Title	Global and local tests to assess stationarity of Markov transition models
Author(s)	Rodrigues de Lara, Idemauro Antonio; Hinde, John; Taconeli, Cesar Augusto
Publication Date	2018-02-09
Publication Information	Rodrigues de Lara, Idemauro Antonio, Hinde, John, & Taconeli, Cesar Augusto. (2019). Global and local tests to assess stationarity of Markov transition models. Communications in Statistics - Simulation and Computation, 48(4), 1019-1039. doi: 10.1080/03610918.2017.1406504
Publisher	Taylor & Francis
Link to publisher's version	https://doi.org/10.1080/03610918.2017.1406504
Item record	http://hdl.handle.net/10379/15793
DOI	http://dx.doi.org/10.1080/03610918.2017.1406504

Global and local tests to assess stationarity of Markov transition models

Running Head: Tests to assess stationarity of transition models

Idemauro Antonio Rodrigues de Lara, John Hinde and Cesar Augusto Taconeli
Exact Sciencies Department, Luiz de Queiroz College of Agriculture, University of São Paulo
idemauro@usp.br

Key Words: longitudinal categorical data, transition probabilities, generalized linear models, simulation study.

Abstract

We present global and local likelihood-based tests to evaluate stationarity in transition models. Three motivational studies are considered. A simulation study was carried out to assess the performance of the proposed tests. The results showed that they present good performance with the control of the type-I error, especially for ordinal responses, and control of the type-II error, especially for the nominal case, and asymptotically they are close to the classical test performance. They can be executed in a single framework without the need to estimate the transition probabilities, incorporating both categorical and continuous covariates, and used to identify sources of non-stationarity.

1. Introduction

The class of transition models are based on the Generalized Linear Model (GLM). This methodology is useful for the analysis of longitudinal data, especially, with categorical data. In these cases, the possible dependence within longitudinal data is incorporated through a Markov-type stochastic process, where the response categories form the state-space, i.e., $S = \{1, 2, 3, \ldots, k\}$, for a k-category response. Additionally, for an ordinal response, the state-space is considered to have the natural ordering of the integers. Here, we consider a discrete-time process, where $\tau = \{0, 1, 2, \ldots, T\}$, corresponds to the set of specific time points at which the data are observed. The first-order Markov assumption for responses $\{Y_\tau\}$ is described by the conditional probability:

$$P(Y_t = b \mid Y_{(t-1)} = a, Y_{(t-2)} = c, \ldots, Y_{(0)} = u) = P(Y_t = b \mid Y_{(t-1)} = a) = \pi_{ab}(t-1,t) \quad (1)$$

with $a, b, c, \ldots, u \in S = \{1, 2, \ldots, k\}$ and $t \in \tau = \{0, 1, \ldots, T\}$. This assumption (1) defines the transition probabilities in a Markov chain and it means that an individual's state at the time t does not depend the complete history of the process but only on the state at time $t-1$ (Jones and Smith, 2001). These probabilities may be represented in a one-step $k \times k$ transition matrix:

$$\boldsymbol{P}(t-1,t) = \begin{pmatrix} \pi_{11}(t-1,t) & \pi_{12}(t-1,t) & \ldots & \pi_{1k}(t-1,t) \\ \pi_{21}(t-1,t) & \pi_{22}(t-1,t) & \ldots & \pi_{2k}(t-1,t) \\ \vdots & \vdots & \ldots & \vdots \\ \pi_{k1}(t-1,t) & \pi_{k2}(t-1,t) & \ldots & \pi_{kk}(t-1,t) \end{pmatrix}.$$

In order to simplify the mathematical notation we hereafter write $\pi_{ab}(t-1,t) = \pi_{ab}(t)$ and $\boldsymbol{P}(t-1,t) = \boldsymbol{P}(t)$, where the argument t indicates time dependence of the transition probabilities. Also, if these transition probabilities are homogeneous over time, we have $\pi_{ab}(t) = \pi_{ab}$ for all $t \in \tau$. In this case we can say the process is stationary and there is a unique transition matrix, $\boldsymbol{P}$. This assumption is very important in transition models because it simplifies the matrix of transition probabilities as well as the number of unknown parameters. Hence, general hypotheses of interest are:

$$\begin{aligned} \mathrm{H}_0 : \pi_{ab}(t) &= \pi_{ab} \text{ for all } t = 1, 2, \ldots T, \text{ for all } a, b \in S = \{1, 2, \ldots, k\}; \\ \mathrm{H}_1 : \pi_{ab}(t) &\neq \pi_{ab}(s) \text{ for some } t \neq s \text{ and some } a, b \in S. \end{aligned} \quad (2)$$

In matrix notation, the hypotheses (2) are $\mathrm{H}_0 : \boldsymbol{P}(t) = \boldsymbol{P}$, for all $t = 1, 2, \ldots T$ against $\mathrm{H}_1 : \boldsymbol{P}(t) \neq \boldsymbol{P}(s)$, for some $t \neq s$. To evaluate these hypotheses, Anderson and Goodman (1957) presented the test:

$$\xi = \sum_{t=1}^{T} \sum_{a=1}^{k} \sum_{b=1}^{k} \frac{n_a(t-1)[\hat{\pi}_{ab}(t) - \hat{\pi}_{ab}]^2}{\hat{\pi}_{ab}}, \quad (3)$$

where $n_a(t-1)$ is the number of individuals that are in category a at time $t-1$, and $\hat{\pi}_{ab}$ and $\hat{\pi}_{ab}(t)$ are estimates of transition probabilities under H_0 and H_1, respectively. It is shown that, asymptotically, the ξ statistic has a χ^2_v distribution, for some appropriate degrees of freedom v. This test (3) was originally proposed for nominal data from homogeneous populations (see Anderson and Goodman, 1957), i.e., without the effect of stratifications, giving $v = k(k-1)(T-1)$. Sometimes, the number of states that influence the present individual state is greater than one and this leads to a q-order Markov chain, $q > 1$. For more details on Markov chains and stochastic processes, see Stirzaker (2005) and Jones and Smith (2001). For the estimation of the transition matrix elements, Good (1955), Anderson and Goodman (1957), Goodman (1962), Lindsey (1995, 2004) and Agresti (2012) describe the likelihood estimation procedure, where the estimators $\hat{\pi}_{ab}(t)$ and $\hat{\pi}_{ab}$ coincide with the observed relative frequencies of specific contingency tables.

In this paper, these transition probabilities are estimated using a GLM, which is more flexible as it allows the inclusion of covariates. These corresponds to the so-called Markov transition models (see Diggle *et al.*, 2002 and Molenberghs and Verbeke, 2005),and allow the study of what happens to a response category from one moment of time to another, as well as assessing the effects of covariates on the transition probabilities. The aim of this work is to present a new test to assess stationarity in such transition models and also to identify sources of non-stationarity. This is particularly relevant because when the stationarity assumption is met there are fewer parameters to be estimated and the model is more easily interpreted, but, often in practice, this important condition is not checked.

2. Examples and Methods

2.1. Examples

We present three experimental studies as motivational examples in which transition models can be applied.

2.1.1. Example 1: respiratory data

This study assessed the individual respiratory condition of 111 patients with respiratory problems, at baseline and four further visits during the study period. The patients were followed up by two medical centres and were randomized to receive either an active or placebo treatment. The respiratory condition of each patient was classified according to a set of five ordinal classes, which reflects a response scale of the less favourable to more favourable (1=terrible, 2=bad, 3=moderate, 4=good, 5=excellent). The covariates for this study were sex and age. More details are given by Koch *et al.* (1990).

2.1.2. Example 2: pig behaviour data

This data set was a part of a study developed by Castro (2016), with 124 animals measured monthly over 4 time occasions, from March to July 2014. The design was completely randomized with a 2×4 factorial treatment structure, corresponding to combinations of two environmental enrichment levels (E1: with nvironmental enrichment and E2: without environmental enrichment) and four genetic lineages (L1; L2; L3; L4). In this experiment, environmental enrichment consists of the use of simple objects in the pens (suspended chains and plastic containers of two different sizes). The response variable of interest is a score measuring the degree of lesions at the front of the animal, that were classified as: 1: absence of lesions; 2: moderate degree of lesions; and 3: serious lesions. The lesion degree is an indicator of aggressive behaviour among the animals. More details on this data set and design are available in Castro (2016).

2.1.3. Example 3: agronomic data

The research studies from Pereira *et al.* (2015a, 2015b), involve an experiment on an elephant grass pasture grazed by dairy cows. The experiment was a complete randomized block design with a 2×2 factorial treatment structure, corresponding the combinations of two pre-grazing conditions and two post-grazing heights. The response variable was the type of vegetation observed in the field, which was classified as 1: tussocks, 2: bare ground, and 3: weeds. Observations were taken at 40 points in each one of the four paddocks present in each block over six seasons, from January 2011 to April 2012. As there are always 40 points observed in each paddock we have repeated measures and there were, initially, $40 \times 16 = 640$ points per season. However, in early spring, one of the paddocks was lost and so the total number of observations was 600. In this work, we used this study as motivation and the original data set was used to obtain simulated data with first order stochastic dependence, but without block effects for the sake of simplicity.

2.2. Methods

2.2.1. Transition models

When Y_t represents a response at time t, which depends on the previous responses, i.e., $\{y_0, y_1, \ldots, y_{t-1}\}$, as well as a set of covariates, we have the so called transition models. If $f(y_0, y_1, \ldots, y_T)$ is a joint distribution of the vector $(Y_0, Y_1, \ldots, Y_T)$, disregarding the covariate effects, these models use the factorization:

$$f(y_0, y_1, \ldots, y_T) = f(y_0) f(y_1 \mid y_0) f(y_2 \mid y_0, y_1) \ldots f(y_T \mid y_0, y_1, \ldots, y_{T-1}), \qquad (4)$$

and are, therefore, conditional models (Agresti, 2012). Considering the first-order Markov assumption (1) the expression (4) reduces to

$$f(y_0, y_1, \ldots, y_T) = f(y_0) f(y_1 \mid y_0) f(y_2 \mid y_1) \ldots f(y_T \mid y_{T-1}). \qquad (5)$$

while for general order q dependence we can write (4) as

$$\begin{aligned} f(y_0, y_1, \ldots, y_T) &= f(y_0, \ldots, y_{q-1}) f(y_q \mid y_0, \ldots, y_{q-1}) \ldots f(y_T \mid y_{T-q}, \ldots, y_{T-1}) \\ &= f(\boldsymbol{h}_q) f(y_q \mid \boldsymbol{h}_q) f(y_{q+1} \mid \boldsymbol{h}_{q+1}) \ldots f(y_T \mid \boldsymbol{h}_T). \end{aligned} \qquad (6)$$

where $\boldsymbol{h}_t = (y_{t-1}, y_{t-2}, \ldots, y_{t-q})$.

These concepts can be extended to longitudinal data with covariates by using GLMs and extensions, in which the parameters that relate to covariates and previous responses are defined in the linear predictor (see, for example, Zeger and Liang, 1992; Diggle *et al.*, 2002; Molenberghs and Verbeke, 2005). Due to the decomposition in (6), conditional transition models (conditioning on the initial history $\boldsymbol{h}_q$) can be fitted to data by standard techniques for GLMs with independent data. The data dependence is incorporated by the presence of the previous responses (history) in the linear predictor.

In this context, let $\boldsymbol{y}_i = (y_{i0}, y_{i2}, \ldots, y_{iT})'$ be the $(T+1) \times 1)$ vector of response variables for the i-th individual ($i = 1, 2, \ldots, N$), $\boldsymbol{x}_{it} = (x_{it1}, \ldots, x_{itp})'$ an associated $(p \times 1)$ vector of covariates and $\boldsymbol{h}_{it} = (y_{i(t-1)}, y_{i(t-2)}, \ldots, y_{i(t-q)})$ the $(q \times 1)$ vector of previous responses, i.e., the q-step history for individual i at time t. According to Diggle *et al.* (2002), a

Markov transition model specifies a generalized linear model for the random variable $Y_{it} \mid \boldsymbol{h}_{it}$, that is assumed to have a distribution that belongs to the canonical exponential family and the conditional expectation, $\mu_{it}^C = \mathrm{E}(Y_{it} \mid \boldsymbol{h}_{it})$, is defined as:

$$g(\mu_{it}^C) = \eta_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \sum_{r=1}^{s} \alpha_r f_r^*(\boldsymbol{h}_{it}), \tag{7}$$

where $g(\mu_{it}^C)$ is a link function and f_r^* are functions that define the structure of the transition model in the linear predictor (stochastic dependence). Also, the conditional variance is given by $v_{it}^C = \mathrm{Var}(Y_{it} \mid \boldsymbol{h}_{it}) = \phi v(\mu_{it}^C)$, where $v(.)$ is a variance function and ϕ a known dispersion parameter. The vector $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ represents the weights that the explanatory variables have on transition probabilities, in which $\boldsymbol{\beta}$, of dimension $p \times 1$, is associated with the covariates, and $\boldsymbol{\alpha}$ is associated with the history (the previous responses) and has a dimension that depends on both the order q and the specific form of the functions f_r^*. These parameters are estimated by maximum likelihood and when the model is assumed to be stationary, only one model is fitted using a sum of individual contributions to the likelihood function (Azzalini, 1983; Diggle *et al.*, 2002; Molenberghs and Verbeke, 2005). Thus, considering a stationary transition model of order q, from (6) the contribution to the likelihood function of the i-th individual is given by:

$$f(y_{i0}, y_{i2}, \ldots, y_{i(q-1)}) \prod_{t=q}^{T} f(y_{it} \mid y_{i(t-1)}, y_{i(t-2)}, \ldots, y_{i(t-q)}).$$

Conditioning on the initial history $\boldsymbol{h}_{it}$, a conditional likelihood for the regression parameters $\boldsymbol{\delta}$ in a generalised linear model is given by

$$L(\boldsymbol{\delta}) \propto \prod_{i=1}^{N} f(y_{iq}, \ldots, y_{iT} \mid y_{i0}, \ldots, y_{i(q-1)}) = \prod_{i=1}^{N} \prod_{t=q}^{T} f(y_{it} \mid \boldsymbol{h}_{it}) = \prod_{t=q}^{T} \prod_{i=1}^{N} f(y_{it} \mid \boldsymbol{h}_{it}), \tag{8}$$

which preserves the relation of stochastic dependence through the vector $\boldsymbol{h}_{it}$. To maximize the conditional likelihood function (8), a Newton-Raphson iterative method can be used that reduces to iteratively reweighted least squares, for regressing Y_{it}, $t = q, \ldots, T$, against the $(p+s)$ covariates $(\boldsymbol{x}_{it}, f_1^*(h_{it}), \ldots, f_s^*(h_{it}))$ in the GLM framework. As noted, this procedure is analogous to the estimation of GLMs for independent data, except for the fact that we now have additional parameters and terms in the linear predictor for the stochastic dependence. Diggle *et al.* (2002) give more details about this procedure, as well as its close relationship with the classic procedure for GLM. If all parameters in the model are non-stationary, i.e. vary over $t = q, \dots, T$, then $\boldsymbol{\delta_t}$ can be found from just the t-th factor in the full product likelihood (8), as was done by Anderson and Goodman (1957), but if any parameter components are stationary then it is necessary to use the full likelihood over all times $t = q, \dots, T$.

For the non-stationary case, differing from the proposal of Diggle *et al.* (2002), Ware *et al.* (1988) suggest to fit a model for each occasion as in a cross-sectional study. To fit a first-order transition model we need to add the response category at the preceding time as an additional covariate in the regression model, i.e. $\boldsymbol{x}_{it} = (x_{it1}, x_{it2}, \dots, x_{itp}, x_{it(p+1)})'$ represents the vector of $(p+1)$ covariates associated with the i-th individual at the t-th transition, and $x_{it(p+1)}$ is the previous state. For example, if $\tau = \{0, 1, 2, 3, 4, 5\}$ then there are 6 time occasions, which corresponds to 5 first-order transition models. For the first transition, we consider the response at time $t = 0$ as an additional covariate. In the second transition, the additional covariate is the response at time $t = 1$ and so on. Then, the vectors $\boldsymbol{\delta}_t$ are specified for each occasion and are obtained through the separate maximization of the likelihood functions, one for each transition. Ware *et al.* (1988) presented this theory, originally for ordinal data, but it is also applicable to nominal responses. For more general q-order dependent models we simply needs to incorporate the q previous responses as additional covariates. However, as we have noted above this requires conditioning on the first q transitions and is only feasible if T is sufficiently large. Also, rather that simply using the previous states as transitions we could consider functions of these f_r^*, $r = 1, \dots, s$, as in (7), which allows for more flexible modelling of potential time dependence in the transitions.

It is important to note that in this paper, when we discuss stationarity, we are referring to homogeneous transition probabilities matrices over time, otherwise it is understood that these matrices vary over time (hence the parameter notation with argument t). Thus, assuming that the process is stationary of first order, the fitted single model allows us to estimate, $\hat{\boldsymbol{\delta}}$, and by the invariance principle of likelihood we can estimate the transition probabilities matrix, $\widehat{\boldsymbol{P}}$. On the other hand, assuming no stationarity or non-homogeneity of transition probabilities over time, but still first order dependence, the T fitted models give esitimates $\hat{\boldsymbol{\delta}}(t)$, and therefore, allow us to estimate separate matrices of transition probabilities, $\widehat{\boldsymbol{P}}(t)$, $t = 1, 2, \ldots, T$.

Some classical references for the binary case are Cox (1970), Korn and Whittemore (1979), Azzalini (1983), Bonney (1987), Zeger and Liang (1992), Fitzmaurice and Laird (1993) and Heagerty (2002). For multinomial response (nominal or ordinal) the procedure is similar when we consider extensions of the GLM through generalized logits and proportional odds models (Ware *et al.*, 1988; Lee and Daniels, 2007; De Rooij, 2011). In both cases, the response of the i-th individual on the t-th occasion becomes a $(k \times 1)$ vector, $\boldsymbol{y}_{it} = (y_{it1}, y_{it2}, \ldots, y_{itk})'$, where $\{y_{itj}\}$ represent a set of index variables for the response categories, with $y_{itj} = 1$ if the i-th individual is in the j-th category at the time t, and $y_{itj} = 0$ otherwise, corresponding to a multivariate response.

In the following sections (2.2.2 and 2.2.3) we show the models that can usually be used with first order dependence, with obvious simple extension to higher order. Also, in the section 2.2.4, we present a new procedure to assess stationarity in which we do not use separate models for the assumption of non-stationarity, instead we incorporate into a single structure (linear predictor) additional parameters to explore the dependence, or not, with respect to t. However, in the simulation study, we use the two fitting approaches, since the classical test requires the fitted model under stationarity and the individual fitted models under non-stationarity. When necessary we use a partition of the explanatory vector $\boldsymbol{x}_{it}$ to separate the parts that refer only to the covariates, $\boldsymbol{x}^*_{it} = (x_{it1}, x_{it2}, \ldots, x_{itp})'$, and the previous response(s) giving $\boldsymbol{x}_{it} = (\boldsymbol{x}^*_{it}, x_{it(p+1)})'$. Also, note that we can write $\boldsymbol{x}_{it} = (\boldsymbol{x}^*_{it}, x_{it(p+1)})' = (\boldsymbol{x}^*_{it}, y_{(t-1)})'$, where $y_{(t-1)}$ is the observed response at time $t-1$.

2.2.2. Proportional odds transition model

For ordinal responses we can use the proportional odds model (McCullagh, 1980), that reduces the number of estimated parameters because it assumes the same $\boldsymbol{\delta}_t$ effects for each logit. The proportional odds transition model provides estimates of cumulative probabilities through:

$$\gamma_{ab}(t)(\boldsymbol{x}) = \frac{\exp(\lambda_{bt} + \boldsymbol{\delta}_t'\boldsymbol{x})}{1 + \exp(\lambda_{bt} + \boldsymbol{\delta}_t'\boldsymbol{x})}, \text{ with } b = 1, 2, \ldots, k-1 \text{ and } \quad t = 1, 2, \ldots, T, \tag{9}$$

where $\gamma_{ab}(t)(\boldsymbol{x}) = \mathrm{P}(Y_{jt} \leq b \mid Y_{a(t-1)})(\boldsymbol{x}) = \pi_{a1}(t)(\boldsymbol{x}) + \ldots + \pi_{ab}(t)(\boldsymbol{x})$ and when using the canonical link function,

$$\eta_t = \log\left(\frac{\gamma_{ab}(t)(\boldsymbol{x})}{1 - \gamma_{ab}(t)(\boldsymbol{x})}\right) = \lambda_{bt} + \boldsymbol{\delta}_t'\boldsymbol{x}, \tag{10}$$

in which λ_{bt} is an intercept (there will be one for each level of response), $\boldsymbol{x}$ is set of the covariates values, $\boldsymbol{\delta}_t' = (\beta_{t1}, \ldots, \beta_{tp}, \alpha_t)$ is the vector of the unknown parameters of interest and the index t is to denote the non-stationary process, in which there are T models of first-order. For stationary processes the structure of the model is the same but without the index t.

In this work model (10) is used to analyse and simulate data related to examples 1 and 2, in which we consider a first-order Markov chain and the following linear predictor:

$$\eta_t = \lambda_{bt} + [\beta_t \text{treatment} + \alpha_t \text{previous response}], \tag{11}$$

where treatment represents the "drug" or "enrichment" effect, as is the case. There are other possibilities for the linear predictor (11), including interaction terms, but these were not significant in this study.

2.2.3. Generalized logits transition model

For nominal responses, we can adapt the generalized logits model, that is useful to describe all logits of pairs of response categories with a common reference level (Agresti, 2012). Now, let $\boldsymbol{\delta}_{bt}' = (\beta_{bt1}, \ldots, \beta_{btp}, \alpha_{bt})$ be the vector of unknown parameters that is associated with the category b and let k be the reference response category. Then the generalized logits transition model is written as:

$$\eta_t = \log\left(\frac{\pi_{ab}(t)(\boldsymbol{x})}{\pi_{ak}(t)(\boldsymbol{x})}\right) = \lambda_{bt} + \boldsymbol{\delta}_{bt}'\boldsymbol{x}, \tag{12}$$

in which $b = 1, 2, \ldots, k-1$; $t = 1, 2, \ldots, T$ and λ_{bt} is an intercept as defined for model (9) but here the vector $\boldsymbol{\delta}'_{bt}$ varies with each category of response level as well as depending on the transition t.

Model (12) gives the effect of each covariate on the $(k-1)$ logits and the transition probabilities are given by:

$$\pi_{ab}(t)(\boldsymbol{x}) = \frac{\exp(\lambda_{bt} + \boldsymbol{\delta}'_{bt}\boldsymbol{x})}{1 + \sum_{b=1}^{k-1} \exp(\lambda_{bt} + \boldsymbol{\delta}'_{bt}\boldsymbol{x})}.$$

We used model (12) to analyse the simulated data-sets derived from example 3, with the following functional structure:

$$\eta_t = \lambda_{bt} + [\beta_{bt1}\text{pre-grazing} + \beta_{bt2}\text{post-grazing} + \alpha_{bt}\text{previous response}], \tag{13}$$

i.e., the linear predictor includes the effects of the factorial treatment structure (without interaction) and previous response.

2.2.4. The proposed test to assess stationarity

To apply the test proposed by Anderson and Goodman (1957) it is necessary to fit $(T+1)$ models, the T first order transition models under non stationarity and the transition model supposing stationarity. Moreover, the transition probabilities need to be estimated. In this paper, we propose another approach to assess stationarity in transition models. This strategy is a simple technique, since it can be done by analysing appropriate interaction parameters in the transition model using a stacked structure of the data and calculation of the transition probabilities, under stationarity and non-stationarity, is not required. Working with the conditional likelihood function (8), the idea consists of including an additional covariate for the transition time occasion in the linear predictor (7), i.e. $\boldsymbol{t}^* = (1, 2, \ldots, T)'$, and checking its interaction with other covariates, especially with the previous response. In this context, let

$$\eta_o = \log\left(\frac{\gamma_{ab}(\boldsymbol{x})}{1 - \gamma_{ab}(\boldsymbol{x})}\right) = \lambda_b + \boldsymbol{\delta}'\boldsymbol{x} \tag{14}$$

be the proportional odds transition model for ordinal data and

$$\eta_n = \log\left(\frac{\pi_{ab}(\boldsymbol{x})}{\pi_{ak}(\boldsymbol{x})}\right) = \lambda_b + \boldsymbol{\delta}'_b\boldsymbol{x}, \tag{15}$$

the generalized logits transition model for nominal data, both with longitudinal structure (stacked data). The inclusion of the indices o and n is to differentiate the ordinal and nominal cases, respectively.

Next, we consider models with additional terms in equations (14) and (15), giving nested models in both cases. Then, for the ordinal case, we can consider the following different models to reflect alternative hypotheses of time dependence :

$$\eta_{o(1)} = \lambda_b + \boldsymbol{\delta}'\boldsymbol{x} + \boldsymbol{\beta}^{*'}\boldsymbol{t}^*, \tag{16}$$

$$\eta_{o(2)} = \lambda_b + \boldsymbol{\delta}'\boldsymbol{x} + \boldsymbol{\beta}^{*'}\boldsymbol{t}^* + \boldsymbol{\gamma}'(\boldsymbol{t}^* : \boldsymbol{y}_{t-1}), \tag{17}$$

$$\eta_{o(3)} = \lambda_b + \boldsymbol{\delta}'\boldsymbol{x} + \boldsymbol{\beta}^{*'}\boldsymbol{t}^* + \boldsymbol{\vartheta}'(\boldsymbol{t}^* : \boldsymbol{x}^*), \tag{18}$$

$$\eta_{o(4)} = \lambda_b + \boldsymbol{\delta}'\boldsymbol{x} + \boldsymbol{\beta}^{*'}\boldsymbol{t}^* + \boldsymbol{\gamma}'(\boldsymbol{t}^* : \boldsymbol{y}_{t-1}) + \boldsymbol{\vartheta}'(\boldsymbol{t}^* : \boldsymbol{x}^*). \tag{19}$$

where the models (14), (16), (17) and (19) are nested, as also are the models (14), (16), (18) and (19). Equation (14) corresponds to the predictor of a model under stationarity, while the structures (16), (17), (18) and (19) are variations of this model, in which we incorporate possible sources of non-stationarity related to the dependence on the vector $\boldsymbol{t}^* = (1, 2, \ldots, T)'$. If the process is non-stationary, that is, the transition probabilities are non-homogeneous in time, then the parameters of model are not homogeneous in time, and there will be one or more causes of variations of these parameters under transitions. The sequences of nested models evaluate some possibilities.

In model (16) we evaluated the possibility of the non-stationarity cause due to only the variation of the vector of regression parameters ($\boldsymbol{\beta}^*$) over time, the effect of the Markov covariate is considered to be the same across all transitions. In model (18), we also considered the possibility of interaction of the covariates with transition time, again except for the Markov covariate. The variation of the Markov covariate over transition is evaluated through the inclusion of the interaction term, in the model (17). Therefore, in the models (17) and (18) there are two sources of non-stationarity. Model (19) is a more general (complete) form in which all sources of possible non-stationarity are considered.

For examples 1 and 2, we have that:

i. (14) corresponds to the stationary model with additive effects of treatment and previous response, which is referred to as "OTM0" (Ordinal transition model 0);

The models that incorporate sources of non-stationarity are:

ii. (16) corresponds to the additive effects of treatment, previous response and time factor, which is referred to as "OTM1" (Ordinal transition model 1);

iii. (17) corresponds to the additive effects of treatment, previous response and time factor and the interaction between the previous response and time factor, which is referred to as "OTM2" (Ordinal transition model 2);

iv. (18) corresponds to the additive effects of treatment, previous response and the time factor and the interaction between treatment and time factor, which is referred to as "OTM3" (Ordinal transition model 3);

v. (19) corresponds to the additive effects of treatment, previous response and the time factor as well as the interactions between previous response and time factor and treatment and time factor, which is referred to as "OTM4" (Ordinal transition model 4).

For nominal case, we define a similar set of models with the same interpretation as above, but now denoted the models for non-stationary processes by NTM1, NTM2, NTM3 and NTM4, with the stationary model denoted by NTM0 (15) and respective linear predictors with suffix n instead of o. Again, the models NTM0, NTM1, NTM2 and NTM4 are nested as well as the models NTM0, NTM1, NTM3 and NTM4. Reiterating that OTM0 and NTM0 correspond to stationary structures and the others form (OTM1, OTM2, OTM3, OTM4 and NTM1, NTM2, NTM3, NTM4) are non-stationary structures. In this context, let $\boldsymbol{\varphi} = (\boldsymbol{\beta}^*, \boldsymbol{\gamma}, \boldsymbol{\vartheta})$ and $\boldsymbol{\varphi_b} = (\boldsymbol{\beta}_b^*, \boldsymbol{\gamma}_b, \boldsymbol{\vartheta}_b)$ be the vectors of additional parameters for ordinal and nominal response, respectively. Then, based on equations (14) and (19), we can rewrite the hypotheses (2) for a global test for the ordinal case as:

$$\mathrm{H}_0 : (\boldsymbol{\delta}, \boldsymbol{\varphi}) = (\boldsymbol{\delta}, \mathbf{0}) \quad \text{and} \quad \mathrm{H}_1 : (\boldsymbol{\delta}, \boldsymbol{\varphi}) \neq (\boldsymbol{\delta}, \mathbf{0}), \tag{20}$$

and for the nominal case

$$\mathrm{H}_0 : (\boldsymbol{\delta}_b, \boldsymbol{\varphi}_b) = (\boldsymbol{\delta}_b, \mathbf{0}), \forall \quad b \in S \quad \text{and} \quad \mathrm{H}_1 : (\boldsymbol{\delta}_b, \boldsymbol{\varphi}_b) \neq (\boldsymbol{\delta}_b, \mathbf{0}), \text{for some} \quad b \in S. \tag{21}$$

For nominal response, the sum of the logarithms of the likelihood functions of T separate transitions, as well as the number of parameters, correspond exactly to the values for model NTM4. In the ordinal case, the structure of the proportional odds model does not allow for this identity to hold due to different cut points (thresholds), but the results are quite close, justifying the proposed method. Moreover, comparing model OTM0 (14) (or NTM0 (15)) with models (OTM1, OTM2, OTM3 and OTM4) (or NTM1, NTM2, NTM3, NTM4) corresponds to testing the stationary structure against the different potential non-stationary forms.

First, we can perform the global likelihood ratio test to evaluate hypotheses (20) and (21). If the null hypothesis is not rejected it is indicative that the process is stationary, since the transition probabilities will not change over time, the intrinsic assumption of stationarity for transition models. If the process is stationary, the likelihood test for the sequence of nested models will select models (14) or (15), because the additional parameters due to $\boldsymbol{t}^*$ will not significantly contribute to the log-likelihoods.

Here: Figure 1

However, when the process is non-stationary and we are led to the selection of models involving $\boldsymbol{t}^*$, in addition to the general time-varying structures, OTM4 and NTM4, there will be several possibilities to explore for how additional time-related parameters may influence the transition process. Local tests of the nested hypotheses are useful to select a parsimonious model with fewer parameters, especially for nominal responses where for each additional covariate we have $(k-1)$ additional parameters, and to try to identify the form/source of non-stationarity. Figure 1 represents this scheme for global and local tests.

2.2.5. Simulation study

Two sets of real data served as the basis for the initial simulation process (examples 1 and 3). To generate new data we started by using the estimates of the parameters and the probability transition matrices obtained from these examples. Note that from a non-stationary base, we can generate data for a stationary process, using only a single transition matrix. Also from a stationary base we may obtain data for a non-stationary process. We established some quantiles from the distribution of the test statistic proposed by Anderson and Goodman (1957) to select the first sets with different patterns of stationarity. To simulate data for the non-stationary process, we decided to work with an average level of non-stationarity, which approximately corresponds to the 25th percentile.

From these first sets we then implemented new simulations to get new categorical data (ordinal and nominal) under two scenarios: stationary and non-stationary processes, for which we varied the number of time occasions, T=4, T=5 and T=6 as well as the sample size, N= 100, N=200 N=500 and N=1000. For each scenario we performed 10,000 simulations. In all cases we considered a first-order Markov chain and for the ordinal case, the linear predictor included the effects of treatment and previous response, as described by the equations (11) and (13).

After that, the proposed test as described in Section 2.2.4 and the classical test (Anderson and Goodman, 1957) based on equation (3) were applied. The degree of agreement between the tests was assessed by means of correlation measures. Additionally, for each scenario, we computed the rejection rates of the tests for significance levels of 1%, 5%, and 10% in order to assess type-I and type-II error rates. The computational implementation was made using the `R` system (R Core Team, 2015), with the aid of packages `nnet` (Ripley and Venables, 2016) and `ordinal` (Christensen, 2011) to fit transition models and the package `markovchain` (Spedicato, 2015) to assist in the data simulation process.

3. Results and discussion

3.1. Respiratory data analysis

Figure 2 shows the observed transition frequencies of individuals on respiratory condition with 5 states: terrible, bad, moderate, good and excellent. At each time occasion there are 111 individuals and the total number of first-order transitions is 444. This indicates that there is an increase in the number of individuals in terrible, good and excellent condition.

Here: Figure 2

Table 1 shows the models that were used to assess stationarity in the first example, with the respiratory data. The structure η_o, with the effects of drug and previous response was selected at a significance level of 1%. As this study involved 5 times, there are 4 transitions of first order,with a sum of log-likelihoods of -491.06 (on 36 degrees of freedom), quite close the value -494.44 for the combined single general non-stationary model OTM4. The result for the global likelihood-ratio test is 25.47 on 18 degrees of freedom, and is not significant (p=0.1124). In fact, the classical test statistic (Anderson and Goodman, 1957) for this example is 36.55, on 27 degrees of freedom is also not significant (p=0.1037). The structure η_o was selected by local tests applied in the upper and lower parts of Table 1, in both directions (forward or backward), using a 5% significance level.

Here: Table 1

The results show that in this example the process is stationary, i.e., the transition probabilities are homogeneous over time. Table 2 shows the parameters estimates, standard errors and p-values of the first order stationary transition model in this respiratory condition study. Apart from the intercepts, all parameters are significant in this model.

Here: Table 2

Note the increasing values of α. This result is very common in transition models, in which the previous response is more important to explain the transition of the individuals (Diggle *et al.*, 2002). Using the coefficients of the parameters available in Table 2, we can estimate the transition probability matrices for the groups Active (A) and Placebo (P), that are shown in Figure .

Here: Figure 3

The first row from bottom to top of the Figure 3 describes the transition probabilities from state 1 (terrible) for the conditions: 1 (terrible), 2 (bad), 3 (moderate), 4 (good) and 5 (excellent). The second line describes the transition probabilities from state 2 (bad) for the others conditions and so until the last line that describes transitions from state 5 for others. As the effect of treatment is significant (see Table 2) , the matrices are statistically different. In general, the transition probabilities for states good and excellent (two last columns) are more favorable for the active treatment group.

3.2. Pig behaviour data analysis

Figure 4 illustrates the frequency of animals in the states 1 (absence), 2 (moderate) and 3 (serious), on each occasion of the study period. On each time occasion there are 124 animals and the total number of first-order transitions is 372. A drop in the frequency of animals with serious lesions over time was observed.

Here: Figure 4

Table 3 shows the sequences of nested models for the second example, on pig behaviour. There are 4 times and 3 transitions of first order, with log-likelihood sum of -317.90 (on 15 degrees of freedom), close of the log-likelihood for model OTM4 that is -322.71 (on 13 degrees of freedom). The statistic for the global likelihood-ratio test is 82.34 on 8 degrees of freedom is significant ($p < 0.01$). This result is in agreement of the classical test of Anderson and Goodman (1957), whose statistic is 78.32, on 10 degrees of freedom ($p < 0.01$). Therefore, the process is non-stationary and OTM0 was not selected. However, this result does not mean that the OTM4 is the best model. Note that only the inclusion of a time factor is significant in the local tests, and model ATM1 gives a more parsimonious non-stationary structure. In some applications we will require the full model OTM4, but this can be verified by performing the step-by-step local tests for specific aspects of time dependence..

Here: Table 3

Specifically, since the transitions in the pig behavior study are not homogeneous over time, we select the linear predictor structure through the local tests. From the upper part of Table 3, the first local test comparing for OTM4 vs. OTM2, is significant (L.R. statistic = 8.88 on 2 d.f. and p = 0.0117). Subsequently, a test of OTM4 vs. OTM1, is not significant for the interactions with the time factor (L.R. statistic = 10.15 on 6 d.f. and p = 0.1185). When we compare the models OTM1 vs. OTM0, the result is significant for the time factor (L.R. statistic = 62.56 on 2 d.f, p< 0.001), i.e., the model OTM1 is selected in the first part a the simplest parsimonious model.

From the lower part of Table 3, comparing OTM4 vs. OTM3, is not significant (L.R. statistic = 0.7012 on 4 d.f. and p = 0.9512). Finally, comparing OTM1 vs. OTM3, the likelihood ratio local tests is significant (L.R. statistic = 9.44 on 2 d.f and p = 0.0088). Therefore the selected final linear predictor, that represents non-stationary process in the pig behaviour study is the model OTM3, that includes the additive effects of treatment and previous response and the interaction between treatment and time factor. The same model would be established if we used the forward direction.

Here: Table 4

Table 4 shows the estimates for the parameters of the selected non-stationary model (OTM3). Note that the effects of the time factor, interaction between treatment (enrichment) and time factor and previous response are significant. It means that previous response has a strong influence on pig behaviour and, under transition (time factor) the environment enrichment is important as well. In fact, the enrichment effect is significant in the second transition. According to Castro (2016) there is an explanation for this: At the beginning of the experiment the pigs were learning to play with the objects and the end of study they lost interest. The transition probability matrices are given in Figure .

The first rows from bottom to top of the Figure 5 describe the transition probabilities from state 1 (absence of lesions) for the conditions: 1 (absence of lesions), 2 (moderate) and 3 (serious). The second line describes the transition probabilities from state 2 (moderate) to the other conditions and the third line describes transitions from state 3 to others. Focussing attention on the first column of these matrices, at the second and third transitions, it is possible to note that the transition probabilities for the state “absence of lesions” are higher for the treated group (with environment enrichment).

Here: Figure 5

3.3. Agronomic data analysis

Figure 6 shows the frequency of points in the three categories: tussocks, bare ground and weeds, on each time occasion. There are 600 points and the total number of first-order transitions is 3000. The number of units classified as "bare ground" is greater than those classified "tussocks" and "weeds".

Here: Figure 6

The structure of the data appears to have a pattern with few changes of state, in contrast to the two first examples. This is not a serious problem for the simulation process, because it is still possible to get new data with different patterns of stationarity and dependence. In Table 5 we present the sequences of nested models for this motivational example with simulated data. In this study, there are 5 transitions of first order, whose sum of the log-likelihoods is -2163.83 (on 50 degree of freedom), whose value is the same for model $\eta_{o(4)}$, as here we have no cut-points to be estimated and so both forms of the full model are the equivalent. Thus, the result for the global likelihood-ratio test (η_o versus $\eta_{o(4)}$) is 43.28 (p=0.3330). Here, the inclusion of time is not significant in all possible sequences (local tests), which guides the choice of the stationary structure, η_o, with the effects of treatment and previous response. Moreover, the classical test statistic of Anderson and Goodman (1957) is 43.17, on 40 degrees of freedom and also not significant (p=0.3371).

Here: Table 5

Finally, Table 6 shows the parameter estimates for the first order stationary transition model for the agronomic data. The effects of treatment and previous response were significant.

Here: Table 6

Here: Figure 7

The transition matrices for each combination of pre-grazing and post-grazing are shown in Figure 7. For each combination, the space was occupied by the types 1 (tussocks) and 2 (bare ground), and the transition probability for it coming to be occupied by 3 (weeds) is smaller. Indeed spaces occupied by vegetation 1 (tussocks) as well as those occupied by 3 (weeds) have a greater probability of moving to condition 2 (bare ground).

3.4. Results from simulation study

Next we present the results of the simulation studies. We start by discussing the number of times each model was selected in all scenarios, using the local tests with the significance level of 5%. Also, we consider the two possible sequences of nested models as presented in Section 2.2.4, i.e., (OTM0, OTM1, OTM2, OTM4) or (OTM0, OTM1,OTM3, OTM4) for the ordinal case, and, (NTM0, NTM1, NTM2, NTM4) or (NTM0, NTM1, NTM3, NTM4) for the nominal case.

For the stationary scenario, for ordinal and nominal data, in more than 80% of the simulations, OTM0 or NTM0 were selected in both sequences. There is an effect of sample size and the number of occasions because as they increase this percentage increases to values close to 85%. In contrast, for the non-stationary scenario, smaller percentages of selection of OTM0 and NTM0 were observed, specifically, for $N = 500$ or $N = 1000$, where these numbers are less than 0.07%. Also, for these sample sizes the highest percentages of selection were to OTM2 and OTM4 (or OTM3 and OTM4) for ordinal response and NTM2 and NTM4 (or NTM3 and NTM4) for nominal response, corresponding to more than 95% of selections. For $N = 200$, less than 22% of the simulations selected the models OTM0 or NTM0. Finally, for $N = 100$, we observed the highest percentages of OTM0 or NTM0, but smaller than 48%, and these were considered within the usual variation expected for the simulation process with a moderate degree of non-stationarity.

In fact, as already mentioned in section 2.2.5, our simulations for the non-stationary scenario were made with a moderate degree of non-stationarity in order to assess the performance of the test in not-so-favourable conditions, in contrast with the situation of motivational example 2, in which the degree of non-stationarity is very high. Under this condition we would have inevitably a power function close to 1, even in small samples and $T = 4$. As an illustration, in the scenario with an ordinal response and a strong degree of non-stationarity (10,000 simulations, $N = 100$ and $T = 4$), we had the numbers 736, 83, 2163 and 7018 for models OTM0, OTN1, OTM2, and OTM4, and 705, 75, 2251 and 6969 for models OTM0, OTN1, OTM3, and OTM4, respectively.

Here: Table 7

Table 7 shows the rejection rates for the classical test and the proposed tests (global and local tests) for scenario 1 (assuming stationarity) with ordinal and nominal data. The global likelihood ratio test involves the models OTM0 and OTM4 for ordinal response (or NTM0 and NTM4 for nominal response) and we present three local tests: (1) that involving models OTM0 and OTM1 for ordinal response (or NTM0 and NTM1 for nominal response), (2) that for OTM0 and OTM2 for ordinal response (or NTM0 and NTM2 for nominal response) and (3) that for OTM0 and OTM3 for ordinal response (or NTM0 and NTM3 for nominal response). In this case it is possible to study the test size (type-I error). For nominal responses, the classical test was a little more conservative than the global test. However, for some local tests, especially the local test 1, we observe lower nominal levels than the classical test. Also, as the samples size increases, all tests tend to maintain the level of significance, i.e., they have equivalent performance asymptotically. On the other hand, for ordinal responses, in most cases, the proposed tests are more conservative than the classical test.

Here: Table 8

Table 8 shows the rejection rates for scenario 2 (non-stationarity process). It was possible to study the power of each test. There was clearly an effect of sample size on the power of the tests. Note, for example, at $T = 4$ and $N = 100$, in both scenarios, there is a higher propensity to type-II error, but it decreases as we consider larger sample sizes and/or more time occasions. Specifically, in the ordinal case, for $N = 100$ and $N = 200$, the power of the global test is smaller than the classical one, a likely consequence of the different handling of cut-point estimation for the ordinal scale. On the other hand, in the nominal case, the power of the proposed global test is greater than the classical, but, asymptotically, they are equivalent. From the Table 8 it is possible to notice that local tests also have a good power function, in some cases very close to their global competitors, which shows their efficiency.

We also consider agreement in terms of correlation between the test statistics (classical and global), as measured by the usual Pearson correlation coefficient. For ordinal responses, the correlation values increased with the values of N (sample size) in both scenarios (stationary or not), all correlations were greater than 0.80 at $T = 4$ and $T = 5$ and all correlations were larger than 0.77 at $T = 6$. Also, for nominal responses, the correlation values increased with the values of N (sample size) in both scenarios (stationary or not), and all correlations were larger than 0.97 for all time occasions. It shows there is a strong association between the tests.

4. Conclusion

The assumption of stationarity is important in the use of transition models. When it is satisfied the model is simpler and therefore, there are fewer parameters. However, sometimes, this is not true. In this article we presented an alternative method to assess stationarity in these models for a categorical response. The procedure has been illustrated with three applications, one being a study with a nominal response. Our goal was to show that the procedure is very simple, composed of local and global tests, applied to a nominal or ordinal response but without the necessity of computing the transition probabilities matrices. Also, there are some advantages of the proposed procedure: it can be carried out in a single modelling framework using the stacked form of the data and, therefore, demands less computational effort and it can be applied to evaluate stationarity in longer-range chains. Also, it allows categorical and continuous covariates and the local tests can be used for the selection of a linear predictor that corresponds to a specific non-stationary model.

We verified that when the process is stationary, local tests for the inclusion of an additional covariate for time are not significant. In this case, the global test also selects the stationary model. However, if the process is non-stationary, some local tests are significant for additional parameters, showing that the inclusion of time is important and the global test rejects the null hypothesis. There is a difference between the nominal and ordinal case, because in the latter the number of parameters involved changes a little for global test, but this is not a problem, it is a consequence of the use of the proportional odds model with estimated cut-points.

It is important to note that both tests are valid asymptotically. However, in some situations, for example, with small sample sizes or a large amount of missing data, the convergence of the model fitting can be difficult, especially in non-stationary cases that involve more parameters. For the classical test, the problem is greater because the successive stratification leads to sparse tables and this test cannot applied.

The simulation studies showed that the proposed test presented good performance with the control of type-I and type-II error rates and the results were quite close to the classical test available in the literature (Anderson and Goodman, 1957). It is noteworthy that the studies for scenario 2 were carried out with a moderate degree of non-stationarity, since we wanted to assess the proposed test under not-so-favourable conditions. With a high degree of non-stationarity, the proposed test has the best performance, even for small sample sizes.

Acknowledgment

This work was supported by São Paulo Research Foundation (FAPESP), Brazil, grant number $2015/02628 - 2$.

Bibliography

Agresti, A. (2012). *Categorical Data Analysis*. New York: Wiley 3rd ed.

Anderson, T.W.; Goodman, L.A. (1957). Statistical Inference about Markov Chains. *Annals of Mathematical Statistics*. Ann Arbor, **28**, 89–110.

Azzalini, A. (1983). Maximum likelihood estimation of order n for stationary stochastic processes. *Biometrika*, **70**, 381–388.

Bonney, G. E.(1987). Logistic Regression for dependent binary observations. *Biometrics*, **43**, 951–973.

Castro, A.C. (2016). *Comportamento e desempenho sexual de suínos reprodutores em ambientes enriquecidos*, PhD. dissertation. Brazil: University of São Paulo.

Christensen,R.H. (2011). *Analysis of ordinal data with cumulative link models estimation with the R-package ordinal*. Available from: `http://www.R-project.org`.

Cox, D.R. (1970). *The analysis of binary data*. London: Methuen.

De Rooij, M. (2011). Transitional ideal point models for longitudinal multinomial outcomes. *Statistical Modelling*, **11(2)**, 115–135.

Diggle, P.J.; Heagerty, P.J.; Liang, K.Y.; Zeger, S.L. (2002). *Analysis of longitudinal data*. New York: Oxford University Press.

Fitzmaurice, G.M.; Laird, N.M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, **80**, 141–151.

Good, I.J. (1955). The likelihood ratio test for Markov chains. *Biometrika*, **42**, 531–533.

Goodman, L.A. (1962). Statistical Methods for analysing processes of change. *American Journal of Sociology*, **68**, 57–78.

Heagerty, P.J. (2002). Marginalized transition models and likelihood inference for longitudinal data. *Biometrics*, **58**, 342-351.

Jones, P.W.; Smith P. (2001). *Stochastic Processes*. Arnold: London.

Koch, G.C.; Carr, G.J; Amara, I.A; Stokes, M.E.; Uryniak, T.J. Categorical Data Analysis. In: BERRY, D.A. (1990). *Statistical Methodology in the Pharmaceutical Sciences.* New

York: Marcel Dekker, **13**, 389–473.

Korn, E.L.; Whittemore, A.S. (1979). Methods for analysing panel studies of acute health effects of air pollution. *Biometrics*, **35**, 715–802.

Lee, K.; Daniels, M. J. (2007). A class of Markov models for longitudinal ordinal data. *Biometrics*, **63**, 1060–1067.

Lindsey, J.K. (1995). *Modelling frequency and count data.* Oxford: Oxford University Press.

Lindsey, J.K. (2004). *Statistical analysis of stochastic processes in time.* New York: Cambridge University Press.

McCullagh, P. (1980). Regression Methods for Ordinal Data. *Journal of The Royal Statistical Society.* **42**, 109–142.

Molenberghs, G. Verbeke, G. (2005).*Models for discrete longitudinal data.* New York: Springer-Verlag.

Pereira, L.E.T.; Paiva, A.J.; Geremia, E.V.; da Silva, S.C. (2015a). Grazing management and tussock distribution in elephant grass. *Grass and Forage Science*,**70(3)**, 406–417.

Pereira, L.E.T.; Paiva, A.J.; Geremia, E.V.; da Silva, S.C. (2015b). Regrowth patterns of elephant grass (*Pennisetum purpureum Schum*) subjected to strategies of intermittent stocking management. *Grass and Forage Science*, **70(1)**, 195–204.

R Development Core Team. (2015). *R: A language and environment for statistical computing 3.2.* Vienna, Austria. Available from: `http://www.R-project.org`

Ripley, B., Venables, W. (2016). Package `nnet`: Feed-forward neural networks and multinomial log-linear models. Available from: `http://www.R-project.org/`

Spedicato, G.A. (2015). `markovchain`: discrete time Markov chains made easy. Available from: `http://www.R-project.org/`

Stirzaker D. (2005). **Stochastic Processes and Models**. Oxford.

Ware, J. H.; Lipsitz, S.; Speizer, F. E. (1988). Issues in the Analysis of Repeated Categorical Outcomes. *Statistics in Medicine*,**7**, 95–107.

Zeger, S.L.; Liang, K.Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*, **II**, 1825–1839.

Tables

Table 1: Analysis of the nested models to assess stationarity in the respiratory study.

Models	Log-likelihood	Number of parameters	D.F.	p-value
η_o (OTM0)	-507.17	9		
$\eta_{o(1)}$ (OTM1)	-504.72	12	3	0.1782
$\eta_{o(2)}$ (OTM2)	-500.06	24	12	0.6756
$\eta_{o(4)}$ (OTM4)	-494.44	27	3	0.0104
η_o (OTM0)	-507.17	9		
$\eta_{o(1)}$ (OTM1)	-504.72	12	3	0.1782
$\eta_{o(3)}$ (OTM3)	-499.20	15	3	0.0115
$\eta_{o(4)}$ (OTM4)	-494.44	27	12	0.6581

Table 2: Parameter estimates of the proportional odds stationary transition model of first order fitted to the respiratory condition study data.

Parameters	Estimates	Standard errors	p-value
λ_2	0.9834	0.5137	
λ_3	2.3490	0.5455	
λ_4	4.3310	0.5634	
λ_5	5.7059	0.5731	
β (Placebo)	-0.6550	0.1899	0.0005
$\alpha(2)$	3.1353	0.5781	< 0.001
$\alpha(3)$	4.1805	0.5579	< 0.001
$\alpha(4)$	5.4542	0.5773	< 0.001
$\alpha(5)$	6.9298	0.5933	< 0.001

Table 3: Analysis of the nested models to assess stationarity in the pig behavior study.

Models	Log-likelihood	Number of parameters	D.F.	p-value
η_o (OTM0)	-359.07	5		
$\eta_{o(1)}$ (OTM1)	-327.79	7	2	< 0.001
$\eta_{o(2)}$ (OTM2)	-327.16	11	4	0.8680
$\eta_{o(4)}$ (OTM4)	-322.71	13	2	0.0117
η_o (OTM0)	-359.07	5		
$\eta_{o(1)}$ (OTM1)	-327.79	7	2	< 0.001
$\eta_{o(3)}$ (OTM3)	-323.07	9	2	0.0088
$\eta_{o(4)}$ (OTM4)	-322.71	13	4	0.9511

Table 4: Parameter estimates and standard errors (s.e.) for the three first order proportional odds transition models, under a non-stationary process, fitted to the pig behaviour data.

Parameters	Estimates	Standard errors	p-value
λ_2	-1.2811	0.3485	
λ_3	1.5811	0.3462	
β(E2)	-0.0120	0.3595	0.9733
$\beta^*(t^*2)$	-2.8284	0.3958	< 0.001
$\beta^*(t^*3)$	-1.6235	0.3864	< 0.001
$\alpha(2)$	0.6820	0.2829	0.0159
$\alpha(3)$	1.3013	0.3303	< 0.001
ϑ(E2:t^*2)	1.5462	0.5284	0.0034
ϑ(E2:t^*3)	0.3596	0.5041	0.4755

Table 5: Analysis of the nested models to assess stationarity in the agronomic study.

Models	Log-likelihood	Number of parameters	D.F.	p-value
η_o (NTM0)	-2185.47	10		
$\eta_{o(1)}$ (NTM1)	-2180.35	18	8	0.2491
$\eta_{o(2)}$ (NTM2)	-2169.51	34	16	0.1538
$\eta_{o(4)}$ (NTM4)	-2163.83	50	16	0.7860
η_o (NTM0)	-2185.47	10		
$\eta_{o(1)}$ (NTM1)	-2180.35	18	8	0.2491
$\eta_{o(3)}$ (NTM3)	-2174.65	34	16	0.7442
$\eta_{o(4)}$ (NTM4)	-2163.83	50	16	0.1547

Table 6: Parameter estimates for the generalized logits transition model of first order fitted to the agronomic data.

Parameters	Estimates	Standard errors	p-value
λ_2	1.4053	0.1851	< 0.001
λ_3	2.0842	0.1792	< 0.001
β_{21} (pre2)	0.3706	0.1678	0.0272
β_{31} (pre2)	0.1055	0.1648	0.5220
β_{22} (post2)	0.6751	0.1720	< 0.001
β_{32} (post2)	0.5129	0.1691	0.0024
$\alpha_2(2)$	-0.0217	0.1920	0.9097
$\alpha_2(3)$	-0.2662	0.1879	0.1566
$\alpha_3(2)$	-0.6557	0.2343	0.0051
$\alpha_3(3)$	-1.0512	0.2303	< 0.001

Table 7: Rejection rates for the classical and proposed tests, resulting from 10,000 simulations, for the scenario 1 (test size).

Time		T=4			T=5			T=6		
Level		10%	5%	1%	10%	5%	1%	10%	5%	1%
					Ordinal Data					
Sample Size	Tests									
	Classical	0.1396	0.0799	0.0181	0.1333	0.0724	0.0194	0.1416	0.0784	0.0199
	Local 1	0.1009	0.0491	0.0100	0.1022	0.0482	0.0099	0.0981	0.0522	0.0094
100	Local 2	0.1142	0.0581	0.0116	0.1079	0.0545	0.0088	0.1065	0.0566	0.0109
	Local 3	0.1067	0.0569	0.0129	0.1064	0.0540	0.0108	0.1029	0.0525	0.0111
	Global	0.1221	0.0625	0.0139	0.1137	0.0559	0.0121	0.1176	0.0604	0.0143
	Classical	0.1167	0.0599	0.0117	0.1227	0.0652	0.0135	0.1258	0.0636	0.0137
	Local 1	0.0970	0.0503	0.0099	0.1048	0.0522	0.0090	0.1023	0.0526	0.0118
200	Local 2	0.1089	0.0566	0.0111	0.1152	0.0599	0.0116	0.1193	0.0603	0.0121
	Local 3	0.1004	0.0518	0.0095	0.1055	0.0515	0.0088	0.1016	0.0522	0.0112
	Global	0.1110	0.0544	0.0119	0.1150	0.0599	0.0130	0.1213	0.0622	0.0120
	Classical	0.1043	0.0534	0.0109	0.1101	0.0528	0.0118	0.1101	0.0581	0.0131
	Local 1	0.1008	0.0517	0.0110	0.1007	0.0517	0.0118	0.0978	0.0490	0.0103
500	Local 2	0.1028	0.0509	0.0113	0.1037	0.0548	0.0118	0.0103	0.0438	0.0118
	Local 3	0.1029	0.0519	0.0097	0.1038	0.0524	0.0113	0.1037	0.0505	0.0084
	Global	0.1006	0.0494	0.0108	0.1054	0.0555	0.0102	0.1044	0.0551	0.0122
	Classical	0.1032	0.0525	0.0113	0.1099	0.0573	0.0115	0.1039	0.0501	0.0095
	Local 1	0.0995	0.0535	0.0107	0.1032	0.0534	0.0122	0.0993	0.0470	0.0087
1000	Local 2	0.1061	0.0538	0.0108	0.1073	0.0537	0.0105	0.1032	0.0505	0.0091
	Local 3	0.0988	0.0478	0.0094	0.1030	0.0533	0.0110	0.1025	0.0490	0.0094
	Global	0.1024	0.0484	0.0112	0.1104	0.0557	0.0103	0.1055	0.0504	0.0091
					Nominal Data					
Sample Size	Tests									
	Classical	0.1359	0.0744	0.0166	0.1253	0.0663	0.0145	0.1374	0.0751	0.0190
	Local 1	0.0898	0.0452	0.0102	0.1124	0.0563	0.0116	0.0788	0.0371	0.0068
100	Local 2	0.1486	0.0795	0.0160	0.1400	0.0732	0.0156	0.1401	0.0711	0.0155
	Local 3	0.1218	0.0674	0.0152	0.1185	0.0657	0.0134	0.1214	0.0640	0.0129
	Global	0.1630	0.0910	0.0206	0.1436	0.0810	0.0188	0.1675	0.0894	0.0212
	Classical	0.1253	0.0663	0.0145	0.1251	0.0660	0.0148	0.1337	0.0688	0.0166
	Local 1	0.1124	0.0563	0.0116	0.1052	0.0522	0.0097	0.1029	0.0520	0.0097
200	Local 2	0.1400	0.0732	0.0156	0.1358	0.0717	0.0158	0.1436	0.0803	0.0167
	Local 3	0.1185	0.0657	0.0134	0.1200	0.0632	0.0133	0.1218	0.0645	0.0146
	Global	0.1436	0.0810	0.0188	0.1456	0.0769	0.0185	0.1578	0.0844	0.0201
	Classical	0.1034	0.0500	0.0113	0.1108	0.0570	0.0112	0.1163	0.0590	0.0112
	Local 1	0.1022	0.0525	0.0094	0.0989	0.0495	0.0097	0.1051	0.0529	0.0109
500	Local 2	0.1067	0.0555	0.0119	0.1103	0.0564	0.0118	0.1147	0.0571	0.0110
	Local 3	0.1030	0.0504	0.0113	0.1067	0.0544	0.0103	0.1122	0.0584	0.0123
	Global	0.1092	0.0544	0.0123	0.1162	0.0613	0.0124	0.1225	0.0616	0.0122
	Classical	0.1017	0.0472	0.0082	0.1031	0.0509	0.0096	0.1047	0.0534	0.0115
	Local 1	0.0989	0.0500	0.0098	0.1016	0.0494	0.0093	0.1025	0.0521	0.0113
1000	Local 2	0.1042	0.0532	0.0093	0.1019	0.0503	0.0099	0.1085	0.0572	0.0127
	Local 3	0.0975	0.0482	0.0074	0.1033	0.0521	0.0097	0.1098	0.0519	0.0098
	Global	0.1039	0.0495	0.0083	0.1062	0.0532	0.0097	0.1074	0.0545	0.0115

Table 8: Rejection rates for the classical (1) and proposed (2) tests, resulting from 10,000 simulations, for the scenario 2 (test power).

Time		T=4			T=5			T=6		
Level		10%	5%	1%	10%	5%	1%	10%	5%	1%
					Ordinal Data					
Sample Size	Tests									
	Classical	0.7257	0.6090	0.3636	0.8079	0.7094	0.4792	0.8403	0.7426	0.5086
	Local 1	0.2329	0.1418	0.0438	0.3935	0.2795	0.1135	0.1400	0.0749	0.0177
100	Local 2	0.5226	0.3911	0.1690	0.6682	0.5433	0.2986	0.5634	0.4217	0.1958
	Local 3	0.5287	0.4035	0.1968	0.5443	0.4214	0.2132	0.2164	0.1339	0.0413
	Global	0.6981	0.5755	0.3287	0.7528	0.6389	0.3898	0.6133	0.4704	0.2263
	Classical	0.9670	0.9386	0.8162	0.9881	0.9734	0.9117	0.9959	0.9907	0.9546
	Local 1	0.3793	0.2593	0.1032	0.6462	0.5315	0.3037	0.1890	0.1118	0.0279
200	Local 2	0.8513	0.7652	0.5438	0.9524	0.9097	0.7696	0.9210	0.8633	0.6773
	Local 3	0.8210	0.7230	0.4913	0.8432	0.7563	0.5445	0.3579	0.2362	0.0904
	Global	0.9596	0.9214	0.7899	0.9793	0.9548	0.8642	0.9448	0.8946	0.7366
	Classical	1.0000	0.9999	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Local 1	0.7021	0.5829	0.3397	0.9548	0.9158	0.7915	0.3329	0.2296	0.0757
500	Local 2	0.9987	0.9962	0.9828	1.0000	1.0000	0.9998	1.0000	0.9999	0.9991
	Local 3	0.9963	0.9913	0.9585	0.9976	0.9935	0.9749	0.7144	0.5942	0.3409
	Global	1.0000	0.9999	0.9994	1.0000	1.0000	0.9999	1.0000	1.0000	0.9998
	Classical	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Local 1	0.9291	0.8754	0.7091	0.9993	0.9979	0.9885	0.5722	0.4336	0.1984
1000	Local 2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Local 3	1.0000	1.0000	0.9998	1.0000	1.0000	1.0000	0.9605	0.9261	0.7938
	Global	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
					Nominal Data					
Sample Size	Tests									
	Classical	0.4654	0.3344	0.1423	0.5655	0.4324	0.2101	0.6586	0.5218	0.2815
	Local 1	0.2115	0.1257	0.0374	0.1799	0.1010	0.0243	0.2313	0.1441	0.0454
100	Local 2	0.3898	0.2605	0.0945	0.4860	0.3445	0.1408	0.6170	0.4796	0.2369
	Local 3	0.4046	0.2781	0.1075	0.4038	0.2819	0.1066	0.4430	0.3076	0.1250
	Global	0.5355	0.3953	0.1811	0.6291	0.4933	0.2485	0.7248	0.5960	0.3360
	Classical	0.7932	0.6876	0.4465	0.8833	0.8080	0.5990	0.9518	0.9092	0.7635
	Local 1	0.3920	0.2771	0.1186	0.3364	0.2308	0.0871	0.4918	0.3679	0.1648
200	Local 2	0.6469	0.5189	0.2801	0.7654	0.6499	0.4136	0.9124	0.8459	0.6530
	Local 3	0.6881	0.5634	0.3255	0.7100	0.5847	0.3430	0.7755	0.6679	0.4244
	Global	0.8252	0.7286	0.4961	0.9006	0.8305	0.6420	0.9635	0.9291	0.8043
	Classical	0.9972	0.9940	0.9709	1.0000	0.9994	0.9961	1.0000	1.0000	1.0000
	Local 1	0.7321	0.6203	0.3930	0.6684	0.5436	0.3131	0.8733	0.7944	0.5944
500	Local 2	0.9616	0.9276	0.8026	0.9934	0.9840	0.9378	0.9996	0.9995	0.9963
	Local 3	0.9808	0.9622	0.8773	0.9877	0.9735	0.9062	0.9966	0.9915	0.9586
	Global	0.9971	0.9947	0.9735	0.9999	0.9995	0.9964	1.0000	1.0000	1.0000
	Classical	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Local 1	0.9614	0.9262	0.8088	0.9361	0.8844	0.7197	0.9948	0.9885	0.9519
1000	Local 2	1.0000	0.9994	0.9956	1.0000	1.0000	0.9999	1.0000	1.0000	1.0000
	Local 3	1.0000	1.0000	0.9990	1.0000	1.0000	0.9999	1.0000	1.0000	1.0000
	Global	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000