Received: 16 December 2019 | Revised: 18 May 2020 | Accepted: 9 June 2020

DOI: 10.1002/bimj.201900394



Biometrical Journal

RESEARCH PAPER

# A generalized transition model for grouped longitudinal categorical data

**Idemauro A. R. Lara**[1] | **Rafael A. Moral**[2] | **Cesar A. Taconeli**[3] | **Carolina Reigada**[4] | **John Hinde**[5]

[1] Department of Exact Sciences, University of São Paulo, Piracicaba, Brazil

[2] Department of Mathematics and Statistics, Maynooth University, Maynooth, Ireland

[3] Department of Statistics, Federal University of Paraná, Curitiba, Brazil

[4] Departament of Ecology and Evolutionary Biology, Federal University of São Carlos, São Carlos, Brazil

[5] School of Mathematics, Statistics, and Applied Mathematics, NUI Galway, Galway, Ireland

**Correspondence**
Rafael A. Moral, Department of Mathematics and Statistics, Maynooth University, Maynooth, Ireland.
Email: rafael.deandrademoral@mu.ie

**Funding information**
Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: 18/06288-0

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

**Abstract**

Transition models are an important framework that can be used to model longitudinal categorical data. They are particularly useful when the primary interest is in prediction. The available methods for this class of models are suitable for the cases in which responses are recorded individually over time. However, in many areas, it is common for categorical data to be recorded as groups, that is, different categories with a number of individuals in each. As motivation we consider a study in insect movement and another in pig behaviou. The first study was developed to understand the movement patterns of female adults of *Diaphorina citri*, a pest of citrus plantations. The second study investigated how hogs behaved under the influence of environmental enrichment. In both studies, the number of individuals in different response categories was observed over time. We propose a new framework for considering the time dependence in the linear predictor of a generalized logit transition model using a quantitative response, corresponding to the number of individuals in each category. We use maximum likelihood estimation and present the results of the fitted models under stationarity and non-stationarity assumptions, and use recently proposed tests to assess non-stationarity. We evaluated the performance of the proposed model using simulation studies under different scenarios, and concluded that our modeling framework represents a flexible alternative to analyze grouped longitudinal categorical data.

**KEYWORDS**

discrete stochastic process, generalized logit models, multinomial distribution, tests for stationarity

## 1 | INTRODUCTION

Discrete data are very common in experiments where the focus of interest is the occurrence of an event that is recorded as a count. Depending upon the focus of the study, the counts may be classified according to certain specific response categories. When one of the objectives of the experiment is to study the existence of an association between one or more explanatory variables (factors) and such a discrete response variable, Agresti (2013) describes how the appropriate analysis is through the use of a regression model for discrete data, which requires the fitting of different discrete distributions such

  

as Poisson, binomial, negative binomial, and multinomial. These models belong to the class of generalized linear models and extensions, and, as in the normal case, there are assumptions that have to be met when making inference using them.

Another issue that must be considered refers to the type of study: cross-sectional or longitudinal. The planning of longitudinal studies deserves special attention because they involve collecting repeated observations on the same experimental unit, giving correlated observations. Following Diggle, Heagerty, Liang, and Zeger (2002) the analysis of longitudinal data needs to take into account this dependence structure over time within each experimental unit, typically by proposing models to describe dependence for serially correlated data. Moreover, Zeger, Liang, and Albert (1988) emphasize that the choice of model for the analysis of longitudinal data should take into account not only the nature of the response variable but also the scientific hypotheses of interest. In this context, marginal models through generalized estimating equations (GEE) and subject-specific mixed models are two classical approaches commonly used for longitudinal data analysis. In particular, marginal models are used when the interest is in the average response over time, whereas in mixed models, the natural heterogeneity of individuals is considered through the inclusion of appropriate random effects, being useful to tackle overdispersion problems. Extensions of these two classes of models to categorized data are also already used in the literature, for example, Hedeker (2003) shows the inclusion of random effects for the generalized logistic model, when there is an excess of zero counts and Touloumis, Agresti, and Kateri (2013) describe GEE for categorized responses using local odds ratios parameterizations.

Transition models are also an important framework that can be used to model longitudinal categorical data. These models are particularly useful when the interest is in prediction, although dependence can also be measured by a regression coefficient for the past state (Ware, Lipsitz, & Speizer, 1988). They should be used when the interest is in what happens to the categorical responses from one moment to another. Marginal models and mixed models are not able to capture these changes of responses over time and the effects involved (Rodrigues de Lara, Hinde, De Castro, & Da Silva, 2016). In these cases, the possible dependence within longitudinal data is commonly incorporated through a Markov type stochastic process, where an individual's current state is influenced by the state of the individual on the previous occasion, but does not depend on any more general history of the individual's responses (Stirzaker, 2005). In this case, the individual's previous response generally strongly influences his current response and is an important covariate in any prediction. This Markov type dependency is a strong assumption and has to be checked statistically and, in practice, this can be verified by fitting appropriate nested models that can be compared using likelihood-ratio tests. Since these models are based on stochastic processes, they can be divided into two classes. The first is discrete time, that is, when observations of response categories are measured at predefined occasions, usually equally spaced over time. In contrast, in the class of continuous time processes, the set of time values is not predefined, but rather experimentally measured together with the category response variable for more details see Lindsey (2004)). In this case, Kalbfleisch and Lawless (1985) discuss that the interest is also in the transition time, for which the transition intensities can be estimated, parameters by which it is possible to describe average residence time and category change. Continuous time transition models are also known as multi-state models for which semi-parametric methods are also available (Meira-Machado, de Uña-Àlvarez, Cadarso-Suarez, & Andersen, 2009).

Whether the time parameter is discrete or continuous, the available methods for these classes of models are suitable for the situation in which responses are recorded over time for each individual (Diggle et al., 2002; Molenberghs & Verbeke, 2005), with the data file presenting a "stacked structure" of states for each individual at measured times (De Rooij, 2011). However, in the discrete time framework, rather than recording data at the individual level we may simply have data at the grouped category level, that is, at each time point the number of individuals in each of the different categories; for example, this is common when conducting choice tests. The two motivational studies in this work describe experimental setups designed to observe the behavior of groups of individuals over time. In the first study, groups of 90 female insects (Asian citrus psyllid) were observed in relation to their choice of four plant species. In the second example, groups of 16 male pigs had their behavior observed when confined in pens and subjected to two treatments: with and without environmental enrichment. Similar situations are very common in experiments in entomology and zootechnics, in which, often, the observation of individual behavior is not possible, but the recording of what happens to the group from one occasion to another is perfectly possible.

Then, when the collected data in the longitudinal study are arranged in a contingency table, where the response corresponds to frequencies over the many categories at each time point, we are dealing with grouped longitudinal multinomial data. Although there are non-parametric methods and log-linear models for contingency table data analysis (see Lindsey (1995)), methodologies for grouped data using transition models are still scarce in the literature. In this context, the purpose of this paper is to extend the methodology of transition models for the analysis of multinomial grouped data over discrete time. We propose a new framework for considering the time dependence in the linear predictor of a generalized

logit transition model using a quantitative response. We present a simulation study and illustrate our methodology using the two real examples mentioned above.

## 2 | MODELING

Initially, we define a transition model for discrete individual level data, and then present the extension to the grouped case.

### 2.1 | Transition models

Let $\mathbf{y}_i = (y_{i0}, y_{i1}, \dots, y_{in_i})'$ be the $(n_i + 1)$-dimensional vector of response variables for the $i$th individual, with $i = 1, 2, \dots, N$. Also let $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})'$ be the $p$-dimensional vector of associated covariates, which may or not depend on time, and $\mathbf{h}_{it} = (y_{i(t-1)}, y_{i(t-2)}, \dots, y_{i(t-q)})$, the $q$-dimensional vector of previous responses, is the (partial) history. A transition model is specified for $Y_{it} \mid \mathbf{h}_{it}$ and the linear predictor is given by:

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

where $g(\cdot)$ is a link function, $\mu_{it} = \mathrm{E}(Y_{it} \mid \mathbf{h}_{it})$ and $f_r^*$ are functions that define the structure of the transition model in the linear predictor. The vector $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ represents the weights of the transition probabilities, in which $\boldsymbol{\beta}$, of dimension $p \times 1$, is associated with the covariates, and $\boldsymbol{\alpha}$ is associated with the previous responses and has a dimension that depends on both the order $q$ and the specific form of the functions $f_r^*$. Because these conditional transition models simply have additional (known) history dependent covariates in the linear predictor, the estimation process is the same as the one used for generalized linear models and their extensions, that is, their parameters are estimated via maximum likelihood where the joint likelihood can be formed from the conditionals $Y_{it} \mid \mathbf{h}_{it}$, see Diggle et al. (2002).

For a multinomial response, we consider a discrete time process, that is, the set $\tau = \{0, 1, \dots, T\}$ represents the times on which a process was observed. Moreover, we have a discrete state space, that is, the set $S = \{1, 2, \dots, k\}$ represents the response categories. The first-order Markov assumption establishes that the probability of the current event, $Y(t)$, given the process history depends only on the previous state of the process, $Y(t-1)$, is

$$\begin{aligned} P[Y(t) = b \mid Y(0) = y_0, Y(1) = y_1, \dots, Y(t-1) = a] &= P[Y(t) = b \mid \mathbf{h}_t)] = \\ &= P[Y(t) = b \mid Y(t-1) = a], \end{aligned} \tag{2}$$

which is called the transition probability from $a$ to $b$ at time $\tau$, with $a, b \in S$ and $t \in \tau$. These probabilities can be combined into a $k \times k$ matrix, called the transition probability matrix. They are of central importance as they describe the probabilistic process for changes in response categories over time and, moreover, these probabilities themselves may or may not be homogeneous over time. When they are not homogeneous over time, in this text, we say that the process is non-stationary with time dependent transition probabilities. To simplify the notation for these one-step transition probabilities, we write $\pi_{ab}(t-1, t) = \pi_{ab}(t)$ for the transition probability at time $t$ from state $a$ to state $b$, and $\mathbf{P}(t)$ for the respective transition probability matrix:

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

$\{\mathbf{P}(t) : t \in \tau\}$ represents the set of $T$ first-order transition matrices and when these depend upon additional covariates $\mathbf{x}$ we write $\mathbf{P}(t; \mathbf{x})$. Sometimes, the process is considered homogeneous over time, so then we have that the $T$ transition matrices are identical and so can write, $\mathbf{P}(t) = \mathbf{P}$ for all $t \in \tau$; this represents a stationary process.

To estimate these matrices, especially when there is a set of covariates $\mathbf{x}$, we use the generalized logit model (Agresti, 2013), which is given by:

$$\eta_{bt} = \log\left(\frac{\pi_{ab}(t;\mathbf{x})}{\pi_{ak}(t;\mathbf{x})}\right) = \lambda_{bt} + \boldsymbol{\delta}'_{bt}\mathbf{x}, \tag{3}$$

in which $\boldsymbol{\delta}'_{bt} = (\beta_{bt1}, \ldots, \beta_{btp}, \alpha_{bt})$ is the vector of unknown parameters, associated with category $b$ ($b = 1, 2, \ldots, k-1$), with $k$ the reference category. In general, $k$ is the first or last category, depending on the statistical software used. However, it is possible to fix it according to the researcher's interest. Finally, $\lambda_{bt}$ is an intercept.

In model (3) the vector $\boldsymbol{\delta}'_{bt}$ varies with each category of response level as well as depending on the $t$th time transition ($t = 1, 2, \ldots, T$). The parameters, $\lambda_{bt}$ and $\boldsymbol{\delta}'_{bt}$ (model 3) are estimated by maximum likelihood, using an iterative procedure as an extension of the classical method for logit models. This procedure for fitting generalized logit models is implemented in several computational packages such as `VGAM` (Yee, 2010), `drm` (Jokinen, 2013), `mlogit` (Croissant, 2013), and nnet (Ripley & Venables, 2016) available for R software (R Core Team, 2019). The estimated probabilities from the fitted model allow for the prediction of future events, given the history and the effects of covariates. The predicted transition probabilities can be written as:

$$\hat{\pi}_{ab}(t;\mathbf{x}) = \frac{\exp(\hat{\lambda}_{bt} + \hat{\boldsymbol{\delta}}'_{bt}\mathbf{x})}{1 + \sum_{b=1}^{k-1}\exp(\hat{\lambda}_{bt} + \hat{\boldsymbol{\delta}}'_{bt}\mathbf{x})}. \tag{4}$$

As is known, the maximum likelihood estimators are normally distributed asymptotically, which allows for the construction of confidence intervals (CIs) for the model parameters $\boldsymbol{\delta}'_{bt}$, as well as for the predicted transition probabilities $\pi_{ab}(t)$. As an illustration, consider the stationary model:

$$\eta_b = \text{logit}(\pi_{ab}) = \log\left(\frac{\pi_{ab}}{\pi_{ak}}\right) = \lambda_b + \boldsymbol{\beta}'_b x + \boldsymbol{\alpha}'_b y_{(t-1)}, \tag{5}$$

where $x$ represents a dummy variable (presence or absence of a treatment effect) and $y_{(t-1)}$ the previous response in the linear predictor. Also suppose that $a, b \in S = \{1, 2, 3\}$. For a fixed previous state, $a \in S$, an asymptotic 95% confidence interval for some $\beta_b$ is given by: $\hat{\beta}_b \pm 1.96\sqrt{\widehat{Var}(\hat{\beta}_b)} \quad \forall \quad b = 1, 2$ and category 3 is the reference. Analogously, an asymptotic CI can be constructed for the Markov parameter $\alpha_b$.

Now, to obtain asymptotic confidence intervals for the predicted transition probabilities, note that from Equation (5), for $a$ fixed, we have two logits:

$$\text{logit}(\pi_{a1}) = \log\left(\frac{\pi_{a1}}{\pi_{a3}}\right) = \lambda_1 + \beta_1 x + \alpha_1 y_{(t-1)}$$

and

$$\text{logit}(\pi_{a2}) = \log\left(\frac{\pi_{a2}}{\pi_{a3}}\right) = \lambda_2 + \beta_2 x + \alpha_2 y_{(t-1)}.$$

From expression (4) it follows that:

$$\hat{\pi}_{a1} = \frac{\exp(\hat{\lambda}_1 + \hat{\beta}_1 + \hat{\alpha}_1)}{\exp(\hat{\lambda}_1 + \hat{\beta}_1 + \hat{\alpha}_1) + \exp(\hat{\lambda}_2 + \hat{\beta}_2 + \hat{\alpha}_2)} \tag{6}$$

and

$$\hat{\pi}_{a2} = \frac{\exp(\hat{\lambda}_2 + \hat{\beta}_2 + \hat{\alpha}_2)}{\exp(\hat{\lambda}_1 + \hat{\beta}_1 + \hat{\alpha}_1) + \exp(\hat{\lambda}_2 + \hat{\beta}_2 + \hat{\alpha}_2)}. \tag{7}$$

The variance of logit estimator is given by:

$$\widehat{Var}[\text{logit}(\hat{\pi}_{ab})] = \widehat{Var}(\hat{\lambda}_b) + \widehat{Var}(\hat{\beta}_b) + \widehat{Var}(\hat{\alpha}_b) + 2\widehat{Cov}(\hat{\lambda}_b, \hat{\beta}_b) + 2\widehat{Cov}(\hat{\lambda}_b, \hat{\alpha}_b) + 2\widehat{Cov}(\hat{\beta}_b, \hat{\alpha}_b),$$

where the variances and covariances can be obtained from the estimated Fisher information matrix. Thus, a 95% asymptotic confidence interval for some $\text{logit}(\pi_{ab})$, $b = 1, 2$, can be written as follows:

$$\text{logit}(\hat{\pi}_{ab}) \pm 1.96\sqrt{\widehat{Var}[\text{logit}(\hat{\pi}_{ab})]}. \tag{8}$$

Finally, we apply the inverse logit transformation to the CI in expression (8) to obtain the CI for the predicted transitions probabilities $\pi_{a1}$ and $\pi_{a2}$.

## 2.2 | Tests to assess stationarity

A relevant modeling issue is the assumption of stationarity, because in that case we have a much simpler process with fewer parameters. Anderson and Goodman (1957) proposed a classical test to assess stationarity, based on the estimated transition probabilities under stationarity and non-stationarity. The null hypothesis is $\text{H}_0 : \mathbf{P}(t; \mathbf{x}) = \mathbf{P}(\mathbf{x})$ for all $t \in \tau$ and it was originally proposed for nominal data in homogeneous samples where all of the covariates are categorical. Here, we use likelihood-ratio tests to assess stationarity, as recently proposed by De Lara, Hinde, and Taconeli (2017). The null hypothesis is

$$\text{H}_0 : \boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \boldsymbol{\delta}_3 = \cdots = \boldsymbol{\delta}_T, \tag{9}$$

and the alternative hypothesis is that at least one pair of parameter vectors are different. In this version it is not necessary to estimate the transition probabilities, since the model coefficients are directly associated with them. Indeed, for the non-stationary process, the $T$ vectors $\boldsymbol{\delta}_t = (\boldsymbol{\beta}_t, \boldsymbol{\alpha}_t)$ of interest are necessary to define the $\mathbf{P}(t; \mathbf{x})$ matrices as shown by Ware et al. (1988), whereas in the stationary process only one vector $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ defines the model and associated transition matrix $\mathbf{P}(\mathbf{x})$, and estimation involves the sum of individual contributions to the likelihood function (Azzalini, 1983; Diggle et al., 2002). The test statistic is written as

$$\Lambda = -2\left[\log(\text{L}(\boldsymbol{\delta}_0, \mathbf{x})) - \sum_{t=1}^{T} \log(\text{L}_t(\hat{\boldsymbol{\delta}}_t, \mathbf{x}))\right], \tag{10}$$

where $\text{L}(\boldsymbol{\delta}_0, \mathbf{x}) = \sup[\text{L}(\boldsymbol{\delta}, \mathbf{x}) \mid \boldsymbol{\delta} \in \boldsymbol{\Theta}_0]$, with $\boldsymbol{\Theta}_0$ the parameter subspace associated with the null hypothesis $\text{H}_0$, and the contribution under the unrestricted parameter space, $\boldsymbol{\Theta}$, is given by the sum of the log-likelihoods over the $T$ individual transitions. Asymptotically, under $\text{H}_0$, $\Lambda \sim \chi^2_{\upsilon}$, with degrees of freedom $\upsilon = \dim(\boldsymbol{\Theta}) - \dim(\boldsymbol{\Theta}_0)$.

Alternatively, we can use global and local tests, as proposed by Rodrigues de Lara, Hinde, and Taconeli (2018). In this case we include an additional covariate for the transition time occasion in the linear predictor and assess the significance of its interaction with other covariates. To illustrate this, let the transition model for nominal data under stationarity be

$$\eta_{bt}^{(s)} = \lambda_b + \boldsymbol{\beta}_b'(\text{treatment}) + \boldsymbol{\alpha}_b' y_{(t-1)}, \tag{11}$$

which includes the treatment effect and the previous response in the linear predictor. Now, we add the time transition factor in the linear predictor and all interactions with this extra factor giving

$$\begin{aligned} \eta_{bt}^{(ns)} &= \lambda_b + \boldsymbol{\beta}_b'(\text{treatment}) + \boldsymbol{\alpha}_b' y_{(t-1)} + \\ &+ \boldsymbol{\vartheta}_b'(\text{time}) + \boldsymbol{\alpha}_b^{*\prime}(\text{time} \times y_{(t-1)}) + \boldsymbol{\beta}_b^{*\prime}(\text{time} \times \text{treatment}). \end{aligned} \tag{12}$$

Model (12) includes additional parameters for the time transition dependence, namely $(\boldsymbol{\vartheta}_b', \boldsymbol{\alpha}_b^{*\prime}, \boldsymbol{\beta}_b^{*\prime})$. Now the global null hypothesis can be written as

$$\text{H}_0 : (\boldsymbol{\delta}_b, \boldsymbol{\vartheta}_b', \boldsymbol{\alpha}_b^{*\prime}, \boldsymbol{\beta}_b^{*\prime}) = (\boldsymbol{\delta}_{b_0}, \mathbf{0}, \mathbf{0}, \mathbf{0}), \tag{13}$$

where $\mathbf{0}$ is a null vector and $\boldsymbol{\delta}_{b_0}$ denotes the vector of coefficients associated with covariates and previous responses under the null hypothesis. The hypotheses (13) and (9) are equivalent and the likelihood-ratio test for this hypothesis (13) is a global test, whose statistic value is the same as for Equation (10).

**TABLE 1** Grouped data: Structure of a $N \times T \times k$ contingency table that can be analyzed using transition models for grouped data

| Units | Time | Covariates | Categories ($S$) | | | | Unit |
|---|---|---|---|---|---|---|---|
| $i$ | $t$ | $\mathbf{x}_{it}$ | 1 | 2 | ... | k | totals |
| 1 | 0 | $\mathbf{x}_{10}$ | $w_{101}$ | $w_{102}$ | ... | $w_{10k}$ | $m_{10}$ |
| 1 | 1 | $\mathbf{x}_{11}$ | $w_{111}$ | $w_{112}$ | ... | $w_{11k}$ | $m_{11}$ |
| 1 | 2 | $\mathbf{x}_{12}$ | $w_{121}$ | $w_{122}$ | ... | $w_{12k}$ | $m_{12}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋱ | ⋮ | ⋮ |
| 1 | $T$ | $\mathbf{x}_{1T}$ | $w_{1T1}$ | $w_{1T2}$ | ... | $w_{1Tk}$ | $m_{1T}$ |
| 2 | 0 | $\mathbf{x}_{20}$ | $w_{201}$ | $w_{202}$ | ... | $w_{20k}$ | $m_{20}$ |
| 2 | 1 | $\mathbf{x}_{21}$ | $w_{211}$ | $w_{212}$ | ... | $w_{21k}$ | $m_{21}$ |
| 2 | 2 | $\mathbf{x}_{22}$ | $w_{221}$ | $w_{222}$ | ... | $w_{22k}$ | $m_{22}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋱ | ⋮ | ⋮ |
| 2 | $T$ | $\mathbf{x}_{2T}$ | $w_{2T1}$ | $w_{2T2}$ | ... | $w_{2Tk}$ | $m_{2T}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋱ | ⋮ | ⋮ |
| $N$ | 0 | $\mathbf{x}_{N0}$ | $w_{N01}$ | $w_{N02}$ | ... | $w_{N0k}$ | $m_{N0}$ |
| $N$ | 1 | $\mathbf{x}_{N1}$ | $w_{N11}$ | $w_{N12}$ | ... | $w_{N1k}$ | $m_{N1}$ |
| $N$ | 2 | $\mathbf{x}_{N2}$ | $w_{N21}$ | $w_{N22}$ | ... | $w_{N2k}$ | $m_{N2}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋱ | ⋮ | |
| $N$ | $T$ | $\mathbf{x}_{NT}$ | $w_{NT1}$ | $w_{NT2}$ | ... | $w_{NTk}$ | $m_{NT}$ |

If the test is significant, several structures can be formulated, corresponding to submodels of Equation (12), and using likelihood-ratio tests for nested models (local tests) we can select the best or most parsimonious model. If the selected model includes the time factor, then there is evidence that the process is non-stationary, which means that the transition probabilities are not homogeneous, due to the linear predictor changing over time. We give more details of these tests in our applications. Previous simulation studies developed by Rodrigues de Lara et al. (2018) have shown that the likelihood-ratio test using the transition model coefficients maintains the level of significance and has power compatible with the classical test. In addition, the general framework is flexible and can be applied with quantitative and categorical variables in the linear predictor.

## 2.3 | Extension for grouped data

Consider now the situation where at time point $t$ each observational or experimental unit $i$ corresponds to a number of individuals, $m_{it}$. For a closed system with complete observation of individuals at each time point $m_{it}$ is constant for all $t \in \tau$, however, in many applications this is not the case and so we do not make this restrictive assumption. Writing $w_{itj}$ to represent the associated count for the $i$th unit at the $t$th time for the $j$th response category associated with a $p$-dimensional vector of covariates (see Table 1), the total individual count at each time point is

$$m_{it} = \sum_{j=1}^{k} w_{itj}.$$

The $w_{itj}$ individuals of each category at time $t \in \tau$ can freely move to other categories at future times. Since we do not have individual information, we do not know which individuals changed from one state to another, only the numbers that were observed in each category at each time point. Now, to define the transition model, we consider the vector $\mathbf{y}_{it} = (1, 2, \dots, k)'$ as a multivariate response for each unit $i$ at time $t$, with an associated weights vector $\mathbf{w}_{it} = (w_{it1}, w_{it2}, \dots, w_{itk})'$ of counts. To incorporate first-order time dependence, we define $\mathbf{h}_{it} = (w_{i(t-1)1}, w_{i(t-1)2}, \dots, w_{i(t-1)k})$ as a vector of explanatory variables represented by the observed frequencies of each of the categories at the previous time. Then, the generalized transition model for grouped longitudinal categorical data is given by

$$\eta_{bt} = \log\left(\frac{\pi_{\mathbf{h}b}(t;\mathbf{x})}{\pi_{\mathbf{h}k}(t;\mathbf{x})}\right) = \lambda_{bt} + \boldsymbol{\beta}'_{bt}\mathbf{x} + \boldsymbol{\alpha}'_{bt}\mathbf{h}. \tag{14}$$

Note that (14) is a little different from the usual generalized logit transition model (3), because previous information here is not the category of response at the previous time but rather the observed frequencies for each of them. Therefore, these counts are incorporated as additional quantitative explanatory variables in the model. Consequently, the transition probability to category $b$ ($b = 1, 2, \ldots, k-1$) at time $t$ is given by

$$\pi_{\mathbf{h}b}(t;\mathbf{x}) = \frac{\exp(\lambda_{bt} + \boldsymbol{\beta}'_{bt}\mathbf{x} + \boldsymbol{\alpha}'_{bt}\mathbf{h})}{1 + \sum_{b=1}^{k-1}\exp(\lambda_{bt} + \boldsymbol{\beta}'_{bt}\mathbf{x} + \boldsymbol{\alpha}'_{bt}\mathbf{h})}, \tag{15}$$

that is, $\pi_{\mathbf{h}b}(t,\mathbf{x})$ does not describe the probability of transition from a particular state to $b \in S$, as in the usual transition model (3) for individual data. In contrast, we now have the total history of all $k$ categories at the previous time. We can interpret (15) in the following way: $\pi_{\mathbf{h}b}(t;\mathbf{x})$ is the probability of transition to category $b$ at time $t$ given that at time $t-1$ unit $i$ had $m_{i(t-1)}$ individuals arranged as frequencies $(w_{i(t-1)1}, w_{i(t-1)2}, \ldots, w_{i(t-1)k})$ across the $k$ categories. Therefore, as in the traditional case, this probability expresses the propensity for a category change given the history of the experimental or observational unit $i$. We can write the likelihood function at the $t$th transition as

$$\begin{aligned} L_t(\boldsymbol{\lambda},\boldsymbol{\beta},\boldsymbol{\alpha}) &= \prod_{i=1}^{N}\left[\prod_{b=1}^{k}[\pi_{\mathbf{h}b}(t;\mathbf{x})]^{w_{itb}}\right] \\ &= \prod_{i=1}^{N}\left[\prod_{b=1}^{k}\left[\frac{\exp(\lambda_{bt} + \boldsymbol{\beta}'_{bt}\mathbf{x} + \boldsymbol{\alpha}'_{bt}\mathbf{h})}{1 + \sum_{b=1}^{k-1}\exp(\lambda_{bt} + \boldsymbol{\beta}'_{bt}\mathbf{x} + \boldsymbol{\alpha}'_{bt}\mathbf{h})}\right]^{w_{itb}}\right]. \end{aligned} \tag{16}$$

Note that for the situation where the units $i$ are single individuals, the weights comprise a vector of zeros with a 1 occurring only for the category to which $y_i$ belongs to at time $t$, and hence model (16) provides a generalization of this framework for grouped data, whose log-likelihood is given by:

$$\begin{aligned} l_t &= \log\left\{\prod_{i=1}^{N}\left[\prod_{b=1}^{k}[\pi_{\mathbf{h}b}(t;\mathbf{x})]^{w_{itb}}\right]\right\} \\ &= \sum_{i=1}^{N}\left[\sum_{b=1}^{k-1} w_{itb}(\lambda_{bt} + \boldsymbol{\beta}'_{bt}\mathbf{x} + \boldsymbol{\alpha}'_{bt}\mathbf{h}) - \log\left(1 + \sum_{b=1}^{k-1}\exp(\lambda_{bt} + \boldsymbol{\beta}'_{bt}\mathbf{x} + \boldsymbol{\alpha}'_{bt}\mathbf{h})\right)\right]. \end{aligned} \tag{17}$$

As is common with discrete models, the maximum likelihood estimators of Equation (16) or (17) do not have a closed analytical form, that is, they are not directly maximized. Therefore, iterative numerical methods are needed. Here, to maximize the function (16 or 17) we have used an extension of the standard procedure of generalized logit models, that is, iteratively weighted least squares (Newton–Raphson). As already explained in Section 2, it is also possible to adapt standard packages, available for R software (R Core Team, 2019), to fit transition models with individual or grouped data.

Here one must work with a stacked structure for the data, creating a vector for the response variable and their respective weights, as well as incorporating the category frequencies at the previous time as additional covariates. The computational implementation was made using the `nnet` package (Ripley & Venables, 2016) available for the `R` system (R Core Team, 2019). An issue with this is the fact that the design matrix is ill-conditioned when the total number of individuals in all groups remain unchanged for all observation units. In fact, according to Mandel (1982), the ill-conditioning is not a statistical problem of the adopted model but is induced by linear combinations of the regression matrix columns. The author shows that in this case, when changing the regressor matrix for its singular value decomposition, the ill-conditioning is reflected in the singular values. This suggested examining the condition number of the model matrix, which is the ratio between the largest and smallest singular values. As alternatives to this problem, Johnson and Wichern (2007) suggests using the technique of principal components. Another possible solution is to adopt the compositional data transformation technique proposed by Egozcue, Pawlowsky-Glahn, Mateu-Figueras, and Barceló-Vidal (2003). Although they are distinct procedures, both aim to obtain orthogonal regressive variables. This problem however does

not affect predictions or statistical tests related to the transition model for grouped data. We explain this in more detail in Section 5.

## 3 | SIMULATION STUDIES

A simulation study was carried out to assess the performance of the proposed model structure for grouped longitudinal categorical data. We considered a response variable with four nominal response categories, called "A," "B," "C," and "D," a treatment factor with two levels and 5, 10, and 20 replicates (depending on the simulation scenario), and three time occasions. The data were simulated under two first-order Markov chain processes, using the following two functional structures:

1. Stationary process:

$$\eta_{bt} = \log\left(\frac{\pi_{\mathbf{h}b}}{\pi_{\mathbf{h}k}}\right) = \lambda_b + \boldsymbol{\beta}_b'(\text{treatment}) + \boldsymbol{\alpha}_b'\mathbf{w}, \quad b = 1, \ldots, k-1, \tag{18}$$

   which corresponds to an additive structure in the linear predictor for the treatment effect and the Markov covariate, $\mathbf{w} = (w_{(t-1)1}, w_{(t-1)2}, w_{(t-1)3}, w_{(t-1)4})$, of the frequencies of categories "A," "B," "C," and "D," respectively, at the previous time point;
2. Non-stationary process:

$$\eta_{bt} = \log\left(\frac{\pi_{\mathbf{h}b}(t)}{\pi_{\mathbf{h}k}(t)}\right) = \lambda_b + \boldsymbol{\beta}_b'(\text{treatment}) + \boldsymbol{\alpha}_b'\mathbf{w} + \\ + \boldsymbol{\vartheta}_b'(\text{time}) + \boldsymbol{\beta}_b^{*\prime}(\text{time} \times \text{treatment}) + \boldsymbol{\alpha}_b^{*\prime}(\text{time} \times \mathbf{w}), \tag{19}$$

   which corresponds to the inclusion of the interaction between the additive structure of the stationary model and the transition time, a factor included to account for time dependence.

The true parameter values used are presented in the Appendix. For each scenario we performed 10,000 simulations taking the group sizes $m_{it}$ to be fixed over time and equal for each experimental unit, that is, $m_{it} = m$ with three different group size settings ($m = 15, 30$, and 60). The discrepancy between the parameter values and the estimates in each scenario was evaluated using the bias and mean squared error (MSE). The computational implementation was made in the `R` software, using the `markovchain` package (Spedicato, 2017).

We carried out an additional simulation study to compare the efficiency of the estimates associated with the treatments ($\beta_b$ and $\beta_b^*$), when data are available at the individual as compared to the grouped level. The set up was the same as the one described above, but we conducted it for a situation where the individual data is available following the methodology in Rodrigues de Lara et al. (2016). In addition to bias and MSE, we also computed the coverage rate of the 95% asymptotic confidence intervals and their average width. All codes used to produce the simulation studies are made available as Supporting Information.

The results obtained for the stationary and non-stationary scenarios are presented in the Appendix. For the stationary scenario, all parameters have low MSE values that decrease when the groups size $m$ and the number of replicates increase, which is what is to be expected as we have more data, and indicates a good overall performance. When looking at the non-stationary scenario, all parameters have low bias and MSEs, except $\beta_b$ and $\beta_b^*$, both related to the treatment effect. This may be due to the fact that this model has many more parameters than the stationary one. However, note that again the bias and MSE decreases with group size $m$ and number of replicates.

When comparing the coverage rate of the CIs between individual and grouped level data, all coverage rates are very close to the nominal value of 95%. However the CI width for the individual level data is consistently smaller than the CI width for the grouped level data, indicating the loss of efficiency when we no longer have the detailed individual information. This is typically an intrinsic aspect of the design where, for a variety of reasons, it may be impossible to collect data at the individual level. However, the simulation studies show that the modeling framework proposed in this paper is capable of recovering well the true parameter values, and the CIs behave as expected.

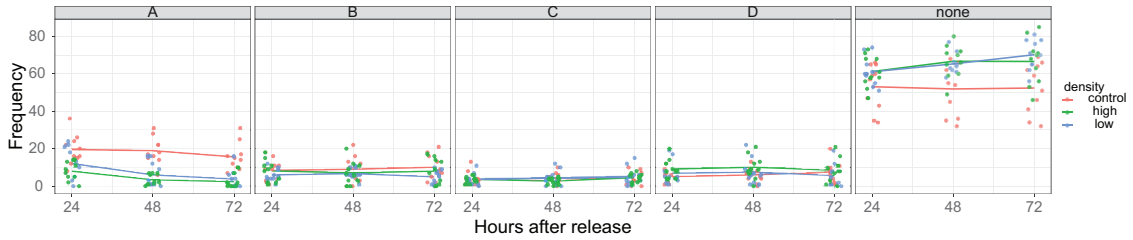


**FIGURE 1** Psyllid movement data: observed frequency of psyllids that are observed in each of the four plants (or none) over time. The lines connect the means over time

## 4 | APPLICATIONS

### 4.1 | Psyllid movement data

The Asian citrus psyllid, *Diaphorina citri*, is a pest of citrus plantations and a vector of bacteria that cause Huanglongbing, or the citrus greening disease. To understand the movement patterns of female adults of *D. citri* for feeding and egg oviposition in Orange Jessamine plants (*Murraya paniculata*) (considered an alternative host plant for the pest), an experiment was set up in the Department of Entomology and Acarology of the University of São Paulo (Piracicaba, Brazil), using a completely randomized design with three treatments and ten replicates. Each experimental unit consisted of a nylon cage (200×60×40 cm) with a row of four Orange Jessamine plants placed at equal intervals of 50 cm, namely "A," "B," "C," and "D," with "A" the closest plant to the entrance and "D" the furthest away. In each experimental unit, 90 female adults were released on plant "A," and after 24, 48, and 72 hr the numbers of insects on each plant and flying were observed. To investigate whether the presence of *D. citri* nymphs (previous infestation) affected adult female movement, three treatments were used: "high density," in which plant "A" received 35 *D. citri* nymphs prior to the release of the adults, "low density," in which plant "A" had 15 nymphs, and control, in which there were no nymphs on any of the plants. The observed frequencies by category and times are shown in Figure 1. The frequencies observed of the category "None" are much higher than the others, and seem to increase over time for the "high density" and "low density" treatments, while remaining constant for the control treatment.

We started by fitting the transition model under stationarity,

$$\eta_{bt} = \lambda_b + \boldsymbol{\beta}_b'(\text{density}) + \boldsymbol{\alpha}_b'\mathbf{w}$$

and non-stationarity,

$$\eta_{bt} = \lambda_b + \boldsymbol{\beta}_b'(\text{density}) + \boldsymbol{\alpha}_b'\mathbf{w} + \boldsymbol{\vartheta}_b'(\text{time}) + \boldsymbol{\alpha}_b^{*\prime}(\text{time} \times \mathbf{w}) + \boldsymbol{\beta}_b^{*\prime}(\text{time} \times \text{density}),$$

and used the likelihood-ratio tests to assess stationarity. The maximized log-likelihoods for the fitted models were −16440.72 (under stationarity) and −16413.05 (under non-stationarity), indicating that the process is non-stationary ($\Lambda = 55.34$ on 32 df, $p = 0.0064$), that is, the transition probabilities are not homogeneous over time. We then fitted models reflecting different non-stationary structures, following the work of Rodrigues de Lara et al. (2018), namely

$$\text{Model P1: } \eta_{bt} = \lambda_b + \boldsymbol{\beta}_b'(\text{density}) + \boldsymbol{\alpha}_b'\mathbf{w} + \boldsymbol{\vartheta}_b'(\text{time})$$
$$\text{Model P2: } \eta_{bt} = \lambda_b + \boldsymbol{\beta}_b'(\text{density}) + \boldsymbol{\alpha}_b'\mathbf{w} + \boldsymbol{\vartheta}_b'(\text{time}) + \boldsymbol{\alpha}_b^{*\prime}(\text{time} \times \mathbf{w})$$
$$\text{Model P3: } \eta_{bt} = \lambda_b + \boldsymbol{\beta}_b'(\text{density}) + \boldsymbol{\alpha}_b'\mathbf{w} + \boldsymbol{\vartheta}_b'(\text{time}) + \boldsymbol{\beta}_b^{*\prime}(\text{time} \times \text{density}).$$

Likelihood-ratio tests indicated that model P3 represented an improvement in the fit over model P1 ($\Lambda = 33.76$ on 8 df, $p < 0.0001$), but model P2 did not ($\Lambda = 18.20$ on 16 df, $p = 0.3125$). When testing model P3 versus the full non-stationary model, the test indicated no improvement ($\Lambda = 20.25$ on 16 df, $p = 0.2091$), and hence model P3 was selected as the final model. The parameter estimates and respective standard errors for model P3 are shown in Table 2.

Although the number of insects observed in each category at the previous time (Markov covariate) must be considered to study the transition behavior, there is an effect of the interaction between time and nymph density (treatment). This

**TABLE 2** Psyllid movement data: parameter estimates (standard errors) for the fitted non-stationary model P3

| | Category | | | |
|---|---|---|---|---|
| **Parameter** | **B** | **C** | **D** | **None** |
| $\lambda_b$ (Intercept) | −0.0011 (0.0001) | 0.0003 (0.0001) | 0.0003 (0.0001) | −0.0015 (0.0001) |
| $\alpha_b$ (A) | −0.0256 (0.0036) | −0.0402 (0.0047) | −0.0012 (0.0035) | −0.0531 (0.0040) |
| $\alpha_b$ (B) | −0.0487 (0.0062) | −0.0027 (0.0074) | −0.0061 (0.0060) | −0.0640 (0.0066) |
| $\alpha_b$ (C) | −0.0628 (0.0107) | 0.0146 (0.0126) | −0.0347 (0.0106) | −0.0482 (0.0114) |
| $\alpha_b$ (D) | 0.0272 (0.0072) | 0.0578 (0.0083) | 0.0330 (0.0071) | 0.0258 (0.0075) |
| $\beta_b$ (High) | 1.2022 (0.1432) | 1.0425 (0.1459) | −1.0941 (0.1323) | 3.8417 (0.1494) |
| $\beta_b$ (Low) | 2.3252 (0.1143) | −0.4440 (0.1461) | −1.6145 (0.1105) | 1.0109 (0.1465) |
| $\vartheta_b$ (Time) | 0.0719 (0.1322) | 0.1090 (0.1317) | 0.2049 (0.0978) | 0.2732 (0.1571) |
| $\beta_b^*$ (Time×High) | −0.6214 (0.2041) | −0.5511 (0.2049) | −0.4637 (0.1851) | −0.6222 (0.2049) |
| $\beta_b^*$ (Time×Low) | −0.1527 (0.1656) | −0.0664 (0.2045) | −0.5545 (0.1627) | −0.5508 (0.2061) |

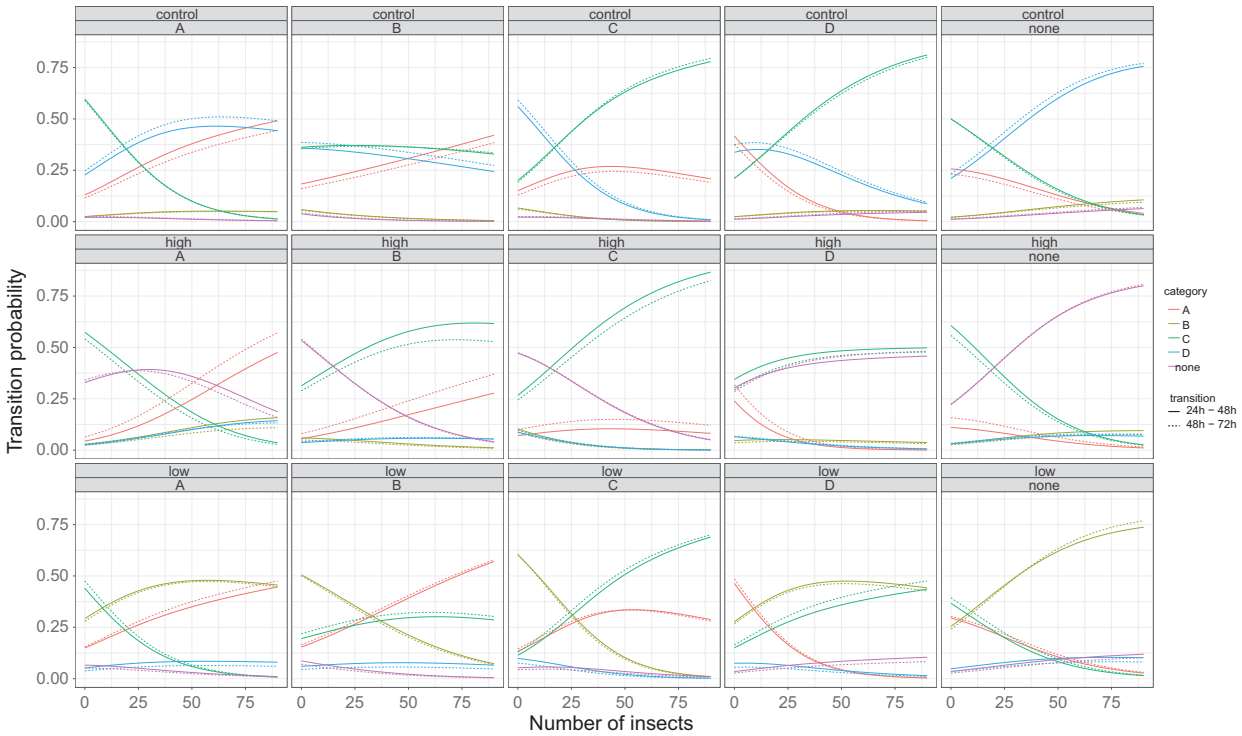


**FIGURE 2** Psyllid movement data: predicted transition probabilities for each state and treatment level, obtained from the fitted non-stationary generalized transition model P3. The *x*-axis represents the number of insects observed at the previous time on the plant (or none) in the facet label, with the remaining insects being evenly distributed across the other categories, totalling 90

influences the transition probabilities, making them non-homogeneous over time. We obtained the predicted transition probabilities based on the non-stationary model P3. We varied the number of insects that belong to each of the five categories to create a grid, while equally dividing the remaining insects between the other four categories (e.g., for 70 insects in category "A," the remaining categories will have 5 insects each, totalling 90). We then created a plot of the predictions to aid in the understanding of how the different treatments are influencing the transition probabilities, see Figure 2. We observe that when there is a high density of nymphs on the first plant ("A"), the transition probability to the "none" category is larger when compared to the low density and control. When looking at the low density, the transition probability to plant "B" is larger compared to the control. Also, for the control it appears that the more the insect advances inside the cage, it is more likely to remain on the same plant, that is, "C" or "D."

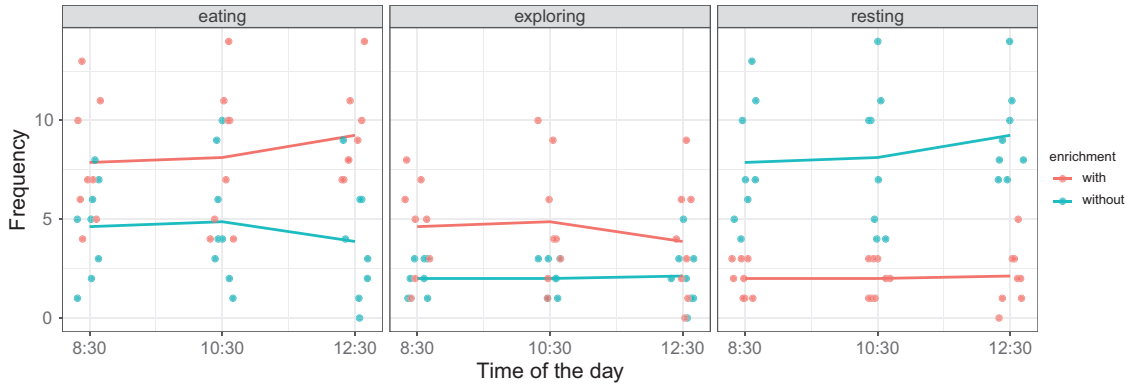


**FIGURE 3** Pig behavior data: Observed frequencies of hogs displaying one of three behaviors: "resting," "eating," or "exploring," when submitted to two rearing conditions: with and without environmental enrichment. The lines connect the means over time

**TABLE 3** Pig behavior data: Parameter estimates (standard errors) for the fitted stationary model

| | Category | |
|---|---|---|
| **Parameter** | **Exploring** | **Resting** |
| $\lambda_b$ (Intercept) | 1.3742 (1.7205) | −2.0394 (1.6589) |
| $\alpha_b$ (Eating) | −0.0880 (0.2268) | −0.0086 (0.2139) |
| $\alpha_b$ (Resting) | −0.1847 (0.1128) | 0.0637 (0.1062) |
| $\alpha_b$ (Exploring) | −0.0919 (0.1065) | 0.0167 (0.1038) |
| $\beta_b$ (Without enrichment) | −0.1068 (0.2595) | 2.1468 (0.2445) |

## 4.2 | Pig behavior data

Castro (2016) studied the effects of environmental enrichment on the behavioral activity in hogs (male pigs used for breeding), at a commercial farm in Brazil in 2014. It involved 256 animals that were exposed to one of the two rearing conditions: with and without environmental enrichment. This enrichment consisted of the inclusion of simple objects that encouraged the animals to play and were intended to reduce stress. The experiment was set up in a completely randomized design with eight replicates. Each experimental unit initially consisted of a housing pen with 16 pigs, and observations were taken three times a day for 1 month, namely at 8.30 a.m., 10.30 a.m., and 12.30 p.m. On each occasion, the researchers registered how many animals were displaying one of three behaviors: "resting," "eating," or "exploring." Here we work with data obtained on the 21st day since the start of the experiment, that is, after 3 weeks. Because there was mortality throughout the experiment, some of the experimental units had fewer than 16 pigs. An exploratory plot representing the observed proportions of behavioral activity by category over time per each treatment is shown in Figure 3. It appears that the pigs reared with environmental enrichment are more active than their peers who were reared without it (they were found resting with a lower frequency).

We followed the same procedure as in the previous application: we fitted the transition models under stationarity and non-stationarity, using the likelihood-ratio test to assess stationarity. The log-likelihoods associated with the stationary and non-stationary models were −452.13 and −447.16, respectively. The likelihood-ratio test statistic was 9.93 on 10 degrees of freedom, and hence we failed to reject the null hypothesis that the process was stationary ($p = 0.4463$). Under the stationary structure, the Markov covariates of previous state frequencies were significant ($\Lambda = 14.35$ on 6 df, $p = 0.0260$), as also was the environmental enrichment effect ($\Lambda = 108.60$ on 2 df, $p < 0.0001$). The parameter estimates and standard errors of the fitted transition model are presented in Table 3.

In the estimation process, the "eating" and "with enrichment" categories were taken as reference levels. The transition probabilities in this example are homogeneous over time and vary only according to the levels of treatment and the numbers of animals in each behavior state at the previous time. Looking at the predicted transition probabilities for each state and enrichment level, it is clear that in the absence of environmental enrichment pig behavior is affected with hogs tending to rest more, which translates into less activity (see Figure 4).

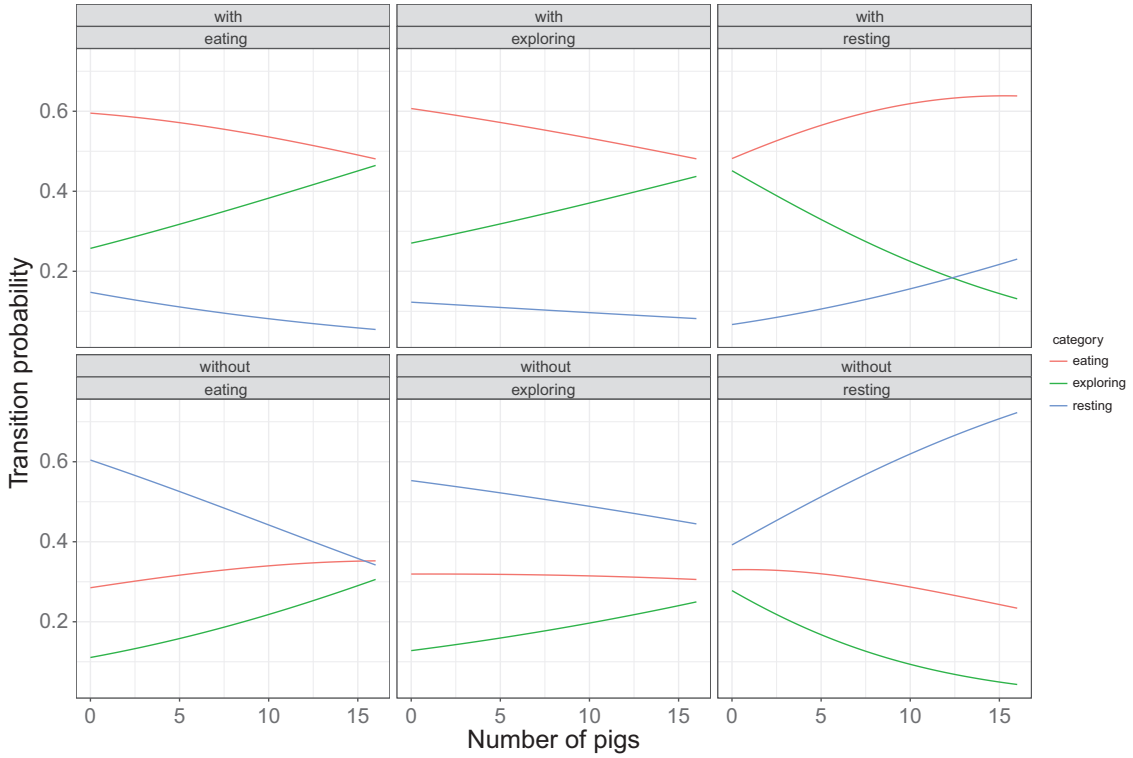


**FIGURE 4** Pig behavior data: Predicted transition probabilities for each state and environmental enrichment level, obtained from the fitted stationary generalized transition model. The $x$-axis represents the number of animals at the previous time presenting the behavior for the category given in the facet label, with the remaining animals evenly distributed across the other categories, totalling 16

## 5 | DISCUSSION

Our modeling framework presents an approach to analyze grouped longitudinal categorical data using transition models. In a transition model, it is important to incorporate dependence into the stochastic process and, when we have grouped data, this can be achieved by including the vector of previous frequencies as predictors. As in the individual case, in general, the frequencies of categories at the previous time may be interpreted as weights in the transition probabilities, but now they depend on a set of quantitative variables. Consequently, it becomes difficult to show the transition probability matrices, because they now vary with different compositions of the weights vector. However, we can represent them elegantly using graphics, in which it is possible to describe the transition probabilities for each category and interpret their effect for different treatments, etc.

Although the assumption of stationarity (or homogeneity of transition probabilities) may facilitate estimation by reducing the number of parameters, the predictions provided by this model can be misleading if this assumption is false. Therefore, it is very important to verify whether the process is stationary or not, as well to identify the sources of non-stationarity. Local tests proposed by Rodrigues de Lara et al. (2018) can be helpful at this stage, since a full non-stationary model may be overparameterized, and hence reduced alternatives may provide better solutions and simpler interpretations.

A potential issue in the specification of this model is the fact that the design matrix is ill-conditioned when the sum of the weights across the rows are $m_{it}$, which poses a multicollinearity problem. This can affect the standard errors of the regression coefficients. However, predictions and the likelihood-ratio test described above are not affected, and these are in fact often the two important objectives when using transition models. If there is an interest in computing confidence intervals for the regression coefficients, then a simple yet effective solution to this problem is to obtain the principal components of the partition of the design matrix referring to the $w_{itb}$ variables, and use them as the Markov covariates instead of the raw counts. If the $m_{it}$ are constant, then the $k$th principal component will be a vector of zeros and therefore irrelevant. This transformation does not change the log-likelihood or the model predictions, but solves the ill-conditioning problem for parameter estimates by inducing orthogonality.

Another alternative is to employ the isometric log-ratio transform (ILR) for compositional data (Egozcue et al., 2003), which reduces the weights matrix to an orthogonal space with $k-1$ columns, and again solves the ill-conditioning problem. The resulting log-likelihoods are not identical to the original model, but are approximately so. One issue with this method, however, is that whenever there is a zero present for any of the categories, the ILR transform will be equivalent to the centroid of the simplex space, and this influences prediction.

A general problem that can affect estimation is the sample and group sizes. As is well known, the generalized logit model has one parameter for each response category and, therefore, with small samples, or excess zeros, may result in estimation problems. Despite this, the estimation is possible and for our motivational studies and simulation processes we did not have any problems of this sort. Future studies will consider this type of transition model extension for ordinal categorical responses, and the impact of limiting factors such as the presence of many non-structural zeros which leads to sparse tables.

## ACKNOWLEDGMENT

I.A.R.L. would like to thank FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil) for financial support. (Proc. no. 18/06288-0).

## CONFLICT OF INTEREST

The authors have declared no conflict of interest.

## OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

## REFERENCES

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). New York: Wiley.

Anderson, T. W. & Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, *28*, 89–110.

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

Castro, A. C. (2016) *Comportamento e desempenho sexual de suínos reprodutores em ambientes enriquecidos* [unpublished doctoral dissertation]. University of São Paulo, Piracicaba, Brazil.

Croissant, Y. (2013). *Multinomial logit model*. Available from http://www.r-project.org

De Lara, I. A. R., Hinde, J., & Taconeli, C. A. (2017). An alternative method for evaluating stationarity in transition models. *Journal Of Statistical Computation and Simulation*, *87*, 1–19, https://doi.org/10.1080/00949655.2017.1351562

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.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, M., & Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, *35*, 279–300.

Hedeker, D. (2003). A mixed-effcts multinomial logistic regression model. *Statistics in Medicine*, *22*, 1433–1446.

Johnson, W. & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). New Jersey: Prentice Hall.

Jokinen, J. (2013). *Regression and association model for repeated categorical data*. Available from http://www.helsinki.fi/⟨jtjokine/drm

Kalbfleisch, J. D. & Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, *80*(392), 863–871.

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.

Mandel, J. (1982). Use of the singular value decomposition in regression analysis. *The American Statistcian*, *36*(1), 15–24.

Meira-Machado, L., de Uña-Àlvarez, J., Cadarso-Suarez, C., & Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical methods in Medical Research*, *18*(2), 195–222.

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

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

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

Rodrigues de Lara, I. A., Hinde, J. P., De Castro, A. C., & Da Silva, I. J. O. (2016). A proportional odds transition model for ordinal responses with an application to pig behaviour. *Journal of Applied Statistics*, *44*(6), 1031–1046. https://doi.org/10.1080/02664763.2016.1191623

Rodrigues de Lara, I. A., Hinde, J., & Taconeli, C. A. (2018). Global and local tests to assess stationarity of Markov transition models. *Communications in Statistics-Simulation and Computation*, *48*, 1–21. https://doi.org/10.1080/03610918.2017.1406504

Spedicato, G. A. (2017). Discrete time Markov chains with R. *The R Journal*, *9*, 84–104.

Stirzaker, D. (2005). *Stochastic processes and models*. Oxford: Oxford University Press.

Touloumis, A., Agresti, A., & Kateri, M. (2013). GEE for multinomial responses using a local odds ratios parameterization. *Biometrics*, *69*, 633–640.

Ware, J. H., Lipsitz, S., & Speizer, F. E. (1988). Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine*, *7*, 95–107.

Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, *32*, 1–34.

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, *44*, 1049–1060.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Lara IAR, Moral RA, Taconeli CA, Reigada C, Hinde J. A generalized transition model for grouped longitudinal categorical data. *Biometrical Journal*. 2020;1–21. https://doi.org/10.1002/bimj.201900394

## APPENDIX

### A.1 | Results from the simulation studies

Here we present all results from the simulation studies. The true values used in the simulations are presented in Table A.1, and Tables A.2–A.7 display the bias and mean squared errors (MSE) for all parameters resulting from the simulation

**TABLE A.1** True parameter values used for the simulation study for the stationary and non-stationary models, in which four categories of response were considered (A, B, C, D with level A taken as the reference category) and a treatment effect with two levels (1 and 2)

| | **Stationary model** | | |
|---|---|---|---|
| **Parameter** | **B** | **C** | **D** |
| $\lambda_b$ (Intercept) | 0.0003 | 0.0005 | 0.001 |
| $\alpha_b$ (A) | 0.05 | 0.08 | 0.09 |
| $\alpha_b$ (B) | −0.01 | −0.01 | −0.02 |
| $\alpha_b$ (C) | −0.02 | −0.05 | 0.01 |
| $\alpha_b$ (D) | −0.006 | 0.004 | −0.03 |
| $\beta_b$ (Treatment level 2) | −0.8 | −0.7 | −1.4 |
| | **Non-stationary model** | | |
| **Parameter** | **B** | **C** | **D** |
| $\lambda_b$ (Intercept) | −0.0002 | 0.0002 | 0.0008 |
| $\vartheta_b$ (time) | −0.0002 | −0.0001 | −0.0002 |
| $\alpha_b$ (A) | 0.05 | 0.14 | 0.02 |
| $\alpha_b$ (B) | 0.04 | −0.04 | 0.02 |
| $\alpha_b$ (C) | −0.07 | −0.1 | −0.01 |
| $\alpha_b$ (D) | −0.02 | 0.02 | 0.003 |
| $\beta_b$ (Treatment level 2) | −0.8 | −0.8 | −1.5 |
| $\alpha^*_b$ (Time×A) | −0.17 | −0.134 | −0.05 |
| $\alpha^*_b$ (Time×B) | 0.08 | 0.06 | 0.06 |
| $\alpha^*_b$ (Time×C) | −0.01 | 0.06 | −0.02 |
| $\alpha^*_b$ (Time×D) | 0.09 | 0.05 | −0.002 |
| $\beta^*_b$ (Time×Treatment level 2) | 2.8 | 1.3 | 2.1 |

**TABLE A.2** Stationary scenario with five replicates per treatment

| | $m = 15$ | | | | | |
|---|---|---|---|---|---|---|
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000512 | −0.000977 | −0.002284 | 0.000023 | 0.000023 | 0.000027 |
| $\alpha_b$ (A) | −0.005730 | −0.006635 | −0.006140 | 0.005540 | 0.005161 | 0.006349 |
| $\alpha_b$ (B) | 0.011496 | 0.000893 | 0.000747 | 0.008209 | 0.008061 | 0.009455 |
| $\alpha_b$ (C) | −0.000555 | 0.010247 | −0.001139 | 0.007976 | 0.007484 | 0.008893 |
| $\alpha_b$ (D) | −0.003391 | −0.002653 | 0.007268 | 0.007332 | 0.007085 | 0.008130 |
| $\beta_b$ (Treatment level 2) | −0.000555 | 0.010247 | −0.001139 | 0.007976 | 0.007484 | 0.008893 |
| | $m = 30$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000159 | −0.000308 | −0.000706 | 0.000001 | 0.000001 | 0.000001 |
| $\alpha_b$ (A) | −0.001805 | −0.002161 | −0.002496 | 0.001207 | 0.001139 | 0.001277 |
| $\alpha_b$ (B) | 0.004263 | −0.000217 | −0.000447 | 0.002049 | 0.001882 | 0.002099 |
| $\alpha_b$ (C) | −0.000127 | 0.003286 | 0.000055 | 0.001894 | 0.001717 | 0.001902 |
| $\alpha_b$ (D) | −0.002109 | −0.001152 | 0.001718 | 0.001559 | 0.001388 | 0.001575 |
| $\beta_b$ (Treatment level 2) | −0.000127 | 0.003286 | 0.000055 | 0.001894 | 0.001717 | 0.001902 |
| | $m = 60$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | 0.000072 | 0.000099 | 0.000159 | 0.000000 | 0.000000 | 0.000000 |
| $\alpha_b$ (A) | 0.000051 | −0.000323 | −0.000478 | 0.000310 | 0.000274 | 0.000269 |
| $\alpha_b$ (B) | 0.001008 | −0.000060 | 0.000293 | 0.000487 | 0.000455 | 0.000444 |
| $\alpha_b$ (C) | −0.000266 | 0.000528 | −0.000085 | 0.000427 | 0.000392 | 0.000371 |
| $\alpha_b$ (D) | −0.000466 | −0.000200 | −0.000180 | 0.000248 | 0.000220 | 0.000226 |
| $\beta_b$ (Treatment level 2) | −0.000266 | 0.000528 | −0.000085 | 0.000427 | 0.000392 | 0.000371 |

**TABLE A.3** Stationary scenario with 10 replicates per treatment

| | $m = 15$ | | | | | |
|---|---|---|---|---|---|---|
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000615 | −0.001019 | −0.002341 | 0.000010 | 0.000011 | 0.000015 |
| $\alpha_b$ (A) | −0.002732 | −0.003323 | −0.003779 | 0.002346 | 0.002141 | 0.002661 |
| $\alpha_b$ (B) | 0.005498 | 0.000926 | −0.000151 | 0.003362 | 0.003195 | 0.003690 |
| $\alpha_b$ (C) | −0.000796 | 0.005020 | −0.000589 | 0.003293 | 0.003083 | 0.003530 |
| $\alpha_b$ (D) | −0.001693 | −0.001407 | 0.004398 | 0.002991 | 0.002723 | 0.003252 |
| $\beta_b$ (Treatment level 2) | −0.000796 | 0.005020 | −0.000589 | 0.003293 | 0.003083 | 0.003530 |
| | $m = 30$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000162 | −0.000302 | −0.000677 | 0.000000 | 0.000000 | 0.000001 |
| $\alpha_b$ (A) | −0.001049 | −0.001069 | −0.001229 | 0.000517 | 0.000472 | 0.000513 |
| $\alpha_b$ (B) | 0.001815 | −0.000493 | −0.000271 | 0.000845 | 0.000785 | 0.000882 |
| $\alpha_b$ (C) | −0.000252 | 0.002014 | −0.000233 | 0.000761 | 0.000727 | 0.000746 |
| $\alpha_b$ (D) | −0.000359 | −0.000522 | 0.001425 | 0.000623 | 0.000571 | 0.000639 |
| $\beta_b$ (Treatment level 2) | −0.000252 | 0.002014 | −0.000233 | 0.000761 | 0.000727 | 0.000746 |
| | $m = 60$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | 0.000067 | 0.000097 | 0.000160 | 0.000000 | 0.000000 | 0.000000 |
| $\alpha_b$ (A) | −0.000101 | −0.000403 | −0.000358 | 0.000131 | 0.000113 | 0.000117 |
| $\alpha_b$ (B) | 0.000368 | −0.000025 | −0.000025 | 0.000200 | 0.000188 | 0.000189 |
| $\alpha_b$ (C) | −0.000201 | 0.000261 | −0.000156 | 0.000176 | 0.000156 | 0.000147 |
| $\alpha_b$ (D) | −0.000074 | 0.000006 | 0.000147 | 0.000101 | 0.000092 | 0.000092 |
| $\beta_b$ (Treatment level 2) | −0.000201 | 0.000261 | −0.000156 | 0.000176 | 0.000156 | 0.000147 |

**TABLE A.4** Stationary scenario with 20 replicates per treatment

| | $m = 15$ | | | | | |
|---|---|---|---|---|---|---|
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000590 | −0.001063 | −0.002302 | 0.000005 | 0.000006 | 0.000010 |
| $\alpha_b$ (A) | −0.001230 | −0.001275 | −0.002300 | 0.001077 | 0.000994 | 0.001191 |
| $\alpha_b$ (B) | 0.002317 | 0.000162 | 0.000023 | 0.001519 | 0.001474 | 0.001638 |
| $\alpha_b$ (C) | 0.000471 | 0.001873 | 0.000326 | 0.001468 | 0.001423 | 0.001522 |
| $\alpha_b$ (D) | −0.000910 | −0.000203 | 0.002427 | 0.001362 | 0.001277 | 0.001483 |
| $\beta_b$ (Treatment level 2) | 0.000471 | 0.001873 | 0.000326 | 0.001468 | 0.001423 | 0.001522 |
| | $m = 30$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000162 | −0.000303 | −0.000669 | 0.000000 | 0.000000 | 0.000001 |
| $\alpha_b$ (A) | −0.000189 | −0.000441 | −0.000547 | 0.000247 | 0.000216 | 0.000249 |
| $\alpha_b$ (B) | 0.001162 | −0.000002 | 0.000094 | 0.000385 | 0.000355 | 0.000392 |
| $\alpha_b$ (C) | −0.000275 | 0.000529 | −0.000203 | 0.000358 | 0.000326 | 0.000349 |
| $\alpha_b$ (D) | −0.000565 | −0.000176 | 0.000596 | 0.000292 | 0.000265 | 0.000297 |
| $\beta_b$ (Treatment level 2) | −0.000275 | 0.000529 | −0.000203 | 0.000358 | 0.000326 | 0.000349 |
| | $m = 60$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | 0.000068 | 0.000100 | 0.000166 | 0.000000 | 0.000000 | 0.000000 |
| $\alpha_b$ (A) | 0.000009 | −0.000163 | −0.000086 | 0.000060 | 0.000052 | 0.000054 |
| $\alpha_b$ (B) | 0.000249 | 0.000073 | 0.000025 | 0.000096 | 0.000086 | 0.000085 |
| $\alpha_b$ (C) | −0.000040 | 0.000202 | −0.000033 | 0.000080 | 0.000073 | 0.000067 |
| $\alpha_b$ (D) | −0.000109 | −0.000106 | 0.000029 | 0.000045 | 0.000041 | 0.000042 |
| $\beta_b$ (Treatment level 2) | −0.000040 | 0.000202 | −0.000033 | 0.000080 | 0.000073 | 0.000067 |

studies for stationary and non-stationary models, in which four categories of response were considered (A, B, C, D, with A as the reference category) and a treatment effect with two levels (1 and 2). The size of the groups ($m$) varied between 15, 30, and 60. Finally, Tables A.8–A.10 present the confidence interval width and coverage rate for each parameter associated with the treatment effect, for the stationary and non-stationary scenarios. These tables compare the performance of the models fitted to the individual level data versus the models fitted to the grouped level data.

**TABLE A.5** Non-stationary scenario with five replicates per treatment

| | $m = 15$ | | | | | |
|---|---|---|---|---|---|---|
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000079 | 0.001846 | −0.001307 | 0.000068 | 0.000075 | 0.000074 |
| $\vartheta_b$ (time) | 0.000169 | −0.002944 | 0.000211 | 0.000163 | 0.000179 | 0.000172 |
| $\alpha_b$ (A) | −0.004643 | −0.013579 | 0.008584 | 0.031552 | 0.033251 | 0.045743 |
| $\alpha_b$ (B) | −0.010219 | 0.002751 | −0.004954 | 0.030798 | 0.033736 | 0.044403 |
| $\alpha_b$ (C) | 0.007937 | 0.009046 | 0.003854 | 0.030172 | 0.033125 | 0.049084 |
| $\alpha_b$ (D) | −0.000798 | 0.039226 | −0.000340 | 0.030498 | 0.033916 | 0.037783 |
| $\beta_b$ (Treatment level 2) | 0.032260 | −0.006983 | 0.077624 | 0.430344 | 0.448966 | 0.561741 |
| $\alpha_b^*$ (A) | 0.017207 | 0.011987 | −0.001454 | 0.070431 | 0.077631 | 0.088585 |
| $\alpha_b^*$ (B) | −0.009535 | −0.008155 | −0.000224 | 0.083593 | 0.095158 | 0.105165 |
| $\alpha_b^*$ (C) | −0.005261 | −0.010407 | −0.003004 | 0.072328 | 0.082262 | 0.100964 |
| $\alpha_b^*$ (D) | −0.005868 | −0.041276 | 0.000955 | 0.081002 | 0.092059 | 0.091754 |
| $\beta_b^*$ (Time×Treatment level 2) | −0.137483 | −0.053540 | −0.155052 | 2.093177 | 2.355302 | 2.474267 |
| | $m = 30$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000164 | 0.000977 | −0.000339 | 0.000002 | 0.000003 | 0.000002 |
| $\vartheta_b$ (Time) | 0.000134 | −0.001364 | 0.000189 | 0.000005 | 0.000007 | 0.000005 |
| $\alpha_b$ (A) | −0.005988 | −0.006965 | 0.004626 | 0.007229 | 0.008209 | 0.008056 |
| $\alpha_b$ (B) | −0.006847 | −0.000431 | −0.003673 | 0.006837 | 0.007820 | 0.007741 |
| $\alpha_b$ (C) | 0.005214 | 0.004038 | 0.003331 | 0.007210 | 0.007966 | 0.007868 |
| $\alpha_b$ (D) | −0.001036 | 0.038238 | 0.000836 | 0.007068 | 0.009230 | 0.007822 |
| $\beta_b$ (Treatment level 2) | 0.009415 | −0.027531 | 0.007415 | 0.187504 | 0.217812 | 0.209387 |
| $\alpha_b^*$ (A) | 0.015755 | 0.007726 | 0.002758 | 0.013785 | 0.014434 | 0.014390 |
| $\alpha_b^*$ (B) | −0.001922 | 0.000792 | 0.005144 | 0.015850 | 0.017712 | 0.017370 |
| $\alpha_b^*$ (C) | −0.006204 | −0.010720 | −0.003054 | 0.014493 | 0.015056 | 0.014938 |
| $\alpha_b^*$ (D) | −0.007047 | −0.040827 | −0.003109 | 0.015057 | 0.017633 | 0.016083 |
| $\beta_b^*$ (Time×Treatment level 2) | −0.041821 | −0.009301 | −0.058030 | 1.226131 | 1.229573 | 1.221885 |
| | $m = 60$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000185 | 0.000619 | 0.000208 | 0.000000 | 0.000000 | 0.000000 |
| $\vartheta_b$ (Time) | 0.000001 | −0.000723 | −0.000002 | 0.000000 | 0.000001 | 0.000000 |
| $\alpha_b$ (A) | −0.004991 | −0.004742 | 0.003660 | 0.001803 | 0.002327 | 0.001690 |
| $\alpha_b$ (B) | −0.005756 | −0.000435 | −0.002609 | 0.001805 | 0.002300 | 0.001667 |
| $\alpha_b$ (C) | 0.004196 | 0.001951 | 0.003928 | 0.001876 | 0.002286 | 0.001689 |
| $\alpha_b$ (D) | −0.002704 | 0.037596 | −0.000147 | 0.001720 | 0.003680 | 0.001665 |
| $\beta_b$ (Treatment level 2) | −0.001857 | −0.037069 | −0.019304 | 0.096146 | 0.131678 | 0.096636 |
| $\alpha_b^*$ (A) | 0.011856 | 0.004597 | 0.001139 | 0.003562 | 0.003685 | 0.003128 |
| $\alpha_b^*$ (B) | −0.001767 | 0.001117 | 0.005221 | 0.003951 | 0.004245 | 0.003758 |
| $\alpha_b^*$ (C) | −0.006102 | −0.009655 | −0.005031 | 0.003240 | 0.003471 | 0.002907 |
| $\alpha_b^*$ (D) | −0.002186 | −0.038381 | 0.000520 | 0.003206 | 0.005086 | 0.003115 |
| $\beta_b^*$ (Time×Treatment level 2) | 0.000793 | 0.024735 | 0.008448 | 1.132276 | 0.990963 | 1.035086 |

**TABLE A.6** Non-stationary scenario with 10 replicates per treatment

| | $m = 15$ | | | | | |
|---|---|---|---|---|---|---|
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000183 | 0.001650 | −0.001419 | 0.000021 | 0.000025 | 0.000023 |
| $\vartheta_b$ (Time) | 0.000358 | −0.002633 | 0.000437 | 0.000046 | 0.000054 | 0.000045 |
| $\alpha_b$ (A) | −0.007311 | −0.006617 | 0.003222 | 0.008090 | 0.008287 | 0.009579 |
| $\alpha_b$ (B) | −0.004727 | 0.000235 | −0.001907 | 0.007986 | 0.008601 | 0.009367 |
| $\alpha_b$ (C) | 0.004630 | 0.002766 | 0.003214 | 0.008013 | 0.008572 | 0.008984 |
| $\alpha_b$ (D) | −0.001873 | 0.038120 | 0.000934 | 0.007580 | 0.009892 | 0.009193 |
| $\beta_b$ (Treatment level 2) | 0.005273 | −0.028315 | 0.009258 | 0.129831 | 0.138814 | 0.157644 |
| $\alpha^*_b$ (A) | 0.016278 | 0.007604 | 0.004194 | 0.016754 | 0.017211 | 0.018529 |
| $\alpha^*_b$ (B) | −0.003248 | 0.000387 | 0.003608 | 0.018164 | 0.020420 | 0.021004 |
| $\alpha^*_b$ (C) | −0.008237 | −0.010611 | −0.005716 | 0.017499 | 0.019468 | 0.019160 |
| $\alpha^*_b$ (D) | −0.005410 | −0.040565 | −0.002425 | 0.018271 | 0.021394 | 0.020842 |
| $\beta^*_b$ (Time×Treatment level 2) | −0.019846 | −0.004296 | −0.042713 | 0.524576 | 0.592597 | 0.591808 |
| | $m = 30$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000172 | 0.000969 | −0.000342 | 0.000001 | 0.000002 | 0.000001 |
| $\vartheta_b$ (Time) | 0.000160 | −0.001343 | 0.000200 | 0.000002 | 0.000003 | 0.000002 |
| $\alpha_b$ (A) | −0.004958 | −0.003690 | 0.003921 | 0.001993 | 0.002083 | 0.002172 |
| $\alpha_b$ (B) | −0.004826 | −0.001412 | −0.002686 | 0.001878 | 0.002117 | 0.002086 |
| $\alpha_b$ (C) | 0.003084 | 0.002262 | 0.003526 | 0.001940 | 0.002202 | 0.002056 |
| $\alpha_b$ (D) | −0.002197 | 0.037490 | 0.000255 | 0.001935 | 0.003528 | 0.002055 |
| $\beta_b$ (Treatment level 2) | −0.005417 | −0.038943 | −0.015283 | 0.064429 | 0.073870 | 0.071370 |
| $\alpha^*_b$ (A) | 0.010779 | 0.003048 | 0.000850 | 0.003813 | 0.003843 | 0.003976 |
| $\alpha^*_b$ (B) | −0.001774 | 0.002221 | 0.005629 | 0.004317 | 0.004579 | 0.004635 |
| $\alpha^*_b$ (C) | −0.005249 | −0.009664 | −0.004188 | 0.003808 | 0.004144 | 0.003891 |
| $\alpha^*_b$ (D) | −0.002390 | −0.038003 | −0.000246 | 0.004040 | 0.005819 | 0.004279 |
| $\beta^*_b$ (Time×Treatment level 2) | 0.019825 | 0.023153 | −0.004039 | 0.357508 | 0.371852 | 0.367164 |
| | $m = 60$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000185 | 0.000617 | 0.000211 | 0.000000 | 0.000000 | 0.000000 |
| $\vartheta_b$ (Time) | −0.000005 | −0.000713 | 0.000002 | 0.000000 | 0.000001 | 0.000000 |
| $\alpha_b$ (A) | −0.004365 | −0.002348 | 0.003918 | 0.000520 | 0.000630 | 0.000494 |
| $\alpha_b$ (B) | −0.005202 | −0.001541 | −0.002556 | 0.000519 | 0.000610 | 0.000461 |
| $\alpha_b$ (C) | 0.002857 | 0.000840 | 0.003756 | 0.000506 | 0.000622 | 0.000453 |
| $\alpha_b$ (D) | −0.002526 | 0.037259 | −0.000098 | 0.000504 | 0.001979 | 0.000446 |
| $\beta_b$ (Treatment level 2) | −0.008159 | −0.043979 | −0.028705 | 0.033784 | 0.042331 | 0.032495 |
| $\alpha^*_b$ (A) | 0.009120 | 0.002119 | 0.000376 | 0.001093 | 0.001022 | 0.000908 |
| $\alpha^*_b$ (B) | −0.000203 | 0.003041 | 0.006296 | 0.001101 | 0.001167 | 0.001100 |
| $\alpha^*_b$ (C) | −0.005933 | −0.008925 | −0.005062 | 0.000884 | 0.001008 | 0.000788 |
| $\alpha^*_b$ (D) | −0.001589 | −0.037991 | 0.000472 | 0.000902 | 0.002395 | 0.000855 |
| $\beta^*_b$ (Time×Treatment level 2) | 0.043872 | 0.029732 | 0.021620 | 0.321577 | 0.293104 | 0.307585 |

**TABLE A.7** Non-stationary scenario with 20 replicates per treatment

| | $m = 15$ | | | | | |
|---|---|---|---|---|---|---|
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000185 | 0.001609 | −0.001390 | 0.000009 | 0.000012 | 0.000011 |
| $\vartheta_b$ (Time) | 0.000478 | −0.002529 | 0.000517 | 0.000020 | 0.000026 | 0.000019 |
| $\alpha_b$ (A) | −0.005417 | −0.003005 | 0.004100 | 0.003194 | 0.003237 | 0.003740 |
| $\alpha_b$ (B) | −0.005744 | −0.002203 | −0.002618 | 0.003203 | 0.003393 | 0.003598 |
| $\alpha_b$ (C) | 0.003299 | 0.002368 | 0.004255 | 0.003182 | 0.003430 | 0.003501 |
| $\alpha_b$ (D) | −0.001441 | 0.036722 | 0.000163 | 0.003145 | 0.004637 | 0.003523 |
| $\beta_b$ (Treatment level 2) | −0.002278 | −0.035800 | −0.017915 | 0.055593 | 0.059410 | 0.068905 |
| $\alpha_b^*$ (A) | 0.011769 | 0.003300 | 0.000383 | 0.006545 | 0.006739 | 0.007407 |
| $\alpha_b^*$ (B) | −0.001232 | 0.001640 | 0.004024 | 0.007400 | 0.008231 | 0.008285 |
| $\alpha_b^*$ (C) | −0.004656 | −0.008944 | −0.004130 | 0.006961 | 0.007591 | 0.007716 |
| $\alpha_b^*$ (D) | −0.004705 | −0.037614 | 0.000582 | 0.007313 | 0.009300 | 0.008221 |
| $\beta_b^*$ (Time×Treatment level 2) | 0.015365 | 0.011702 | 0.002749 | 0.218840 | 0.232765 | 0.246963 |
| | $m = 30$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000182 | 0.000955 | −0.000339 | 0.000000 | 0.000001 | 0.000000 |
| $\vartheta_b$ (Time) | 0.000154 | −0.001324 | 0.000203 | 0.000001 | 0.000002 | 0.000001 |
| $\alpha_b$ (A) | −0.004709 | −0.002285 | 0.003600 | 0.000825 | 0.000870 | 0.000881 |
| $\alpha_b$ (B) | −0.004526 | −0.001194 | −0.001979 | 0.000779 | 0.000890 | 0.000806 |
| $\alpha_b$ (C) | 0.002612 | 0.000742 | 0.003261 | 0.000787 | 0.000871 | 0.000795 |
| $\alpha_b$ (D) | −0.002569 | 0.036947 | 0.000221 | 0.000799 | 0.002207 | 0.000819 |
| $\beta_b$ (Treatment level 2) | −0.010015 | −0.043645 | −0.026470 | 0.028680 | 0.032600 | 0.032244 |
| $\alpha_b^*$ (A) | 0.009746 | 0.002413 | 0.001124 | 0.001611 | 0.001587 | 0.001606 |
| $\alpha_b^*$ (B) | −0.000978 | 0.002239 | 0.006212 | 0.001745 | 0.001967 | 0.001889 |
| $\alpha_b^*$ (C) | −0.006003 | −0.008800 | −0.004706 | 0.001552 | 0.001651 | 0.001522 |
| $\alpha_b^*$ (D) | −0.001554 | −0.037678 | −0.000485 | 0.001625 | 0.003113 | 0.001742 |
| $\beta_b^*$ (Time×Treatment level 2) | 0.042086 | 0.025354 | 0.010838 | 0.152398 | 0.154473 | 0.156488 |
| | $m = 60$ | | | | | |
| **Parameter** | **B (bias)** | **C (bias)** | **D (bias)** | **B (MSE)** | **C (MSE)** | **D (MSE)** |
| $\lambda_b$ (Intercept) | −0.000186 | 0.000616 | 0.000214 | 0.000000 | 0.000000 | 0.000000 |
| $\vartheta_b$ (Time) | −0.000005 | −0.000710 | 0.000002 | 0.000000 | 0.000001 | 0.000000 |
| $\alpha_b$ (A) | −0.004492 | −0.001535 | 0.003888 | 0.000226 | 0.000240 | 0.000209 |
| $\alpha_b$ (B) | −0.004624 | −0.001603 | −0.002152 | 0.000220 | 0.000256 | 0.000187 |
| $\alpha_b$ (C) | 0.002698 | 0.000354 | 0.003586 | 0.000212 | 0.000245 | 0.000191 |
| $\alpha_b$ (D) | −0.002883 | 0.036934 | −0.000134 | 0.000203 | 0.001603 | 0.000180 |
| $\beta_b$ (Treatment level 2) | −0.011449 | −0.049124 | −0.035157 | 0.014760 | 0.020159 | 0.015069 |
| $\alpha_b^*$ (A) | 0.008993 | 0.001512 | 0.000454 | 0.000483 | 0.000388 | 0.000367 |
| $\alpha_b^*$ (B) | −0.000657 | 0.002824 | 0.006122 | 0.000433 | 0.000476 | 0.000453 |
| $\alpha_b^*$ (C) | −0.005996 | −0.008392 | −0.004887 | 0.000368 | 0.000432 | 0.000326 |
| $\alpha_b^*$ (D) | −0.000925 | −0.037482 | 0.000375 | 0.000357 | 0.001789 | 0.000348 |
| $\beta_b^*$ (Time×Treatment level 2) | 0.053050 | 0.034962 | 0.027269 | 0.132274 | 0.118883 | 0.123014 |

**TABLE A.8** Comparing average 95% confidence interval width and coverage for the treatment-related parameters ($\beta_b$) for the stationary model fitted to the individual versus grouped level data

| | | | Grouped level data | | Individual level data | |
|---|---|---|---|---|---|---|
| **Replicates** | ***m*** | **Parameter** | **95% CI width** | **95% CI coverage** | **95% CI width** | **95% CI coverage** |
| 5 | 15 | $\beta_B$ | 1.5271 | 0.9489 | 1.2817 | 0.9499 |
| 5 | 15 | $\beta_C$ | 1.4842 | 0.9462 | 1.2606 | 0.9479 |
| 5 | 15 | $\beta_D$ | 1.6340 | 0.9509 | 1.4429 | 0.9488 |
| 5 | 30 | $\beta_B$ | 1.1428 | 0.9462 | 0.8964 | 0.9497 |
| 5 | 30 | $\beta_C$ | 1.1004 | 0.9479 | 0.8814 | 0.9469 |
| 5 | 30 | $\beta_D$ | 1.1697 | 0.9491 | 1.0069 | 0.9466 |
| 5 | 60 | $\beta_B$ | 0.8784 | 0.9499 | 0.6306 | 0.9488 |
| 5 | 60 | $\beta_C$ | 0.8345 | 0.9537 | 0.6201 | 0.9443 |
| 5 | 60 | $\beta_D$ | 0.8370 | 0.9516 | 0.7067 | 0.9504 |
| 10 | 15 | $\beta_B$ | 1.0212 | 0.9477 | 0.8964 | 0.9497 |
| 10 | 15 | $\beta_C$ | 0.9921 | 0.9509 | 0.8814 | 0.9469 |
| 10 | 15 | $\beta_D$ | 1.0909 | 0.9512 | 1.0069 | 0.9466 |
| 10 | 30 | $\beta_B$ | 0.7672 | 0.9475 | 0.6306 | 0.9488 |
| 10 | 30 | $\beta_C$ | 0.7390 | 0.9455 | 0.6201 | 0.9443 |
| 10 | 30 | $\beta_D$ | 0.7861 | 0.9490 | 0.7067 | 0.9504 |
| 10 | 60 | $\beta_B$ | 0.5943 | 0.9471 | 0.4446 | 0.9517 |
| 10 | 60 | $\beta_C$ | 0.5647 | 0.9520 | 0.4373 | 0.9498 |
| 10 | 60 | $\beta_D$ | 0.5668 | 0.9490 | 0.4982 | 0.9494 |
| 20 | 15 | $\beta_B$ | 0.7037 | 0.9490 | 0.6306 | 0.9488 |
| 20 | 15 | $\beta_C$ | 0.6845 | 0.9439 | 0.6201 | 0.9443 |
| 20 | 15 | $\beta_D$ | 0.7514 | 0.9457 | 0.7067 | 0.9504 |
| 20 | 30 | $\beta_B$ | 0.5305 | 0.9477 | 0.4446 | 0.9517 |
| 20 | 30 | $\beta_C$ | 0.5111 | 0.9485 | 0.4373 | 0.9498 |
| 20 | 30 | $\beta_D$ | 0.5433 | 0.9490 | 0.4982 | 0.9494 |
| 20 | 60 | $\beta_B$ | 0.4126 | 0.9514 | 0.3140 | 0.9471 |
| 20 | 60 | $\beta_C$ | 0.3922 | 0.9473 | 0.3089 | 0.9497 |
| 20 | 60 | $\beta_D$ | 0.3935 | 0.9499 | 0.3517 | 0.9528 |

**TABLE A.9** Comparing average 95% confidence interval width and coverage for the treatment-related parameters ($\beta_b$) for the non-stationary model fitted to the individual versus grouped level data. This table presents the results using 5 and 10 replicates

| | | | Grouped level data | | Individual level data | |
|---|---|---|---|---|---|---|
| **Replicates** | ***m*** | **Parameter** | **95% CI width** | **95% CI coverage** | **95% CI width** | **95% CI coverage** |
| 5 | 15 | $\beta_B$ | 2.3817 | 0.9469 | 1.8221 | 0.9465 |
| 5 | 15 | $\beta_C$ | 2.4512 | 0.9486 | 1.8268 | 0.9457 |
| 5 | 15 | $\beta_D$ | 2.6616 | 0.9497 | 2.1241 | 0.9556 |
| 5 | 15 | $\beta_B^*$ | 5.0565 | 0.9478 | 2.9547 | 0.9477 |
| 5 | 15 | $\beta_C^*$ | 5.3097 | 0.9504 | 3.1627 | 0.9499 |
| 5 | 15 | $\beta_D^*$ | 5.4123 | 0.9475 | 3.3233 | 0.9483 |
| 5 | 30 | $\beta_B$ | 1.6507 | 0.9521 | 1.2596 | 0.9491 |
| 5 | 30 | $\beta_C$ | 1.7419 | 0.9517 | 1.2622 | 0.9490 |
| 5 | 30 | $\beta_D$ | 1.7422 | 0.9508 | 1.4559 | 0.9502 |
| 5 | 30 | $\beta_B^*$ | 4.1111 | 0.9486 | 2.0231 | 0.9475 |
| 5 | 30 | $\beta_C^*$ | 4.1699 | 0.9455 | 2.1676 | 0.9487 |
| 5 | 30 | $\beta_D^*$ | 4.2096 | 0.9457 | 2.2706 | 0.9493 |
| 5 | 60 | $\beta_B$ | 1.1864 | 0.9483 | 0.8815 | 0.9481 |
| 5 | 60 | $\beta_C$ | 1.3188 | 0.9464 | 0.8827 | 0.9514 |
| 5 | 60 | $\beta_D$ | 1.1538 | 0.9441 | 1.0137 | 0.9504 |
| 5 | 60 | $\beta_B^*$ | 3.9073 | 0.9508 | 1.4088 | 0.9482 |
| 5 | 60 | $\beta_C^*$ | 3.6799 | 0.9477 | 1.5088 | 0.9516 |
| 5 | 60 | $\beta_D^*$ | 3.7803 | 0.9470 | 1.5782 | 0.9507 |
| 10 | 15 | $\beta_B$ | 1.3880 | 0.9448 | 1.2596 | 0.9491 |
| 10 | 15 | $\beta_C$ | 1.4227 | 0.9498 | 1.2622 | 0.9490 |
| 10 | 15 | $\beta_D$ | 1.5313 | 0.9472 | 1.4559 | 0.9502 |
| 10 | 15 | $\beta_B^*$ | 2.7743 | 0.9515 | 2.0231 | 0.9475 |
| 10 | 15 | $\beta_C^*$ | 2.8956 | 0.9527 | 2.1676 | 0.9487 |
| 10 | 15 | $\beta_D^*$ | 2.9468 | 0.9507 | 2.2706 | 0.9493 |
| 10 | 30 | $\beta_B$ | 0.9832 | 0.9497 | 0.8815 | 0.9481 |
| 10 | 30 | $\beta_C$ | 1.0302 | 0.9473 | 0.8827 | 0.9514 |
| 10 | 30 | $\beta_D$ | 1.0352 | 0.9449 | 1.0137 | 0.9504 |
| 10 | 30 | $\beta_B^*$ | 2.3071 | 0.9512 | 1.4088 | 0.9482 |
| 10 | 30 | $\beta_C^*$ | 2.3302 | 0.9483 | 1.5088 | 0.9516 |
| 10 | 30 | $\beta_D^*$ | 2.3515 | 0.9467 | 1.5782 | 0.9507 |
| 10 | 60 | $\beta_B$ | 0.7099 | 0.9482 | 0.6202 | 0.9534 |
| 10 | 60 | $\beta_C$ | 0.7776 | 0.9444 | 0.6209 | 0.9506 |
| 10 | 60 | $\beta_D$ | 0.6907 | 0.9473 | 0.7114 | 0.9509 |
| 10 | 60 | $\beta_B^*$ | 2.1557 | 0.9465 | 0.9887 | 0.9545 |
| 10 | 60 | $\beta_C^*$ | 2.0440 | 0.9500 | 1.0590 | 0.9485 |
| 10 | 60 | $\beta_D^*$ | 2.0960 | 0.9494 | 1.1069 | 0.9525 |

**TABLE A.10** Comparing average 95% confidence interval width and coverage for the treatment-related parameters ($\beta_b$) for the non-stationary model fitted to the individual versus grouped level data. This table presents the results using 20 replicates

| | | | Grouped level data | | Individual level data | |
|---|---|---|---|---|---|---|
| **Replicates** | ***m*** | **Parameter** | **95% CI width** | **95% CI coverage** | **95% CI width** | **95% CI coverage** |
| 20 | 15 | $\beta_B$ | 0.9238 | 0.9501 | 0.8815 | 0.9481 |
| 20 | 15 | $\beta_C$ | 0.9454 | 0.9466 | 0.8827 | 0.9514 |
| 20 | 15 | $\beta_D$ | 1.0170 | 0.9497 | 1.0137 | 0.9504 |
| 20 | 15 | $\beta_B^*$ | 1.8117 | 0.9457 | 1.4088 | 0.9482 |
| 20 | 15 | $\beta_C^*$ | 1.8873 | 0.9523 | 1.5088 | 0.9516 |
| 20 | 15 | $\beta_D^*$ | 1.9218 | 0.9472 | 1.5782 | 0.9507 |
| 20 | 30 | $\beta_B$ | 0.6571 | 0.9473 | 0.6202 | 0.9534 |
| 20 | 30 | $\beta_C$ | 0.6867 | 0.9464 | 0.6209 | 0.9506 |
| 20 | 30 | $\beta_D$ | 0.6925 | 0.9486 | 0.7114 | 0.9509 |
| 20 | 30 | $\beta_B^*$ | 1.5046 | 0.9468 | 0.9887 | 0.9545 |
| 20 | 30 | $\beta_C^*$ | 1.5178 | 0.9477 | 1.0590 | 0.9485 |
| 20 | 30 | $\beta_D^*$ | 1.5328 | 0.9492 | 1.1069 | 0.9525 |
| 20 | 60 | $\beta_B$ | 0.4756 | 0.9480 | 0.4374 | 0.9456 |
| 20 | 60 | $\beta_C$ | 0.5179 | 0.9304 | 0.4379 | 0.9479 |
| 20 | 60 | $\beta_D$ | 0.4634 | 0.9381 | 0.5016 | 0.9472 |
| 20 | 60 | $\beta_B^*$ | 1.3945 | 0.9454 | 0.6966 | 0.9487 |
| 20 | 60 | $\beta_C^*$ | 1.3265 | 0.9480 | 0.7463 | 0.9506 |
| 20 | 60 | $\beta_D^*$ | 1.3584 | 0.9474 | 0.7799 | 0.9507 |