínos. **Revista Brasileira de Saúde e Produção Animal**, v. 12, n. 2, p. 456-467, 2011.
- CASTILHO, R. A. *et al.* Equations to predict the metabolizable energy of meat and bone meal for growing pigs. **Ciência e Agrotecnologia**, v. 39, n. 6, p. 565-573, 2015.
- COELHO-BARROS, E. A. *et al.* Métodos de estimação em regressão linear múltipla: aplicação a dados clínicos. **Revista Colombiana de Estadística**, v. 31, n. 1, p. 111-129, 2008.
- HESTERBERG, T. *et al.* Least angle and  $l_1$  penalized regression: a review. **Statistics Surveys**, v. 2, p. 61-93, 2008.
- LANGER, C. N. *et al.* Precisão das estimativas em modelos de regressão da energia metabolizável do milho para suínos em crescimento. **Scientia Agraria Paranaensis**, v. 12, p. 331-335, 2013. Suplemento.
- LOVATTO, P. A. *et al.* Meta-análise em pesquisas científicas: enfoque em metodologias. **Revista Brasileira de Zootecnia**, v. 36, p. 285-294, 2007. Suplemento.
- MARIANO, F. C. M. Q. *et al.* Equações de predição de valores energéticos de alimentos obtidas utilizando meta-análise e componentes principais. **Ciência Rural**, v. 42, n. 9, p. 1634-1640, 2012.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. 4<sup>th</sup> ed. New Jersey: John Wiley & Sons, 2006. 612 p.
- MORGAN, C. A. *et al.* The prediction of the energy value of compounded pig foods from chemical analysis. **Animal Feed Science Technology**, v. 17, n. 2, p. 81-107, 1987.
- NASCIMENTO, G. A. J. *et al.* Equações de predição para estimar os valores energéticos de alimentos concentrados de origem vegetal para aves utilizando a metanálise. **Revista Brasileira de Zootecnia**, v. 38, n. 7, p. 1265-1271, 2009.
- NASCIMENTO, G. A. J. *et al.* Equações de predição para estimar valores da energia metabolizável de alimentos concentrados energéticos para aves utilizando meta-análise. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 63, n. 1, p. 222-230, 2011.
- NATIONAL RESEARCH COUNCIL. **Nutrient requirements of swine**. 11<sup>th</sup> ed. Washington, D.C.: The National Academies Press, 2012. 420 p.



NOBLET, J.; PEREZ, J. M. Prediction of digestibility of nutrients and energy values of pig diets from chemical analysis. **Journal of Animal Science**, v. 71, n. 12, p. 3389-3398, 1993.

OREDEIN, A. I.; OLATAYO, T. O.; LOYINMI, A. C. On validating regression models with bootstraps and data splitting techniques. **Global Journal of Science Frontier Research**, v. 11, n. 6, 2011.

PELIZZERI, R. N. *et al.* Avaliação de modelos de predição da energia metabolizável do milho para suínos. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 65, n. 2, p. 460-468, 2013.

POZZA, P. C. *et al.* Composição química, digestibilidade e predição dos valores energéticos da farinha de carne e ossos para suínos. **Acta Scientiarum. Animal Science**, v. 30, n. 1, p. 33-40, 2008.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 26 ago. 2016.

ROSTAGNO, H. S. *et al.* **Tabelas brasileiras para aves e suínos**: composição de alimentos e exigências nutricionais. 3. ed. Viçosa, MG: UFV, 2011. 252 p.

SAKOMURA, N. K.; ROSTAGNO, H. S. **Métodos de pesquisa em nutrição de monogástricos**. 2. ed. Jaboticabal: Funep, 2016. 262 p.

SCALON, J. D.; FREIRE, S. M.; CUNHA, T. A. Validation of models for predicting the use of health technologies. **Medical Decision Making**, v. 18, p. 311-319, 1998.



This is an open-access article distributed under the terms of the Creative Commons Attribution License