Acta Scientiarum

http://periodicos.uem.br/ojs/acta
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v40i1.37792

STATISTICS

Langmuir adsortion isotherm with regular and irregular autoregressive error structures

Cristiane Costa da Fonseca Cintra, Denismar Alves Nogueira* and Luiz Alberto Beijo

Departamento de Estatística, Universidade Federal de Alfenas, Rua Gabriel Monteiro da Silva, 700, 37130-001, Alfenas, Minas Gerais, Brazil.
**Author for correspondence. E-mail: denismar.nogueira@unifal-mg.edu.br*

ABSTRACT. The Langmuir isotherm is a nonlinear regression model, being one of the most applied in adsorption studies. In this type of study, the data are collected over time, which can provide correlated errors; in addition, the collection is not always done in an equidistant way, which may influence the estimation of model parameters. One way of modelling the dependent errors in a regression model is to use an autoregressive process that assumes that the observations are performed at equidistant intervals. However, the definition of the independent variable is often performed at irregular intervals, causing a reduction of information obtained from the dataset. One possible improvement in the adjustment quality of these models is the use of the irregular autoregressive process. The objective of this work was to compare the estimates of isotherm parameters with different irregular and regular autoregressive error structures, considering the positive autocorrelation in different sample sizes, error autocorrelation values and positioning of non-equidistant observations. It was found that there is a need to respect the assumptions of the model. The irregular autoregressive model is more appropriate because it is mostly more precise and accurate, especially when non-equidistance occurs in the initial third.

Keywords: nonlinear model; Monte Carlo simulation; estimators; precision; accuracy; autoregressive error.

Isoterma de adsorção de langmuir com estruturas de erros autorregressivos regulares e irregulares

RESUMO. A isoterma de Langmuir é um modelo de regressão não linear, sendo um dos mais aplicados em estudos de adsorção. Neste tipo de estudo os dados são coletados ao longo do tempo, o que pode proporcionar erros correlacionados, além disso, a coleta nem sempre é feita de forma equidistante, o que pode influenciar a estimação dos parâmetros do modelo. Uma forma de modelar os erros dependentes em um modelo de regressão é utilizar um processo autorregressivo que supõe que as observações sejam realizadas em intervalos equidistantes. Entretanto, a definição da variável independente muitas vezes é realizada em intervalos irregulares, ocasionando a redução de informações obtidas do conjunto de dados. Uma possível melhoria da qualidade do ajuste destes modelos é o uso do processo autorregressivo irregular. Este trabalho teve como objetivo comparar as estimativas dos parâmetros da isoterma com diferentes estruturas de erros autorregressivos, regulares e irregulares, considerando a autocorrelação positiva em diferentes tamanhos amostrais, valores de autocorrelação do erro e posicionamentos das observações não equidistantes. Verificou-se que há necessidade de se respeitar as pressuposições do modelo. O modelo autorregressivo irregular é mais indicado por ser, na maioria, mais preciso e acurado, principalmente quando a não equidistância ocorre no terço inicial.

Palavras-chave: modelo não-linear; simulação Monte Carlo; estimadores; precisão; acurácia; erros autorregressivos.

Introduction

By adjusting a nonlinear regression model, it is assumed that the model errors are uncorrelated, that is, they are independent of each other. In addition, the errors must be identically distributed with normal distribution mean zero and constant variance. If there is correlation between the errors, it is said that they are auto correlated and the coefficient of autocorrelation ρ determines the degree of correlation existing, that is, it measures its co variability.

When working with statistical modelling, the idea is that, when performing the adjustment of the usual model, it is verified whether it meets the assumptions of the regression methodology, which can be verified by graphical analysis of errors and confirmed analytically by tests of independence, normality and homogeneity of variances. If the dependence of errors is characterized, one way to adjust models with dependent errors, bypassing the problem of autocorrelation is to use time series techniques such as the autoregressive process, which

in turn assumes that the observations are performed at equidistant intervals. However, in some studies, the definition of the independent variable is often performed at different or non-equidistant intervals. By ignoring this irregularity in the measurements, you can have, for example, the reduction of information obtained from the dataset with possible problems of super or underestimation of model parameters. An alternative to improve the quality of the adjustment is to consider the irregularity in the data collection by applying the irregular autoregressive process (Vasagam, Jagathnath, & Kollannavar, 2016).

The Langmuir adsorption isotherm is a nonlinear model described by an equation that relates the volume of solute adsorbed to the concentration of the adsorbent in an adsorption phenomenon. Adsorption is a solid-fluid type mass transfer operation in which the ability of some solids to concentrate on their surface certain substances found in liquid or gaseous solutions is explored (Gomide, 1980).

The Langmuir adsorption isotherm is expressed by the Equation 1:

$$y_i = \frac{KCM}{1 + KC} + \varepsilon_i \quad (1)$$

in which y is the amount of solute adsorbed, K is the affinity parameter between the adsorbent and the adsorbed solute and indicates the rate with which the adsorption reaches its maximum value. M is the parameter that indicates the maximum amount of solute that can be adsorbed; C is the concentration of adsorbent (or time) and ε is the random error associated with the model ($\varepsilon_i \sim N(0,\sigma^2)$).

Every adsorbent surface has a limited adsorption capacity, becoming saturated when the limit is reached. In the Langmuir isotherm, this limit is given by parameter M, because as the concentration increases, the amount of solute adsorbed tends to M. The affinity parameter between the adsorbent and the adsorbed solute (K) indicates the rate at which the adsorption reaches its maximum value. The higher the value of K the faster is the saturation of the adsorbent surface.

It can be noticed that, in many studies on adsorption where the modelling is performed by the Langmuir isotherm, errors can be correlated and in addition, data collection is not always performed on a regular basis. The autocorrelation of the errors and the irregularity in the measurements can influence the estimation of parameters of the model in question.

The objective of this study was to compare estimates of the Langmuir isotherm parameters with different autoregressive error structures of order 1, regular and irregular, with different sample sizes, correlations and non-equidistance in the measurements of the independent variable using Monte Carlo simulation. The objective was also to evaluate the scenario ignoring the presence of autocorrelation.

Autoregressive errors

When adjusting a regression model in which the assumption of independence in waste is not met, using classical methods, they produce biased estimates. A regression model (Equation 1) in which autocorrelation is present is known as an autoregressive process and a way of representing how the new error becomes in a process of order 1 is given by Equation 2. Where ε_i is given by Equation 2:

$$\varepsilon_i = \rho\varepsilon_{i-1} - u_i \quad (2)$$

where:
ρ is the coefficient of autocorrelation and u_i an error term with the characteristics of the hypotheses of the regression model, with i =1,...,n.

An autoregressive process is a regression of a variable by itself (Sartoris, 2003). When we have a lag of the variable, that is, when the value of ε_i and its immediately previous value are involved it is named AR(1). The estimation of the coefficient of autocorrelation ρ is performed by the method of the least squares, that is, it is sought ρ that minimizes the sum of squares of the errors.

As it is generally not possible to collect data with equidistant measures, it is common in some studies to ignore this irregularity in the measurements and treat the data as if they were regular, which can introduce biases in the parameter estimates.

An alternative is to use a method to analyse a time series with not equally spaced data. The authors[1] consider a time series y_t that can be decomposed as $y_t = b(t) + \varepsilon_t$, where, $b(t)$ is a function that represents the trend component and ε_t the random noise. To consider the irregularity in the data, the authors take $\{y_t, t_i\}$, a sample of y_t. It is then defined $\varepsilon_{t_i} = y_{t_i} - b_p(t_i)$ and $\{\varepsilon_{t_i}, t_i\}$ is treated as a sample of ε_t. An autoregressive process of order 1 is defined for irregular data such as:

Definition: A time series $\{\varepsilon_{t_i}, t_i\}$ is stationary irregularly sampled if ε_t is stationary and if for every t and $\Delta_i > 0$, $\varepsilon_{t+\Delta_i} = \rho^{\Delta_i}\varepsilon_t + \sigma_{\Delta_i}u_{i+\Delta_i}$ in which $u_{t+\Delta_i} \sim N(0,1)$ and $cov(u_t, u_s) = 0$ for each $t \neq s$

and $\sigma_{\Delta_i}^2 = \sigma^2\left(\frac{1-\rho^{2\Delta_i}}{1-\rho^2}\right)$, for some $\sigma > 0$ (Vasagam, Jagathnath, & Kollannavar, 2016).

In this work, the nonlinear Langmuir isotherm model (Equation 1) was used to represent the trend component and the irregularity in the data was previously defined for the simulation process.

Thus, in the regression model in which the errors are irregular autoregressive (IS-AR(1)), ε_i is given by Equation3:

$$\varepsilon_i = \rho^{\Delta_i}\varepsilon_{i-1} - u_i \quad (3)$$

where:
Δ_i represents the interval between two subsequent observations and μ_i is a random variable with mean zero and constant variance σ^2. The estimation of the autocorrelation coefficient ρ is performed by the least squares method.

Experimental part

Monte Carlo simulation process

The evaluation of different scenarios (120) in the estimation of the parameters of the nonlinear Langmuir Isotherm model was performed considering different sample sizes n ∈ {12, 30, 60, 90}, different autocorrelation values of the errors ρ ∈ {0.2; 0.5; 0.8} and different positioning of the non-equidistant observations, via Monte Carlo simulation. In order to allow non-equidistance, some points of the independent variable C were taken randomly, considering the order. For a better understanding, Figure 1 shows the parts of the independent variable where the data was taken. Nine situations were evaluated: withdraw 25, 50 and 75% from the initial, central and final thirds of the data.

The values of the parameters for the simulation process were K = 0.9 and M = 20 (Fungaro, Yamaura, & Graciano, 2010). Concentration values ranged from zero to ten mg L^{-1}.

One thousand simulations were performed for each of the scenarios using the program R (R Core Team, 2016). The adjustment procedures for the model with independent and correlated error structures were implemented through the *gnls* function of the *nlme* package (Pinheiro, Bates, DebRoy, & Sarkar, 2016). The random error was generated from a normal distribution with zero mean and standard deviation equal to two. Considering the parameterization presented in Equation 1, the errors contained a regular autoregressive structure of order 1 presented in Equation 2 and irregular autoregressive of order 1 presented in Equation 3. In the case of irregular errors, possible problems were diagnosed when simulations involving ρ of small numerical value and the exponents of large values, making the computational process difficult. The solution found was to work with a linear transformation of the distances given by $\delta_i = \frac{\Delta_i}{\min(\Delta_i)}$ and use the optimize function of the R program stats package to calculate the value that minimizes the sum of squares of the errors by replacing Δ_i with δ_i. All simulations were followed in accordance with the Durbin-Watson test (p-value < 5%) and pacf and acf plot.

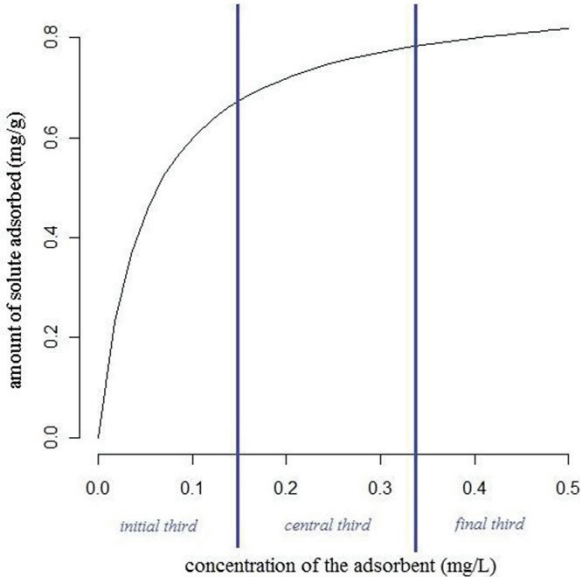

Figure 1. Graphical representation of a Langmuir isotherm model with the concentration values separated by the parts: initial third, middle third and final third.

Evaluation of parameters estimates and evaluators of adjustment quality

The relative mean bias percentage (RMB) was used to verify the accuracy of the estimates of the parameters of the model studied, and the relative mean square error (RMSE) was used for the precision of the estimates. Their values were calculated respectively by the following Equation 4 and 5 (Casella & Berger, 2010):

$$RMB = \frac{\left(\frac{\sum_{i=1}^{S}|\hat{\theta}_i - \theta|}{|\theta|}\right)}{s} \times 100 \quad (4)$$

$$RMSE = \frac{\left(\frac{\sum_{i=1}^{S}(\hat{\theta}_i - \theta)^2}{|\theta|}\right)}{s} \times 100 \quad (5)$$

where:
in which s is the number of simulations ($s = 1000$), θ is the value of the simulated parameter, that is, the

value of ρ, K or M and $\hat{\theta}_i$ is the estimate of parameter θ in the i-th simulation.

In order to measure or evaluate the adjustment quality of the models with regular or irregular autoregressive errors, the Akaike (Akaike, 1974) ($AIC = -2logL(\hat{\theta}) + 2p$) and the Bayesian information (Schwarz, 1978) ($BIC = -2logL(\hat{\theta}) + plog(n)$) information criteria were used, in which p is the number of parameters of the model, $L(\hat{\theta})$ is the maximum of the likelihood function, considering the parameter estimates and n is the sample size.

Cases where AIC values are very close, for different models, Motulsky and Christopoulos (2003) recommended the use of Akaike's weight, which provides a probability that the model with the lowest AIC is the correct model. The probability is given by: $probability = \frac{e^{-0.5\Delta AIC}}{1+e^{-0.5\Delta AIC}}$, in which $\Delta_{AIC} = AIC_B - AIC_A$ being $AIC_B > AIC_A$.

Results and discussion

Evaluation of the parameter estimates

In almost all sample sizes, the IS-AR(1) model was more accurate than the AR(1) model. This statement can be verified in all the results obtained. Figure 2 shows the graphical representation of the results referring to the accuracy of parameter K estimates. The graphs in the first column refer to $\rho = 0.2$, those in the second column to $\rho = 0.5$ and those in the third column to $\rho = 0.8$. It is observed that as autocorrelation increases, accuracy decreases, as well as, as the sample size increases, accuracy increases, regardless of the model. It is also noted that the lack of data in the initial third (Figure 2a) increases the RMB of the estimates. This occurrence is accentuated for sample size 12. A greater difference between the correlated and uncorrelated models is observed in the presence of higher correlation.

Regarding the accuracy of the estimates of *K*, it is noted that, according to Figure 3, as autocorrelation increases, there is an increase of error. As the sample size increases, precision increases, regardless of the model, which was expected. In almost all sample sizes, the IS-AR(1) model was more accurate than the AR(1) model and it can be stated that the lack of data in the initial third decreases the precision of the estimates more than when the loss occurs in the other thirds. Convergence problems of the nonlinear estimation algorithm were observed more expressively for smaller samples and data losses in the initial third.

Figure 4 presents the results regarding the accuracy of the estimates of *M*, which present behaviour similar to the estimates of *K* when there is an increase in autocorrelation, since accuracy decreases, and, as sample size increases, accuracy increases regardless of the model. However, when comparing the accuracy between the estimates of the AR(1) and the IS-AR(1) models, a small percentage difference in the RMB values is observed. Larger differences were observed between the correlated and uncorrelated models as the correlation increases. The region of data loss does not interfere with the estimates of RMB.

As the sample size increases, accuracy increases, mainly in $\rho = 0.2$ and in $\rho = 0.5$, which was expected. Unlike parameter K, not always the IS-AR(1) model was more accurate than the AR(1) model. For a high correlation, accuracy was shown to be similar for all sample sizes. The correlated and uncorrelated models presented similarity for RMSE. As autocorrelation increases, there is an increase of error in parameter M (Figure 5).

Comparison of the autoregressive models AR(1) and IS-AR(1) with the model that does not consider autocorrelation

In order to verify the importance of considering the presence of autocorrelation in the estimation process of the Langmuir isotherm parameters, the accuracy and precision obtained from the AR(1) and IS-AR(1) models were compared with those of the model which does not consider the presence of autocorrelation.

Table 1 shows the mean values of RMB and RMSE obtained in the estimation of parameters *K* and in the three error structures. The incorporation of autocorrelation in the model presented more precise and more accurate estimates, mainly in parameter K. By not considering autocorrelation, the value of RMSE of parameter *K* estimation is, on average, 78% higher than the estimation when autocorrelation is considered and 9% higher in the estimation of parameter *M*. The results corroborate the studies done by the authors (Mazzini, Muniz, Silva, & Aquino, 2005; Pereira, Muniz, & Silva, 2005; Mendes, Muniz, Silva, Mazzini, & Silva, 2009; Porter et al., 2010; Neto, Carvalho, & Mischan, 2013), that considered the structure of autoregressive errors in the adjustment of nonlinear models and obtained more accurate estimates of parameters and more efficient adjustments when compared to the adjustment of the models without the presence of autocorrelation.

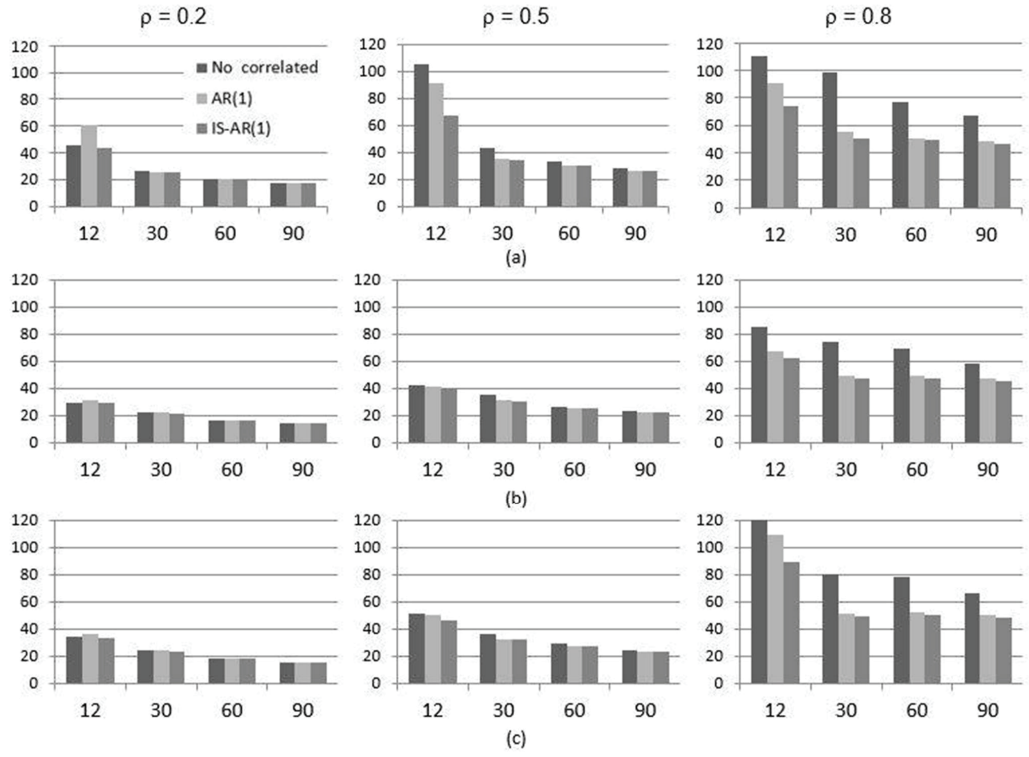

Figure 2. Representation of RMB (%) values of parameter *K* estimates at different autocorrelation values with 50% of data taken from: (a) initial third, (b) central third, and (c) final third.

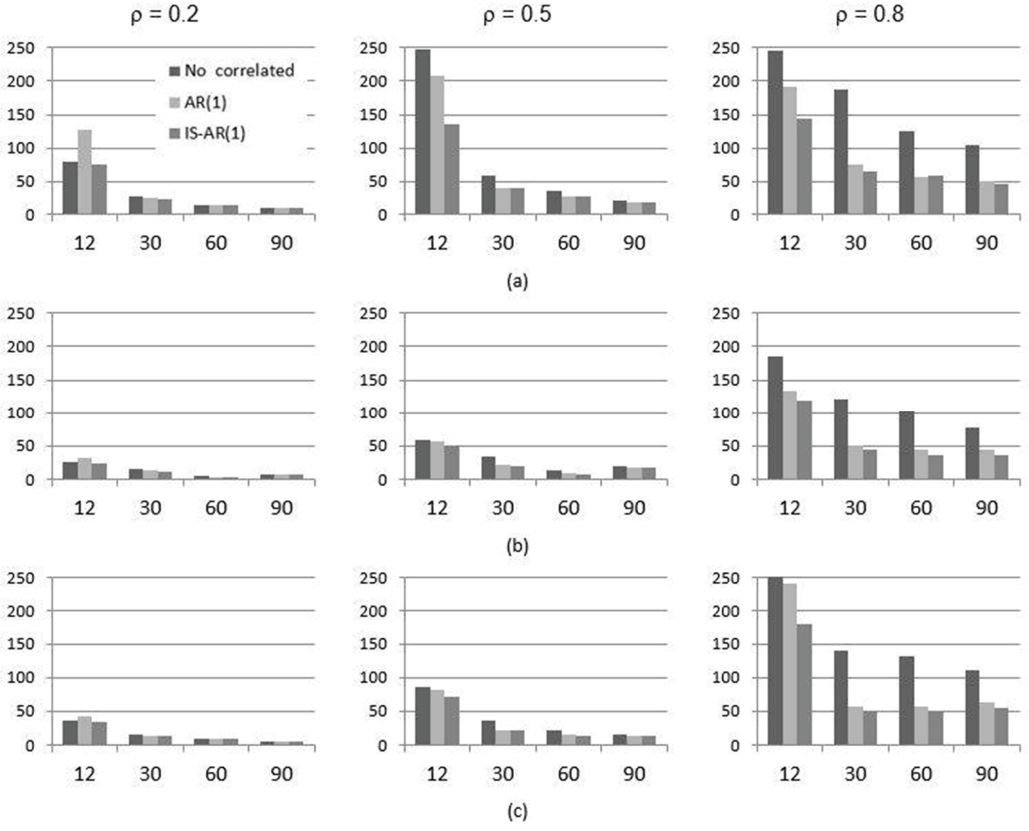

Figure 3. Representation of RMSE (sqrt) values of parameter *K* estimates in different autocorrelation values with 50% of the data taken from: (a) initial third, (b) central third, and (c) final third.

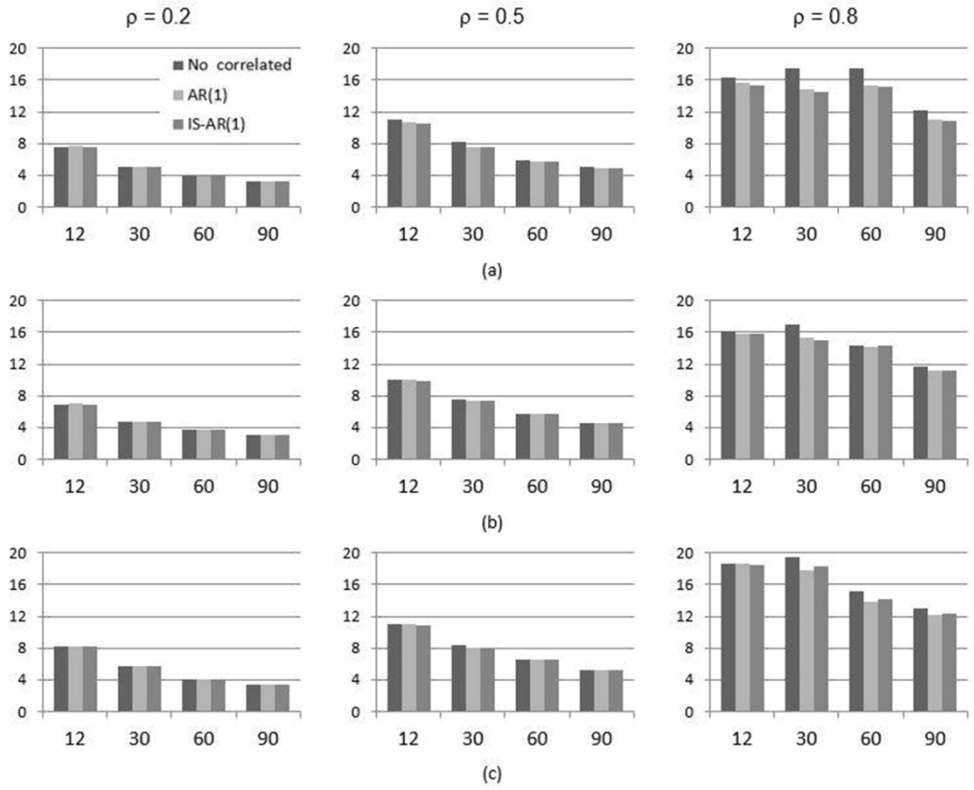

Figure 4. Representation of RMB (%) values of parameter M estimates at different autocorrelation values with 50% of the data taken from: (a) initial third, (B) central third, and (c) final third.

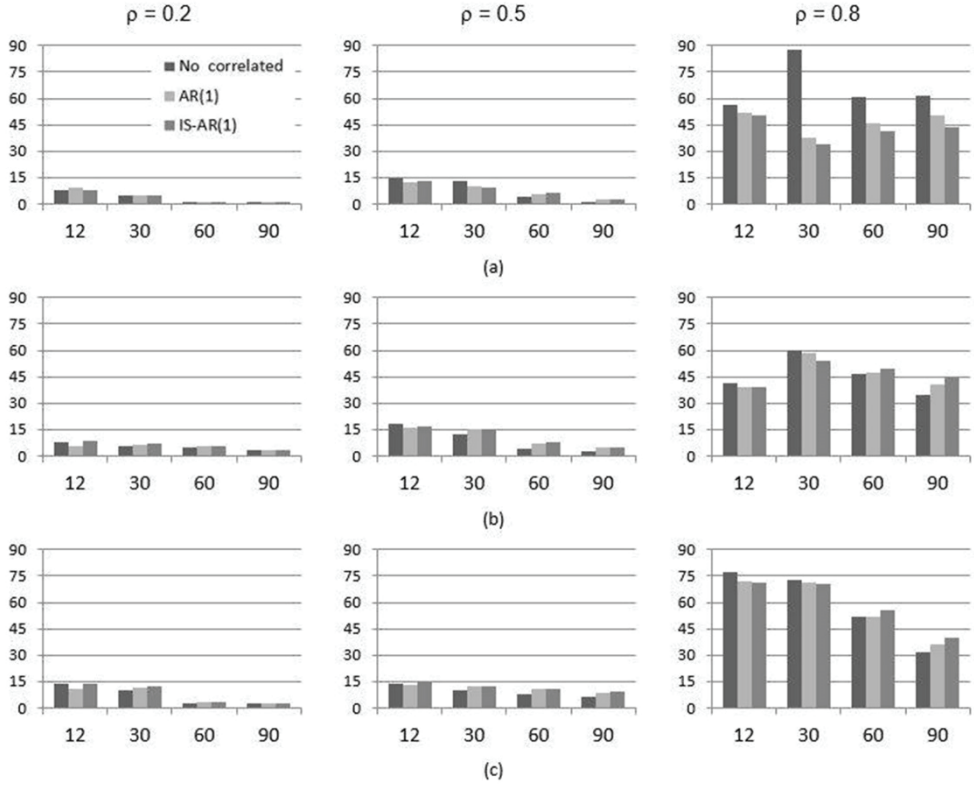

Figure 5. Representation of RMSE (sqrt) values of parameter M estimates in different autocorrelation values with 50% of the data taken from: (a) initial third, (B) central third, and (C) final third.

Table 1. Comparison of mean values of RMB and RMSE(sqrt) for parameters *K* and *M* estimates.

	Parameter	IS-AR(1)	AR(1)	No correlated
RMB (%)	K	36.80	39.17	48.78
	M	8.78	8.83	9.36
RMSE	K	43.43	50.53	77.17
	M	22.14	22.10	24.14

Adjustment quality assessment

The mean AIC and BIC values of the autoregressive models AR(1) and IS-AR(1) were close. However, when the autoregressive models are compared with the model that does not consider autocorrelation, the first ones present smaller values in both the Akaike information criterion and the Bayesian information criterion, as observed in Table 2. Note that, as the sample size increases, the AIC and BIC values of the autocorrelation models distance themselves from the model that does not consider it.

Table 2. Mean values of AIC and BIC of the three models studied by sample size.

	n	IS-AR(1)	AR(1)	No Correlated
AIC	90	317.14	316.90	342.61
	60	211.26	211.08	226.08
	30	105.42	105.24	110.32
	12	41.33	40.74	41.84
BIC	90	324.10	323.85	349.56
	60	216.99	216.81	231.81
	30	109.08	108.90	113.98
	12	42.23	41.65	42.75

The results obtained are in agreement with the authors Prado, Savian, and Muniz (2013), who obtained close values for the Akaike information criterion in models with different structures of autoregressive errors. Other studies (Neto et al., 2013) concluded that adjustments in models with autoregressive error structure of order 1 in relation to models with independent error structures were shown to be better in the study of bovine animals' growth, considering the Akaike (AIC) and the Bayesian (BIC) information criteria.

Since the AIC values between the models AR(1) and IS-AR(1) are very close, the Akaike Weights values were calculated. The results varied between 0.43 and 0.48 for the IS-AR(1) model and, consequently, 0.52 and 0.57 for the AR(1) model, that is, the probability of decision between the two models can be considered close and therefore, according to AIC, any of the models could be chosen. Thus, according to the results of accuracy and precision obtained in the simulation studies, the IS-AR(1) model is more adequate than the AR(1) model in most of the sample sizes studied for parameter K and similar for parameter M.

Conclusion

By adjusting the model with autoregressive error structures, considering the positive autocorrelation and non-equidistance in the independent variable, it is concluded that the estimates of parameter K of the irregular model (IS-AR(1)) are more precise and more accurate than estimates of the regular model (AR(1)) in practically all scenarios studied, regardless of the factors under study. It obtained similar behaviour for parameter M.

Ignoring the presence of autocorrelation between the data in the Langmuir Isotherm produces an increase in bias and a decrease in the precision of its parameter estimates, manly in high correlated.

The IS-AR(1) model is more appropriate for better estimation of parameter *K* of the isotherm.

The results of the simulations of this study show that, if the presence of autocorrelation is identified, the IS-AR(1) and AR(1) models provide an improvement in the adjustment quality. From the AIC and BIC values obtained from the model that does not consider autocorrelation it is noticed that, mainly, in larger samples, the autocorrelation should not be ignored when the Langmuir isotherm model is adjusted.

Excessive data loss and very high autocorrelation values hinder the nonlinear estimation process. Good estimates are reached when the sample number is greater than 30 and the autocorrelation is less than 0.5.

Acknowledgements

The authors acknowledge the financial support by the Brazilian agencie *Fundação de Amparo à Pesquisa do Estado de Minas Gerais*- Fapemig (Process: APQ-02225-13).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on. Automatic Control, 19*(6), 716-723. doi: 10.1109/TAC.1974.1100705

Casella, G., & Berger, R. L. (2010). *Inferência estatística* (2a ed.) São Paulo, SP: Cengage Learning.

Fungaro, D. A., Yamaura, M., & Graciano, J. E. A. (2010). Remoção de íons Zn^{2+}, Cd^{2+} e Pb^{2+} de soluções aquosas usando compósito magnético de zeólita de cinzas de carvão. *Química Nova, 33*(6), 1275-1278. doi: 10.1590/S0100-40422010000600011

Gomide, R. (1980). *Operações unitárias - separações mecânicas*. (3º Vol.). São Paulo, SP: Reynaldo Gomide.

Mazzini, A. R. A., Muniz, J. A., Silva, F. F., & Aquino, L. H. (2005). Curva de crescimento de novilhos Hereford: heterocedasticidade e resíduos autorregressivos. *Ciência Rural, 35*(2), 422-427. doi: 10.1590/S0103-84782005000200028

Mendes, P. N., Muniz, J. A., Silva, F. F., Mazzini, A. R. A., & Silva, N. A. M. (2009). Análise da curva de crescimento difásica de fêmeas Hereford por meio da função não linear de Gompertz. *Ciência Animal Brasileira, 10*(2), 454-461.

Motulsky, H. J., & Christopoulos, A. (2003). *Fitting models to biological data using linear and nonlinear regression: A practical guide to curve fitting*. San Diego, CA: GraphPad Software Inc.

Neto, F. S., Carvalho, L. R., & Mischan, M. M. (2013). Ajuste de modelos não lineares a dados de crescimento com estrutura de erros independentes e autorregressivos de primeira ordem - aplicações. *Revista Brasileira de Biometria, 31*(4), 631-644.

Pereira, J. M., Muniz, J. A., & Silva, C. A. (2005). Nonlinear models to predict nitrogen mineralization in an Oxisol. *Scientia Agricola, 62*(4), 395-400. doi: 10.1590/S0103-90162005000400014

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2016). nlme: Linear and Nonlinear Mixed Effects Models. *R Package Version, 3*, 1-128.

Porter, T., Kebreab, E., Kuhi, H. D., Lopez, S., Strathe, A. B., & France, J. (2010). Flexible alternatives to the Gompertz equation for describing growth with age in turkey hens. *Poultry Science, 89*(2), 371-378. doi: 10.3382/ps.2009-00141

Prado, T. K. L., Savian, T. V., & Muniz, J. A. (2013). Ajuste dos modelos Gompertz e Logístico aos dados de crescimento de frutos de coqueiro anão verde. *Ciência Rural, 43*(5), 803-809.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, AT: R Foundation for Statistical Computing.

Sartoris, A. (2003). *Estatística e introdução à econometria*. São Paulo, SP: Saraiva.

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

Vasagam, N. & K M, Jagathnath K. & Kollannavar, G. (2016). A Software Model for Estimation and Prediction of Unevenly Spaced Livestock Population. *International Journal of Advanced Research in Computer Science and Software Engineering, 6*, 747-751.

Received on June 21, 2017.
Accepted on October 25, 2017.