Analysis of Crossover Designs with repeated measurements using Generalized Estimating Equations

Nelson Alirio Cruz Gutierrez

Universidad Nacional de Colombia
Departamento de Estadística
Facultad de Ciencias
Bogotá, Colombia
2022

Analysis of Crossover Designs with repeated measurements using Generalized Estimating Equations

Nelson Alirio Cruz Gutierrez

Tesis presentada como requisito parcial para optar al título de: Doctor en Ciencias Estadística

Director:
Ph.D. en Estadística, Oscar Orlando Melo Martínez
Codirector:
Ph.D. en Genética-Estadística, Carlos Alberto Martínez Niño

Grupo de Investigación:
Estadística Aplicada en Investigación Experimental, Industria y Biotecnología

Universidad Nacional de Colombia
Departamento de Estadística
Facultad de Ciencias
Bogotá, Colombia
2022

Dedication

While Pastor was telling me about that, he handed me the phone. Then, a man named Campo Elías convinced me to look for Statistics. He did not spend more than 15 minutes, not did he need eloquent words because a 15-year-old adolescent, ignorant of the world, would accept anything.

That afternoon I looked in the math books for statistics, I found bars, lines, frequencies and weather. I wasn't sure what to do, not did I know if I could study.

Thanks to the support of Campo Elias, Rubiola and Alicia, today I present this doctoral thesis.

Acknowledgments

I thank my parents for giving me the life and motivating me to study.
To my brothers for putting up with me for so long and listening to my empty words for hours.
I thank my friends (Andrés, Astrid, Fabian and Luna never stopped believing in me) for all the patience and help in this thesis.
To Oscar for directing my work, helping in its construction and serving as a voice of encouragement.
To Carlos for reviewing this work and helping in its construction.
To Laura for tolerating my late nights while I wrote the same thing 10 times.
To all the people who supported the development of this work.

Resumen: Análisis de diseños crossover con medidas repetidas mediante ecuaciones de estimación generalizadas

Los diseños experimentales crossover se usan ampliamente en medicina, agricultura y otras áreas de las ciencias biológicas. Por las características del diseño crossover, cada unidad experimental tiene observaciones longitudinales y presencia de efectos de arrastre en la variable respuesta. Además, en muchos escenarios no es posible dejar un período de lavado entre aplicaciones de diferentes tratamientos, lo que genera problemas al estimar los efectos del tratamiento sin una especificación adecuada del modelo. Como solución a lo anterior, esta tesis trata sobre diseños crossover sin período de lavado y con medidas repetidas.

En primer lugar, se desarrolla una metodología para el análisis de diseños crossover cuando la variable de respuesta es un conteo de Poisson. Para la estimación se utilizan ecuaciones de estimación generalizadas asumiendo que no existe período de lavado y que la unidad experimental fue observada una vez por período. Además, esta metodología es fácilmente extensible a cualquier variable de respuesta que pertenezca a la familia exponencial.

En un segundo lugar, la metodología anterior se extiende a diseños cruzados con medidas repetidas dentro de cada período, es decir, cuando una unidad experimental es observada más de una vez en cada período. Para este modelo se construye una familia de estructuras de correlación que toman en cuenta las particularidades del diseño, es decir, la correlación entre y dentro de los periodos.

En tercer lugar, se proporciona una extensión de las ecuaciones de estimación generalizadas que incluye un componente paramétrico para modelar los efectos del tratamiento y un componente no paramétrico para modelar los efectos del tiempo y los efectos carry-over. El componente no paramétrico se estima a partir de splines insertados en las ecuaciones de estimación generalizadas. Adicionalmente, se dan los códigos para la aplicación de la metodología en cualquier diseño crossover en el software estadístico R.

Las ventajas de la metodología propuesta se evidencian en ejercicios de simulación y explorando teóricamente las propiedades asintóticas de los estimadores obtenidos. También se compara el rendimiento de la metodología con las metodologías habituales sobre algunos datos reales de diseños cruzados. La metodología construida en esta tesis permite analizar cualquier diseño crossover siempre que la variable respuesta observada pertenezca a la familia exponencial, sin importar si hay periodo de lavado o no. Además, permite modelar medidas repetidas dentro de cada periodo y amplía las estructuras de correlación dentro de las ecuaciones de estimación generalizadas.

Palabras clave: Efecto de arrastre; Diseño cruzado; Ecuaciones de estimación generalizadas; Distribución de Poisson; Datos de conteo de sobredispersión; Correlación de Kronecker; Estimación de splines.

Abstract

Experimental crossover designs are widely used in medicine, agriculture, and other areas of the biological sciences. Due to the characteristics of the crossover design, each experimental unit has longitudinal observations and the presence of drag effects on the response variable. Furthermore, in many scenarios it is not possible to have a washout period between applications of different treatments, which creates problems in estimating treatment effects without a proper model specification. As a solution to this problem, this thesis deals with crossover designs without a washout period and with repeated measures.

First, a methodology is developed for the analysis of crossover designs when the response variable is a Poisson count. For the estimation, generalized estimation equations are used assuming that there is no washout period and that the experimental unit was observed once per period. Furthermore, this methodology is easily extended to any response variable that belongs to the exponential family.

Then, the above methodology is extended to crossover designs with repeated measures within each period, that is, when an experimental unit is observed more than once in each period. For this model, a family of correlation structures that takes into account the particularities of the design, that is, the correlation between and within the periods, is built.

Finally, an extension of the generalized estimating equations is developed. It includes a parametric component to model treatment effects and a nonparametric component to model time effects and carry-over effects. The non-parametric component is estimated from splines inserted into the generalized estimation equations. Additionally, the codes for the application of the methodology in any crossover design in the R statistical software are given.

The advantages of the proposed methodology are evidenced through simulation exercises and, theoretically, by exploring the asymptotic properties of the estimators obtained. The performance of the methodology is also compared with the usual methodologies on some real data from crossover designs. The methodology built in this thesis allows to analyze any crossover design as long as the observed response variable belongs to the exponential family, regardless of whether there is a washout period or not. It also allows modeling repeated measurements within each period and broadens the correlation structures used in the generalized estimation equations.

Keywords: Carry-over effect; Crossover design; Generalized estimating equations; Poisson distribution; Overdispersion count data; Kronecker correlation; Splines estimation

List of Figures

List of Tables

Contents

1. Introduction

Experimental design has been a very useful tool in many research areas, where it is of interest to observe the effect of treatments on experimental units. Generally, each experimental unit is randomly assigned to a treatment and one or more response variables are measured on it (Hinkelmann and Kempthorne, 2005). An experiment is cross-sectional if the measurements on each experimental unit are taken only once and longitudinal or repeated measures if they are taken sequentially in time (Davis, 2002). However, there are occasions in which one treatment is applied to one experimental unit in a first period of time, then in the second period of time, the treatment is changed for another one and so on. That is, each experimental unit is given a succession of treatments over time. These types of designs are known as crossover designs (Jones and Kenward, 2015). For the analysis of crossover experimental designs, repeated measures models with a single observation per period are usually proposed. So, in Madeyski and Kitchenham (2018) and Kitchenham et al. (2018), mixed linear models are used assuming that the response variable is normal and there is a single observation per period, including additive carry-over effects in the model.

Biabani et al. (2018) presented a review of papers on crossover designs used in neuroscience, and in all of them, the response variable was assumed to have a normal distribution, in addition to not including carry-over effects due to the presence of a washout period (a rest period between the application of one treatment and another), even when it only lasts a few days.

In Oh et al. (2003), Curtin (2017) and Li et al. (2018), generalized linear models were used for crossover designs for two periods and two sequences of two treatments, given a continuous (normal or gamma) or binary response variable and where the experimental unit was observed once in each period; they used generalized estimating equations (GEE) for parameter estimation. In Shkedy et al. (2005) and Liu and Li (2016), Bayesian generalized linear models were used to study crossover designs where the response was a lifetime and one observation per period was assumed.

Below, three crossover experiments that do not meet all assumptions for the application of the above mentioned methodologies are presented. In the first experiment, Chard et al. (2019) conducted a pilot study to investigate the impact of providing additional drinking water on the cognitive performance of students in two grades (fifth and sixth) in water-scarce schools in the rural areas from Mali. Forty-seven students were assigned to take the

control treatment (normal conditions) on the first day and received the treatment (controlled hydration) on the second day. 60 received the treatments in reverse order (Hydration on the first day and control on the second). One part of this test assessed visual attention. Two nearly identical images were presented side by side. Students were given one minute to circle the differences between the two images and the number of correct differences identified. The

Sequence		Period 1	Period 2
(1) AB	Ind 1	3 measurements	3 measurements
	⋮	⋮	⋮
	Ind 4	3 measurements	3 measurements
(1) BA	Ind 5	3 measurements	3 measurements
	⋮	⋮	⋮
	Ind 8	3 measurements	3 measurements

Table 1-1.: Structure of the crossover design in cows

second experiment was carried out in the Department of Animal Sciences of the Universidad Nacional de Colombia, which consisted of feeding two diets, A (grass) and B (mixture of grass and paper recycling waste) to a total of eight cows divided into two groups of four each. The first group received diet A from day 1 to day 42 and diet B from day 43 to day 84, while the second group received diet B from day 1 to day 42 and diet A from day 43 to day 84. Measurements of milk production and quality, as well as weight and body condition of the cows were made on days 1, 14, 28, 42, 56, 70 and 84 (Jaime, 2019). The structure of this design is shown in Table **1-1**.

At last, in the third experiment, Jones and Kenward (2015, pg 204) described a study where 3 treatments for blood pressure control were used. Treatment A consisted of a 20 mg dose of a test drug, treatment B was a 40 mg dose of the same drug, and treatment C was a placebo. Each one of six sequences of three periods (ABC, ACB, BCA, BAC, CAB, CBA) were applied to two individuals. During each application period, 10 successive measurements of systolic blood pressure were taken: 30 and 15 minutes before the application, and 15, 30, 45, 60, 75, 90, 120 and 240 minutes after the application, as shown in Table **1-2**. Each treatment was applied for six weeks and for ethical reasons its application was not suspended to leave a washout period.

To analyze such designs, Basu and Santra (2010), Josephy et al. (2015), Hao et al. (2015), Lui (2015), Rosenkranz (2015), Grayling et al. (2018), Madeyski and Kitchenham (2018) and Kitchenham et al. (2018) used mixed models for crossover designs with normal response and a single observation per period, including additively carry-over effects in the model. Diaz et al. (2013) and Forbes et al. (2015) normal mixed models are used to study crossover designs with repeated measures, but calculating the area under the curve to obtain a single

Sequence		Period 1	Period 2	Period 3
(1) ABC	Ind 1	10 measurements	10 measurements	10 measurements
	Ind 2	10 measurements	10 measurements	10 measurements
⋮		⋮	⋮	⋮
(6) CBA	Ind 11	10 measurements	10 measurements	10 measurements
	Ind 12	10 measurements	10 measurements	10 measurements

Table 1-2.: Structure of the blood pressure crossover design

observation per period and without including carry-over effects.

However, none of the above methodologies is suitable for the analysis of the three aforementioned experiments for the following reasons: i) It is not possible to leave a washout period between treatments because the hydration of the children, the feeding of the cows and the the application of blood pressure control medications cannot be suspended at any moment, for which it is necessary to take residual effects into account. ii) In all three studies, the placebo is part of the treatment design, so considering it as a washout period would modify the experiment. iii) For the second and third experiments, it is important to observe the response variables over time, and therefore, the data for each period cannot be summarized in a single datum. iv) In the event that the residual effects of each treatment are significant, the second and third experiments make it necessary to construct curves that explain this effect within the observed measurements. v) The response variable of the first and second experiment could be non-continuous, but rather counts or non-numerical ordinals, which makes the assumption of normality impossible. vi) For the second and third experiments, the correlation structure between each period could be different for the same individual, which makes it necessary to explore new block correlation structures, different from the structures used in repeated measurements.

In addition, generalized linear mixed models (GLMM) or generalized estimating equations (GEE) can be used to analyse longitudinal data. While the GLMM explicitly models within-subject correlation using random effects, the GEE implicitly explains such correlations using sandwich variance estimates (Liang and Zeger, 1986). Therefore, the GEE estimators yield a more robust inference about the fixed effects compared to the GLMM that depend on the distributional assumptions about the random effects (Zhang et al., 2012). Since the main goal of a crossover experiment is to obtain a robust inference about treatment effects, GEE models are more suitable than GLMM models.

In order to overcome the described limitations, a methodology of analysis for a crossover experimental design where the response variable is a Poisson count and there are in fact carry-over effects is laid out in the second chapter. This methodology is extended to data

with overdispersion or underdispersion. The theoretical development for the analysis of cases with few treatments and few periods is presented. Under such conditions, the log-linear link for estimation purposes and the Delta method are considered for the asymptotic inference of the estimators. When the number of periods and sequences increases, an extension of this methodology is proposed using the GEE. In this extension, crossover designs for count data include treatments, sequences, time effects, covariates, and any correlation structures. The most important result of the GEE methodology is that it allows detecting significant effects within the crossover design when the response variable belongs to the exponential family, especially those related to the treatment effects. A simulation study making the comparison between the usual methods of analysis and those developed in the present work is shown, demonstrating the advantage over the usual methods in situations with the presence of carry-over effects. Finally, the analysis of the data obtained in the student hydration study is presented.

In the third chapter, a family of correlation structures for repeated measures crossover designs for Gaussian and non Gaussian responses using GEE is presented. The structure considers two matrices: one that models the between-periods correlation and another that models the intra-period correlation. The general correlation matrix, which is used to build the GEE, corresponds to the Kronecker between these matrices. A procedure to estimate the parameters of the correlation matrix is proposed, its statistical properties are studied and it is compared to standard models using a single correlation matrix. A simulation study showed superior performance of the proposed structure in terms of the quasilikelihood criterion, efficiency, and ability to explain patterns of complex correlation phenomena in longitudinal data from crossover designs. In addition, the methodology was applied to the blood pressure and cow feeding experiments.

Finally, in the fourth chapter, a model for crossover designs with repeated measures within each period was developed. It is obtained using an extension of GEE that includes a parametric component to model treatment effects and a nonparametric component to model time and carry-over effects; the estimation approach for the nonparametric component is based on splines. A simulation study was carried out to explore the properties of the model. The results shown that there is either a carry-over effect or a functional temporal effect, or both, where the proposed model yields better results than the standard models. Regarding the theoretical properties, it was found that the parameter estimation was analogous to weighted least squares. Therefore, model diagnostics can be performed adapting techniques used in multiple regression. The proposed methodology was implemented in the data sets of the experiments on systolic blood pressure and insulin in rabbits. In this study a robust methodology for the analysis of repeated measures crossover designs was developed through GEE. This methodology is applicable to multiple crossover experiments, regardless of the structure of the design or the configuration of the washout period. It also allows the analysis

of any response variable that belongs to the exponential family.
In each of the previous chapter, the proposed methodology offers better results than the traditional one in three key aspects: i) the asymptotic properties of estimators, ii) the simulation results and iii) the analysis of real data. This provides a better framework for the statistical analysis of crossover designs.

2. Analysis of crossover experiments with count data in the presence of carry-over effects

The content of this chapter was published as a research article in the journal Statistica Neerlandica (Cruz et al., 2023a). [1]

2.1. Introduction

Experimental designs are a very critical tool for the analysis of applied treatments on experimental units (Hinkelmann and Kempthorne, 2005). For this kind of studies, it is common the simultaneous observation of all experimental units, commonly named as cross-sectional studies (Melo et al., 2007). The main advantage of cross-sectional studies is the absence of time effects in the analysis, but sometimes it is necessary to evaluate the response variable in several instants of time, which means that investigators must include them (Hardin and Hilbe, 2003). When time is added, this kind of studies is known as longitudinal experiments, in which each unit receives the treatment at the first moment, upholds it throughout the study; for example, in diets for animal growth (Davis, 2002). However, sometimes the experimental unit receives a treatment at the first moment of time, which is changed to another at the second moment, and so on throughout the study. Therefore, each experimental unit receives a sequence of treatments over time (Patterson, 1951). This is important, for example, in chronic disease's treatments in humans, or in crop pest's control, for avoiding drug resistance, and to determine the evolution of disease behavior. These designs are known as crossover designs (Jones and Kenward, 2015).

In the construction methodology of crossover designs, Chalikias and Kounias (2012) made the ψ-optimal crossover designs with two treatments, n subjects and p periods. As well in Bailey et al. (2017), the universally optimality of circular weakly balanced designs for different types of models was proved. Moreover, Chalikias (2017) worked with two-treatment circular and non-circular crossover designs. Furthermore, Chasiotis and Kounias (2021) found and built

[1] https://doi.org/10.1111/stan.12295

the ψ-optimal circular crossover designs with two treatments, any number of experimental subjects, and two, three as well as four periods of equal size for the estimation of contrasts of direct and carry-over effects. At last, a recent paper, Chasiotis (2021) determined and constructed the ψ-optimal designs over the class of all circular repeated measurement's designs with two treatments when the number of both periods and subjects was even. In addition, the author was interested in estimating the contrasts of both direct and carry-over effects jointly, and minimizing the variances of the best linear unbiased estimators of contrasts of direct as well as carry-over effects.

In the context of analysis of experimental crossover designs, most of the analysis assumes normality models as one of its pillars. However, it is often necessary to analyze experimental with response variables as counts, proportions, binary data, among others (Kenward and Jones, 1987). Recent research is focused on the binary data analysis, that is, two possible outcomes for the variable of interest: success or failure. For this kind of analysis, generalized linear models for binary data are used (Jones and Kenward, 2015).

When the response variable in a crossover design is a count, Layard and Arvesen (1978) builds a methodology assuming independence within the experimental units and absence of carry-over effects, Lindsey and Jones (1997) used mixed generalized linear models but without including correlation within subjects, Longford (1998) used linear mixed models including correlation but only for a 2×2 crossover design and supposed absence of carry-over effects and Lui (2016) used a linear log model for 2×2 crossover design, but did not include overdispersion or carry-over effects and assumed independence within subjects. Therefore, in this paper a first way to approach a crossover design that shows this distribution is the log-linear model including carry-over effects in addition to a correlation structure. For a crossover design with Poisson response, estimators are linear combinations of the logarithms of the means. From the Delta method, the asymptotic distribution of these estimators is obtained, and inference is performed. Since it is very restrictive that the mean of the counts is equal to its variance, an extension is given to an answer with overdispersion or subdispersion.

However, this methodology is not adequate if it is found that significant covariates are affecting the crossover design, e.g., blocking sources, where a more complex analysis of a crossover design using generalized estimation equations is developed: model, estimation, and inference. The main contribution of this chapter is a deviance analysis table obtained from the development of the methodology. This table allows to select the significant effects of the design, especially the treatment effect.

In this chapter, the analysis of data obtained in an hydration student study is presented as an example. A comparison between usual methods of analysis and those obtained in the present work, testing the new methodology for effectiveness, assuming the count quality of the response variable.

2.2. Crossover design

A crossover design consists of the following components (Jones and Kenward, 2015): i) Sequences: each of the different combinations of treatments randomly assigned to be applied chronologically to experimental units. The total number of sequences is denoted by S. ii) Treatment: applied to each experimental unit within each sequence in a moment of time, t denoted the total number of treatments. iii) Periods: instant of time in which a treatment is applied following a sequence of treatments. When all sequences have the same length, the number of periods is equal to P, the length of any sequence. v) Experimental units: individuals or elements that receive the treatment, the total of units in the study is denoted by n.

The carry-over effects are sometimes part of the crossover design. If the k-th treatment is applied to an individual in a period l and j periods ahead ($j > 0$), that is, in period $l+j$ the treatment k^* is used, then a residual effect is generated between the treatments or a "drag effect" called carry-over defined as θ_j^k. When $j = 1$, it is called first order carry-over effect, when $j = 2$ second order, and so on (Patterson, 1951).

A crossover design with a total of S sequences to be applied in P different periods of time has been considered. Y_{ij} is the response of the i-th experimental unit in the j-th period. Finally, $\boldsymbol{Y}$ is the vector of all observed responses, as it is shown in equation (2-1):

$$\boldsymbol{Y} = (\boldsymbol{Y}_{11}, \ \ldots, \boldsymbol{Y}_{1P}, \ \ldots, \boldsymbol{Y}_{nP})^t \tag{2-1}$$

The mean $\overline{Y}_{(k)j}$ is defined as the average response in the k-th sequence and the j-th period, and the vector of the means in equation (2-2):

$$\overline{\boldsymbol{Y}} = \left(\overline{Y}_{(1)1}, \ \ldots, \overline{Y}_{(S)1}, \ \ldots, \overline{Y}_{(S)P}\right)^t, \qquad \text{where } \overline{Y}_{(k)j} = \frac{1}{n_k}\sum_{i=1}^{n} y_{ij}\delta_{i(k)} \tag{2-2}$$

where $\delta_{i(k)}$ is an indicator function, its value is 1 if the i-th experimental unit was located in sequence k and 0 otherwise and n_k is the number of experimental units in the k-th sequence. The covariance matrix of $\overline{\boldsymbol{Y}}$ is presented in equation (2-3)

$$\boldsymbol{\Sigma}_{\overline{\boldsymbol{Y}}} = \begin{bmatrix} \frac{1}{n_1}\boldsymbol{\Sigma}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \frac{1}{n_S}\boldsymbol{\Sigma}_S \end{bmatrix} \tag{2-3}$$

where

$$\boldsymbol{\Sigma}_k = \begin{bmatrix} Var(\overline{Y}_{(k)1}) & \sigma_{(k)(12)} & \sigma_{(k)(13)} & \ldots & \sigma_{(k)(1P)} \\ \sigma_{(k)(12)} & Var(\overline{Y}_{(k)2}) & \sigma_{(k)(23)} & \ldots & \sigma_{(k)(2P)} \\ \sigma_{(k)(13)} & \sigma_{(k)(23)} & Var(\overline{Y}_{(k)3}) & \ldots & \sigma_{(k)(3P)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{(k)(1P)} & \sigma_{(k)(2P)} & \sigma_{(k)(3P)} & \ldots & Var(\overline{Y}_{(k)T}) \end{bmatrix} \tag{2-4}$$

with $\sigma_{(k)(jj')} = Cov\left(\overline{Y}_{(k)j}, \overline{Y}_{(k)j}\right)$.
In next section, the Poisson crossover design concepts for analysis are introduced.

2.2.1. Poisson crossover design

Let $\boldsymbol{\mu}$ be the expected mean cell vector of the crossover design, with each element of $\boldsymbol{\mu}$ given by $\mu_{ij} = E(Y_{ij})$, and if $Y_{ij} \sim Poisson(\mu_{ij})$ then the log-linear model is given by

$$\ln(\mu_{(k)j}) = \mu + \tau_{(k)j} + \pi_j + \theta_{(k)j} \tag{2-5}$$

where $\mu_{(k)j} = E(Y_{(k)j})$, $\tau_{(k)j}$ is the treatment effect applied within k-th sequence of the j-th period and $\theta_{(k)j}$ is the carry-over effect within the k-th sequence of the j-th period. If $j = 1$ then $\theta_{(k)1} = 0$, because in period 1 there are no carry-over effects. The estimability conditions, similar to normal and binomial models for the parameters (Jones and Kenward, 2015), are given by:

$$\sum_{k=1}^{S}\sum_{j=1}^{P}\tau_{(k)j} = 0, \quad \sum_{k=1}^{S}\sum_{j=2}^{P}\theta_{(k)j} = 0, \quad \sum_{j=1}^{P}\pi_j = 0 \tag{2-6}$$

Result 1: Under the assumption that $Y_{(k)j} \sim Poisson(\mu_{(k)j})$ and the correct model for the mean is given by the equation (2-5), $\overline{\boldsymbol{Y}}$ is an unbiased and consistent estimator for $\boldsymbol{\mu}$ with normal asymptotic distribution with mean $\boldsymbol{\mu}$ and covariance matrix given in equation (2-3).

Result 1. Proof:

$$E(\overline{\boldsymbol{Y}}) = \left(E(\overline{Y}_{(1)1}),\ \ldots, E(\overline{Y}_{(1)P}),\ \ldots, E(\overline{Y}_{(S)P})\right)^t = \left(\mu_{(1)1},\ \ldots, \mu_{(1)P},\ \ldots, \mu_{(S)P}\right)^t = \boldsymbol{\mu}$$

$$\lim_{n_1,\ldots,n_s\to\infty} Var(\overline{\boldsymbol{Y}}) = \lim_{n_1,\ldots,n_s\to\infty} \boldsymbol{\Sigma}_{\overline{\boldsymbol{Y}}} = \lim_{n_1,\ldots,n_s\to\infty} \bigoplus_{i=1}^{S} \frac{1}{n_i}\boldsymbol{\Sigma}_i$$

$$\lim_{n_1,\ldots,n_s\to\infty} Var(\overline{\boldsymbol{Y}}) = \bigoplus_{i=1}^{S} \lim_{n_i\to\infty} \frac{1}{n_i}\boldsymbol{\Sigma}_i = \bigoplus_{i=1}^{S} (\boldsymbol{\Sigma}_i) \lim_{n_i\to\infty} \frac{1}{n_i} = \mathbf{0}$$

Since $\overline{\boldsymbol{Y}}$ is unbiased and its variance tends to zero when $n_i \to \infty$, then it is consistent. By using the central limit theorem, it is obtained:

$$\overline{\boldsymbol{Y}} \xrightarrow{D} N_{ST}\left(\boldsymbol{\mu}, \bigoplus_{i=1}^{S} \frac{1}{n_i}\boldsymbol{\Sigma}_i\right) \tag{2-7}$$

□

When $Y_{ij} \sim Poisson(\mu_{ij})$ then $Var(Y_{ij}) = E(Y_{ij}) = \mu_{ij}$ and $n_k Var(\overline{Y}_{(k)j}) = \mu_{(k)j}$, for which the equation (2-4) becomes:

$$\boldsymbol{\Sigma}_k = \begin{bmatrix} \mu_{(k)1} & \sigma_{(k)(12)} & \sigma_{(k)(13)} & \cdots & \sigma_{(k)(1P)} \\ \sigma_{(k)(12)} & \mu_{(k)2} & \sigma_{(k)(23)} & \cdots & \sigma_{(k)(2P)} \\ \sigma_{(k)(13)} & \sigma_{(k)(23)} & \mu_{(k)3} & \cdots & \sigma_{(k)(3P)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{(k)(1P)} & \sigma_{(k)(2P)} & \sigma_{(k)(3P)} & \cdots & \mu_{(k)P} \end{bmatrix} \tag{2-8}$$

As many times the counts may present overdispersion or subdispersion using Poisson distribution (Kokonendji et al., 2004), an additional parameter has to be placed on to control this phenomenon. Let $\phi > 0$ and $Var(Y_{ij}) = \phi E(Y_{ij}) = \phi\mu_{ij}$, then if $\phi < 1$, the variance is less than the mean (subdispersion); whereas if $\phi > 1$, then there is overdispersion. The variance given in equation (2-4), with the dispersion parameter, is determined as:

$$\boldsymbol{\Sigma}_{(k)} = \begin{bmatrix} \phi\mu_{(k)1} & \sigma_{(k)(12)} & \sigma_{(k)(13)} & \cdots & \sigma_{(k)(1P)} \\ \sigma_{(k)(12)} & \phi\mu_{(k)2} & \sigma_{(k)(23)} & \cdots & \sigma_{(k)(2P)} \\ \sigma_{(k)(13)} & \sigma_{(k)(23)} & \phi\mu_{(k)3} & \cdots & \sigma_{(k)(3P)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{(k)(1P)} & \sigma_{(k)(2P)} & \sigma_{(k)(3P)} & \cdots & \phi\mu_{(k)P} \end{bmatrix} \tag{2-9}$$

Result 2: Suppose a crossover design with S sequences of length P and response variable Y_{ij} with marginal Poisson distribution and covariance matrix defined in equation (2-3). Let $\boldsymbol{X}$ be the design matrix for the model defined in equation (2-5), $\boldsymbol{\beta} \in \mathbb{R}^p$, the vector with true unknown parameters (μ, τ's, θ's and π's) of the model defined in equation (2-5) and $\mu_{(k)j} = \exp(\boldsymbol{x}^t_{(k)j}\boldsymbol{\beta})$ the log-linear link. If $\boldsymbol{X} = (\boldsymbol{x}_{(1)1}, \boldsymbol{x}_{(1)2}, \ldots, \boldsymbol{x}_{(1)P}, \ldots, \boldsymbol{x}_{(S)P})^t$ is a design matrix of full column range, which contains the factors associated to the vector $\boldsymbol{\beta}$, then exist linear combinations of the vector $(\ln(\overline{Y}_{(1)1}), \ldots, \ln(\overline{Y}_{(1)P}), \ldots, \ln(\overline{Y}_{(S)P}))^t$ which are asymptotically unbiased estimators for each component of the vector $\boldsymbol{\beta}$.

Result 2. Proof: Assume that $\boldsymbol{X}$ is the design matrix and the system of equations (2-10)

$$\begin{aligned} &\begin{pmatrix} \boldsymbol{x}_{(1)1} & \ldots & \boldsymbol{x}_{1T} & \ldots & \boldsymbol{x}_{ST} \end{pmatrix}^t \boldsymbol{\beta} = \begin{pmatrix} \ln(\mu_{(1)1}) & \ldots & \ln(\mu_{1T}) & \ldots & \ln(\mu_{ST}) \end{pmatrix} \\ &\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\mu}^* \end{aligned} \tag{2-10}$$

where $\boldsymbol{\mu}^* = \left(\ln(\mu_{(1)1}), \ln(\mu_{(1)2}), \ldots, \ln(\mu_{(1)P}), \ldots, \ln(\mu_{(S)P})\right)^t$. Then, if $\boldsymbol{X}$ is a full column range, the matrix $\boldsymbol{X}^t\boldsymbol{X}$ is a full-range square matrix for which it exists $(\boldsymbol{X}^t\boldsymbol{X})^{-1}$, and thus, for the equation (2-10):

$$\begin{aligned} \boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta} &= \boldsymbol{X}^t\boldsymbol{\mu}^* \\ (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta} &= (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{\mu}^* \\ \boldsymbol{\beta} &= (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{\mu}^* \end{aligned} \tag{2-11}$$

Let $\boldsymbol{X}_k^*$ be the k-th row of the matrix $(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t$, then, the function h_a is constructed:

$$\begin{aligned} h_a : \mathbb{R}^{SP} &\xrightarrow{h_a} \mathbb{R} \\ \boldsymbol{\mu} &\longrightarrow \boldsymbol{X}_a^*\boldsymbol{\mu^*} \\ h_a(\boldsymbol{\mu}) &= \boldsymbol{X}_a^* \left(\ln(\mu_{(1)1}), \ln(\mu_{(1)2}), \ldots, \ln(\mu_{(1)P}), \ldots, \ln(\mu_{(S)P})\right)^t \end{aligned}$$

and by using the equation (2-11), it is obtained $h_a(\boldsymbol{\mu}) = \beta_a$ where β_a is the a-th component of the $\boldsymbol{\beta}$ vector.
By applying the Result 1, $\overline{\boldsymbol{Y}} \xrightarrow{D} N_{ST}\left(\boldsymbol{\mu}, \bigoplus_{k=1}^{S} \frac{1}{n_k}\boldsymbol{\Sigma}_k\right)$, and using the Delta method (Agresti, 2002, page 279) in the function h_a evaluated in $\overline{\boldsymbol{Y}}$, it is satisfied that $\boldsymbol{X}_a^*\overline{\boldsymbol{Y}}$ is an asymptotically unbiased estimator with normal asymptotic distribution with variance given by:

$$Var(\boldsymbol{X}_a^*\overline{\boldsymbol{Y}}) \xrightarrow{P} \boldsymbol{\nabla}^t_{h_a(\overline{\boldsymbol{Y}})}\boldsymbol{\Sigma}_{\overline{\boldsymbol{Y}}}\boldsymbol{\nabla}_{h_a(\overline{\boldsymbol{Y}})} \tag{2-12}$$

□ .
Now, this methodology is applied to the analysis of a 2×2 crossover design. The construction of the $\boldsymbol{X}$ and $\boldsymbol{\beta}$ matrices for this design is also illustrated.

2×2 crossover design

A 2×2 experimental design is the most common crossover design (Jones and Kenward, 2015), besides being the simplest one. The model structure is presented in Table **2-1**. To analyze this design, the Poisson model given in equation (2-5) is used.

Sequence	Period 1	Period 2
AB	Y_{11}	Y_{12}
	$\vdots$	$\vdots$
	$Y_{n_1 1}$	$Y_{n_1 2}$
BA	$Y_{n_1+1\,1}$	$Y_{n_1+1\,2}$
	$\vdots$	$\vdots$
	Y_{n1}	Y_{n2}

Table 2-1.: 2×2 crossover design

The fixed model effects based in equation (2-5) are as in Table **2-2**, where μ is the average effect, π_1 is the effect of period 1, π_2 of period 2, $\tau_{(1)2}$ is the effect of treatment applied in first sequence and first period, $\tau_{(1)1}$ is the effect of treatment applied in first sequence and second period, $\tau_{(2)1}$ is the effect of treatment applied in second sequence and first period, $\tau_{(2)2}$ is the effect of treatment applied in second sequence and second period, $\theta_{(1)2}$ is the residual effect

Sequence	Period 1	Period 2
AB	$\exp(\mu + \pi_1 + \tau_{(1)1})$	$\exp(\mu + \pi_2 + \tau_{(1)2} + \theta_{(1)2})$
BA	$\exp(\mu + \pi_1 + \tau_{(2)1})$	$\exp(\mu + \pi_2 + \tau_{(2)2} + \theta_{(2)2})$

Table 2-2.: Fixed model effects of 2×2 crossover design

(carry-over) present over treatment applied in first sequence and second period and $\theta_{(2)2}$ is the residual effect (carry-over) present over treatment applied in second sequence and second period. Since $\tau_{(1)1}$ and $\tau_{(2)2}$ are given by the same treatment A, then they are assumed equal and denoted by τ_A, and analogously $\tau_{(2)1}$ and $\tau_{(1)2}$ are assumed equal and will be denoted by τ_B. With these specifications, the estimability restrictions given by the equation (2-6) will be as follows:

$$\pi_2 = -\pi_1 = \pi \qquad \tau_A = -\tau_B = \tau \qquad \theta_{(1)2} = -\theta_{(2)2} = \theta \tag{2-13}$$

Under the model given in Table **2-2** and equation (2-13), the expected values for each cell of Table **2-1** are given by $\mu_{(1)1} = \exp(\mu - \pi - \tau)$, $\mu_{(1)2} = \exp(\mu + \pi + \tau + \theta)$, $\mu_{(2)1} = \exp(\mu - \pi + \tau)$ and $\mu_{(2)2} = \exp(\mu + \pi - \tau - \theta)$. Now consider the following system of equations:

$$\begin{aligned}
\ln(\overline{Y}_{(1)1}) &= \hat{\mu} - \hat{\pi} - \hat{\tau} \\
\ln(\overline{Y}_{(1)2}) &= \hat{\mu} + \hat{\pi} + \hat{\tau} + \hat{\theta} \\
\ln(\overline{Y}_{(2)1}) &= \hat{\mu} - \hat{\pi} + \hat{\tau} \\
\ln(\overline{Y}_{(1)2}) &= \hat{\mu} + \hat{\pi} - \hat{\tau} - \hat{\theta}
\end{aligned}$$

This system of equations can be raised in matrix form to illustrate the use of Result 2, as in the following equation:

$$\begin{bmatrix} 1 & -1 & -1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\pi} \\ \hat{\tau} \\ \hat{\theta} \end{bmatrix} = \begin{bmatrix} \ln(\overline{Y}_{(1)1}) \\ \ln(\overline{Y}_{(1)2}) \\ \ln(\overline{Y}_{(2)1}) \\ \ln(\overline{Y}_{(1)2}) \end{bmatrix} \tag{2-14}$$

$$\boldsymbol{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} \ln(\overline{Y}_{(1)1}) \\ \ln(\overline{Y}_{(1)2}) \\ \ln(\overline{Y}_{(2)1}) \\ \ln(\overline{Y}_{(1)2}) \end{bmatrix}$$

where

$$\boldsymbol{X} = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & 1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} \hat{\mu} \\ \hat{\pi} \\ \hat{\tau} \\ \hat{\theta} \end{bmatrix}$$

Solving the given system of equations in (2-14), it is obtain that:

$$\begin{bmatrix} \hat{\mu} \\ \hat{\pi} \\ \hat{\tau} \\ \hat{\theta} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \ln(\overline{Y}_{(1)1}) \\ \ln(\overline{Y}_{(1)2}) \\ \ln(\overline{Y}_{(2)1}) \\ \ln(\overline{Y}_{(1)2}) \end{bmatrix} = \begin{bmatrix} \frac{1}{4}\ln\left(\overline{Y}_{(1)2}\overline{Y}_{(2)2}\overline{Y}_{(1)1}\overline{Y}_{(2)1}\right) \\ \frac{1}{4}\ln\left(\frac{\overline{Y}_{(1)2}\overline{Y}_{(2)2}}{\overline{Y}_{(1)1}\overline{Y}_{(2)1}}\right) \\ \frac{1}{2}\ln\left(\frac{\overline{Y}_{(1)1}}{\overline{Y}_{(2)1}}\right) \\ \frac{1}{2}\ln\left(\frac{\overline{Y}_{(1)1}\overline{Y}_{(1)2}}{\overline{Y}_{(2)1}\overline{Y}_{(2)2}}\right) \end{bmatrix} \tag{2-15}$$

Therefore, an estimator of the carry-over effect is given by the following equation:

$$\hat{\theta} = \frac{1}{2}\ln\left(\frac{\overline{Y}_{(1)1}\overline{Y}_{(1)2}}{\overline{Y}_{(2)1}\overline{Y}_{(2)2}}\right) \tag{2-16}$$

To find the asymptotic distribution of the estimator $\hat{\theta}$, the Result 2 is used, and the $\overline{\boldsymbol{Y}}$ covariance matrix is given by:

$$\boldsymbol{\Sigma}_{\overline{\boldsymbol{Y}}} = \begin{bmatrix} \frac{\mu_{(1)1}}{n_1} & \frac{\sigma_{1(12)}}{n_1} & 0 & 0 \\ \frac{\sigma_{1(12)}}{n_1} & \frac{\mu_{(1)2}}{n_1} & 0 & 0 \\ 0 & 0 & \frac{\mu_{(2)1}}{n_2} & \frac{\sigma_{2(12)}}{n_2} \\ 0 & 0 & \frac{\sigma_{2(12)}}{n_2} & \frac{\mu_{(2)2}}{n_2} \end{bmatrix} \tag{2-17}$$

where $\sigma_{1(12)}$ is the covariance between the two measurements of the response variable in an individual from the first sequence (AB) of treatments, and $\sigma_{2(12)}$ between the two measurements of the response variable in an individual from the second sequence (BA) of treatments. Using the Delta method (Agresti, 2002, page 279) and since

$$\theta = h(\boldsymbol{\mu}) = \frac{1}{2}\ln\left(\frac{\mu_{(1)1}\mu_{(1)2}}{\mu_{(2)1}\mu_{(2)2}}\right)$$

then the gradient of h is given by:

$$\boldsymbol{\nabla} h^t = \frac{\partial\theta}{\partial\boldsymbol{\mu}} = \left(\frac{1}{\mu_{(1)1}} + \frac{1}{\mu_{(1)2}} - \frac{1}{\mu_{(2)1}} - \frac{1}{\mu_{(2)2}}\right)$$

Therefore, the estimator of $\hat{\theta}$ satisfies that:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N(\boldsymbol{0}, \boldsymbol{\nabla} h^t \boldsymbol{\Sigma}_{\overline{Y}} \boldsymbol{\nabla} h)$$

where

$$\boldsymbol{\nabla} h^t \boldsymbol{\Sigma}_{\overline{Y}} \boldsymbol{\nabla} h = \frac{1}{4n_1}\left(\frac{1}{\mu_{(1)1}} + \frac{1}{\mu_{(1)2}} + \frac{2\sigma_{1(12)}}{\mu_{(1)1}\mu_{(1)2}}\right) + \frac{1}{4n_2}\left(\frac{1}{\mu_{(2)1}} + \frac{1}{\mu_{(2)2}} + \frac{2\sigma_{2(12)}}{\mu_{(2)1}\mu_{(2)2}}\right)$$

Calculating the estimator of the variance of the carry-over effect, it is obtained that

$$\hat{V}(\hat{\boldsymbol{\theta}}) = \frac{1}{4n_1}\left(\frac{1}{\overline{Y}_{(1)1}} + \frac{1}{\overline{Y}_{(1)2}} + \frac{2\hat{\sigma}_{1(12)}}{\overline{Y}_{(1)1}\overline{Y}_{(1)2}}\right) + \frac{1}{4n_2}\left(\frac{1}{\overline{Y}_{(2)1}} + \frac{1}{\overline{Y}_{(2)2}} + \frac{2\hat{\sigma}_{2(12)}}{\overline{Y}_{(2)1}\overline{Y}_{(2)2}}\right)$$

and when there is over-dispersion effect, the variance is:

$$\hat{V}(\hat{\boldsymbol{\theta}}) = \frac{1}{4n_1}\left(\frac{\phi}{\overline{Y}_{(1)1}} + \frac{\phi}{\overline{Y}_{(1)2}} + \frac{2\hat{\sigma}_{1(12)}}{\overline{Y}_{(1)1}\overline{Y}_{(1)2}}\right) + \frac{1}{4n_2}\left(\frac{\phi}{\overline{Y}_{(2)1}} + \frac{\phi}{\overline{Y}_{(2)2}} + \frac{2\hat{\sigma}_{2(12)}}{\overline{Y}_{(2)1}\overline{Y}_{(2)2}}\right)$$

where ϕ is the dispersion parameter of the response variable, its estimator is given by:

$$\hat{\phi} = \frac{1}{4}\left(\frac{\widehat{Var}(\overline{Y}_{(1)1})}{\overline{Y}_{(1)1}} + \frac{\widehat{Var}(\overline{Y}_{(1)2})}{\overline{Y}_{(1)2}} + \frac{\widehat{Var}(\overline{Y}_{(2)1})}{\overline{Y}_{(2)1}} + \frac{\widehat{Var}(\overline{Y}_{(2)2})}{\overline{Y}_{(2)2}}\right)$$

The estimators for each covariance are presented in equations (2-18) and (2-19), respectively,

$$\hat{\sigma}_{1(12)} = \frac{1}{n_1 - 2}\sum_{i=1}^{n_1}(y_{i1} - \overline{Y}_{(1)1})(y_{i2} - \overline{Y}_{(2)2}) \tag{2-18}$$

$$\hat{\sigma}_{2(12)} = \frac{1}{n_2 - 2}\sum_{i=n_1+1}^{n}(y_{i1} - \overline{Y}_{(1)2})(y_{i2} - \overline{Y}_{(2)1}) \tag{2-19}$$

From Equation (2-15), an estimator of the treatment effect is given by:

$$\hat{\tau} = \frac{1}{2}\ln\left(\frac{\overline{Y}_{(1)1}}{\overline{Y}_{(2)1}}\right) \tag{2-20}$$

and using the same reasoning for the carry-over effect, the variance of the estimator is:

$$\hat{V}(\hat{\tau}) = \frac{1}{4n_1\overline{Y}_{(1)1}} + \frac{1}{4n_2\overline{Y}_{(2)1}}$$

This leads to estimate the effect of treatments only with the first period, similar to what occurs in a normal crossover design (Jones and Kenward, 2015), but the information of the second period remains lost. An estimator of period effect is based on (2-15), and it is given by:

$$\hat{\pi} = \frac{1}{4}\ln\left(\frac{\overline{Y}_{(1)2}\overline{Y}_{(2)2}}{\overline{Y}_{(1)1}\overline{Y}_{(2)1}}\right) \tag{2-21}$$

and its variance:

$$\hat{V}(\hat{\pi}) = \frac{1}{16n_1}\left(\frac{1}{\overline{Y}_{(1)1}} + \frac{1}{\overline{Y}_{(1)2}} - \frac{2\hat{\sigma}_{1(12)}}{\overline{Y}_{(1)1}\overline{Y}_{(1)2}}\right) + \frac{1}{16n_2}\left(\frac{1}{\overline{Y}_{(2)1}} + \frac{1}{\overline{Y}_{(2)2}} - \frac{2\hat{\sigma}_{2(12)}}{\overline{Y}_{(2)1}\overline{Y}_{(2)2}}\right)$$

When a common crossover designs with few sequences and few periods are analyzed, the above methodology is adequate because both the parameters and their variances are easily estimated. However, it is observed that as the complexity of the design increases, the estimators become more difficult to handle and the previous methodology does not work well. Therefore, in the next section, a methodology that can be applied to more complex designs is built. It will even allow the integration of covariates.

2.2.2. Analysis using generalized estimation equations (GEE)

In this section, the crossover design has a response with Poisson distribution. This means that the answer belongs to the exponential family (Jorgensen, 1997), so, generalized estimation equations (GEE) are used (Davis, 2002). Let μ_{ij} be the expected value of the response of i-th experimental unit in the j-th period, the linear predictor used is determined by the following equation:

$$\boldsymbol{x}_{ij}^t\boldsymbol{\beta} = \boldsymbol{x}_{ij(\tau)}^t\boldsymbol{\beta}_\tau + \boldsymbol{x}_{ij(\pi)}^t\boldsymbol{\beta}_\pi + \boldsymbol{x}_{ij(\theta_1)}^t\boldsymbol{\beta}_{\theta_1} + \boldsymbol{x}_{ij(\theta_2)}^t\boldsymbol{\beta}_{\theta_2} + \cdots + \boldsymbol{x}_{ij(\theta_{T-1})}^t\boldsymbol{\beta}_{\theta_{T-1}} \tag{2-22}$$

$$\ln(\mu_{ij}) = \boldsymbol{x}_{ij}^t\boldsymbol{\beta}$$

where $\boldsymbol{x}_{ij(\tau)}$ is the vector that indicates the treatments applied to i-th experimental unit in the j-th period , $\boldsymbol{\beta}_\tau$ is the parameter vector of treatments' effects, $\boldsymbol{x}_{ij(\pi)}$ is the vector associated with the periods, $\boldsymbol{\beta}_\pi$ is the parameter vector of effects of the periods, $\boldsymbol{x}_{ij(\theta_m)}$ is the vector associated to the carry-over effect of order m and $\boldsymbol{\beta}_{\theta_m}$ is the m-th vector of carry-over effect, $m = 1, \ldots, P-1$, and $g(\mu_{ij}) = \ln(\mu_{ij}) = \boldsymbol{x}_{ij}^t\boldsymbol{\beta}$.

To illustrate the model given in equation (2-22), the model adapted for a cell mean to a

Sequence	Period 1	Period 2	Period 3
ABA	$\exp(\mu+\pi_1+\tau_1)$	$\exp(\mu+\pi_2+\tau_2+\theta_1)$	$\exp(\mu+\pi_3+\tau_1+\theta_2)$
BAB	$\exp(\mu+\pi_1+\tau_2)$	$\exp(\mu+\pi_2+\tau_1+\theta_2)$	$\exp(\mu+\pi_3+\tau_2+\theta_1)$

Table 2-3.: Fixed model effects of extra-period crossover design

crossover design with an extra period is used. This consists of two sequences of three periods (ABA and BAB) and is shown in Table **2-3**, τ_1 and τ_2 are the effects from treatment A and B, respectively; π_1, π_2 and π_3 are the effects of the periods 1, 2 and 3, respectively; θ_1 and θ_2 are the carry-over of first order. In addition to these effects, covariates measured in each experimental unit, such as age, gender, among others, can be included as covariates.

Generalized estimation equations

The estimation of the vector of parameters $\boldsymbol{\beta}$ is obtained by solving the system of p equations ($p = dim(\boldsymbol{\beta})$) given by (2-23), which is built with the score functions similar to the generalized linear models (Hardin and Hilbe, 2003).

$$\Psi_1(\boldsymbol{\beta}) = \left[\left\{ \sum_{i=1}^{n} \boldsymbol{x}_{li}^t \boldsymbol{D} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \eta} \right) [\boldsymbol{V}(\boldsymbol{\mu}_i)]^{-1} \left(\frac{\boldsymbol{y}_i - \boldsymbol{\mu}_i}{a(\phi)} \right) \right\}_{l=1,\ldots,p} \right]_{p\times 1} \tag{2-23}$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iT})^t$, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iP})^t$, $\boldsymbol{x}_{li}$ the l-th column of the design matrix $\boldsymbol{X}$ and the variance component $\boldsymbol{V}(\boldsymbol{\mu}_i)$, which is defined as:

$$\boldsymbol{V}(\boldsymbol{\mu}_i) = \left[\boldsymbol{D}(V(\mu_{ij}))^{\frac{1}{2}} \boldsymbol{R}(\boldsymbol{\alpha}) \boldsymbol{D}(V(\mu_{ij}))^{\frac{1}{2}} \right]_{T\times T} \tag{2-24}$$

where $\boldsymbol{D}(\cdot)$ is a diagonal matrix, $V(\mu_{ij})$ is the variance function from exponential family, and $R(\boldsymbol{\alpha})$ is a correlation matrix among observations on the same individual of dimension $T \times T$, this matrix has a total of $T + \binom{T}{2}$ correlation parameters to estimate. However, to achieve parsimony and to make the model more suitable, some defined correlation structures can be used (Davis, 2002). As the generator vector of the correlation matrix $\boldsymbol{\alpha}$ is added, estimation equation building given by Hardin and Hilbe (2003) is used:

$$\Psi_2(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \left(\frac{\partial \boldsymbol{\varepsilon}_i}{\partial \boldsymbol{\alpha}} \right)^t \boldsymbol{H}_i^{-1} (\boldsymbol{W}_i - \boldsymbol{\varepsilon}_i) \tag{2-25}$$

where $\boldsymbol{H}_i = \boldsymbol{D}(V(r_{ij}))_{q\times q}$ is a diagonal matrix, $\boldsymbol{\varepsilon}_i = E(\boldsymbol{W}_i)_{q\times 1}$ and

$$\boldsymbol{W}_i = (r_{i1}r_{i2}, r_{i1}r_{i3}, \ldots, r_{i(P-1)}r_{iP})_{q\times 1}^t$$

with r_{ij} as the ij-th Pearson residual, and $q = \binom{P}{2}$. By using the properties of the Poisson distribution, it is determined that (2-23) is equal to:

$$\Psi_1(\boldsymbol{\beta}) = \left[\left\{ \sum_{i=1}^{n} \boldsymbol{x}_{li}^t \boldsymbol{D}(\boldsymbol{\mu}_i) [\boldsymbol{V}(\boldsymbol{\mu}_i)]^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) \right\}_{l=1,\ldots,p} \right]_{p\times 1} \tag{2-26}$$

where $\boldsymbol{V}(\boldsymbol{\mu}_i) = \left[\boldsymbol{D}(\mu_i)^{\frac{1}{2}} \boldsymbol{R}(\boldsymbol{\alpha}) \boldsymbol{D}(\mu_i)^{\frac{1}{2}} \right]_{T\times T}$, and the vector of correlation parameters $\boldsymbol{\alpha}$ is estimated using the equation (2-25).

Estimation

The equations (2-25) and (2-26) are equated to zero, resulting the following system:

$$\Psi(\boldsymbol{\beta}, \boldsymbol{\alpha}) = (\Psi_1^t(\boldsymbol{\beta}), \Psi_2^t(\boldsymbol{\alpha}))^t \tag{2-27}$$

The solution of the systems of equations (2-27) is performed by applying the Fisher Scoring algorithm (Hardin and Hilbe, 2003). The $(m+1)$-th step is defined as:

$$\boldsymbol{\Theta}^{(m+1)} = \boldsymbol{\Theta}^{(m)} + (\boldsymbol{\mathfrak{I}}^{(m)})^{-1}\boldsymbol{U}(\boldsymbol{\Theta}^{(m)}) \tag{2-28}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\beta}^t, \boldsymbol{\alpha}^t)^t$, $\boldsymbol{\mathfrak{I}}$ is the Fisher information matrix for $\boldsymbol{\Theta}$ and $\boldsymbol{U} = \Psi(\boldsymbol{\Theta}) = \Psi(\boldsymbol{\beta}, \boldsymbol{\alpha})$. For more details, see Hardin and Hilbe (2003) and Liang and Zeger (1986).

Residuals

The residuals in a GEE model can be defined in three main ways:

1. **Ordinary residuals**: these are defined as $y_{ij} - \hat{\mu}_{ij}$, where $\hat{\mu}_{ij} = g^{-1}(\boldsymbol{x}_{ij}^t\hat{\boldsymbol{\beta}})$.

2. **Pearson residuals**: these are identified as:

$$r_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{\widehat{Var}(\hat{\mu}_{ij})}}. \tag{2-29}$$

3. **Deviance residuals**: for the Poisson distribution, deviance residuals are defined as:

$$r_{ij} = \begin{cases} \sqrt{2\hat{\mu}_{ij}} & \text{if } y_{ij} = 0 \\ \text{sign}(y_{ij} - \hat{\mu}_{ij})\sqrt{2\left(y_{ij}\ln\left(\frac{y_{ij}}{\hat{\mu}_{ij}}\right) - (y_{ij} - \hat{\mu}_{ij})\right)} & \text{if } y_{ij} > 0 \end{cases} \tag{2-30}$$

These residuals detect possible problems in GEE adaptation. In Hardin and Hilbe (2003), there is a clear discussion on residual analysis and the technique to verify the model fit through the Quasi Information Criterion (QIC), defined by:

$$\text{QIC} = -2\sum_{i=1}^{n}\sum_{j=1}^{P} Q(g^{-1}(\boldsymbol{x}_{ij}^t\hat{\boldsymbol{\beta}})) + 2\text{tr}(\boldsymbol{A}_I^{-1}\mathbf{V}_{MS}) \tag{2-31}$$

where $Q(\cdot)$ is the quasi-likelihood function for the Poisson distribution, $\mathbf{V}_{MS}$ is the sandwich variance for the model and $\boldsymbol{A}_I^{-1}$ is the Fisher information matrix for the independent model ($R(\boldsymbol{\alpha}) = I_T$ in the Equation (2-24)). The QIC has a similar interpretation to the Akaike information criterion (AIC) (Hardin and Hilbe, 2003).

Deviation analysis

Theorem 2.2.1. *Let $\hat{\boldsymbol{\beta}}$ be the solution of the GEE defined in equation (2-27) and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^t, \boldsymbol{\beta}_2^t)^t$. If the null hypothesis $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ is considered, the test statistic T_w defined in*

equation (2-32) *under the assumption of* H_0 *certainty and under some regularity conditions, follows an asymptotic distribution* $W \sim F_w$ *with* F_w *defined in equation* (2-33),

$$T_w = N(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10})^t V_{\hat{\boldsymbol{\beta}}_1}^{-1}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}) \tag{2-32}$$

$$P(W \leq w) = F_W(w) = \left(\prod_{i=1}^{p} b_i\right) \sum_{j=0}^{\infty} \left(a_j \int_0^w \frac{y^{\frac{1}{2}+j-1} e^{-\frac{y}{2c_1}}}{(2c_1)^{\frac{1}{2}+j}\Gamma(\frac{1}{2}+j)} dy \right) \tag{2-33}$$

where $V_{\hat{\boldsymbol{\beta}}_1} = \widehat{Var}(\hat{\boldsymbol{\beta}}_1)$, $b_i = \sqrt{\frac{c_1}{c_i}}$, $a_j = A_j^{(p)}$ $(j = 0, 1, \ldots)$, $A_j^{(i)} = \sum_{k=0}^{j} A_k^{(i-1)} A(c_i, j-k)$ $(i = 2, 3, \ldots, p)$, $A_r^{(2)} = A(c_2, r)$, $A(c_i, r) = \left[\prod_{k=0}^{r-1}\left(\frac{1}{2}+k\right)\left(1-\frac{c_1}{c_r}\right)^r\right](r!)^{-1}$ $(r = 1, 2, \ldots)$, *with* c_i *as the* i*-th eigenvalue of the matrix* $\boldsymbol{Q} = \boldsymbol{Q}_0^{-1}\boldsymbol{Q}_1$. $\boldsymbol{Q}_0$ *and* $\boldsymbol{Q}_1$ *are, respectively,*

$$\boldsymbol{Q}_0 = \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{D}}_i \boldsymbol{V}(\boldsymbol{\mu}_i)^{-1} \boldsymbol{D}_i \tag{2-34}$$

$$\boldsymbol{Q}_1 = \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{D}}_i \boldsymbol{V}(\boldsymbol{\mu}_i)^{-1} Cov(\boldsymbol{Y}_i) \boldsymbol{V}(\boldsymbol{\mu}_i)^{-1} \boldsymbol{D}_i \tag{2-35}$$

where $\boldsymbol{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} = \boldsymbol{D}\left(\frac{\partial \boldsymbol{\mu}_i}{\partial \eta}\right) \boldsymbol{X}_i = \left((\boldsymbol{D}_i^{(1)})_{T\times r}, (\boldsymbol{D}_i^{(2)})_{T\times(p-r)}\right)_{T\times p}$,
$\tilde{\boldsymbol{D}}_i = \tilde{\boldsymbol{D}}_i^{(1)} - \boldsymbol{D}_i^{(2)}\left(\sum_{i=1}^{n} \boldsymbol{D}_i^{(2)t} \boldsymbol{V}(\boldsymbol{\mu}_i)^{-1} \boldsymbol{D}_i^{(2)}\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{D}_i^{(2)t} \boldsymbol{V}(\boldsymbol{\mu}_i)^{-1} \boldsymbol{D}_i^{(1)}\right)$, $r = dim(\boldsymbol{\beta}_1)$, $p = dim(\boldsymbol{\beta})$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1k}, \ldots, \boldsymbol{x}_{iTk})^t$, *and*

$$\boldsymbol{Cov}(\boldsymbol{Y}_i) = \boldsymbol{D}\left(\boldsymbol{\mu}_i\right)^{\frac{1}{2}} \left(\frac{1}{n}\sum_{i=1}^{S}\sum_{i=1}^{n_i} \boldsymbol{D}\left(\boldsymbol{\mu}_i\right)^{-\frac{1}{2}} \boldsymbol{S}_i \boldsymbol{S}_i^t \boldsymbol{D}\left(\boldsymbol{\mu}_i\right)^{-\frac{1}{2}}\right) \boldsymbol{D}\left(\boldsymbol{\mu}_i\right)^{\frac{1}{2}} \tag{2-36}$$

Proof. Let $\hat{\boldsymbol{\beta}}$ be the fixed parameters estimator found by applying the generalized estimation equations, then by Theorem 2 of Liang and Zeger (1986) it is known that $\hat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}$, furthermore:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N(\mathbf{0}, Var(\hat{\boldsymbol{\beta}}))$$

If $\boldsymbol{\beta}$ and its estimator are partitioned as

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \qquad \text{and} \qquad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix},$$

respectively. The variance and covariance matrix of $\hat{\boldsymbol{\beta}}$ is partitioned into:

$$Var(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} Var(\hat{\boldsymbol{\beta}}_1) & Cov(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) \\ Cov(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)^t & Var(\hat{\boldsymbol{\beta}}_2) \end{pmatrix} \tag{2-37}$$

Besides, a robust and consistent estimator for the variance of $\hat{\boldsymbol{\beta}}$ (Pan, 2001b) is given by:

$$\widehat{Var}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{n} \boldsymbol{D}_i^t \boldsymbol{V}(\boldsymbol{\mu}_i)^{-1} \boldsymbol{D}_i\right)^{-1} \left(\sum_{k=1}^{n} \boldsymbol{D}_i^t \boldsymbol{V}(\boldsymbol{\mu}_i)^{-1} \widehat{Cov}(\boldsymbol{y}_i) \boldsymbol{V}(\boldsymbol{\mu}_i)^{-1} \boldsymbol{D}_i\right) \left(\sum_{k=1}^{n} \boldsymbol{D}_i^t \boldsymbol{V}(\boldsymbol{\mu}_i)^{-1} \boldsymbol{D}_i\right) \tag{2-38}$$

where $\widehat{Cov}(\boldsymbol{y}_i)$ was defined in equation (2-36). Also, using the consistent estimator of the variance given in the equation (2-38), the variance partition given in (2-37) and the Result 1 of Rotnitzky and Jewell (1990), it is obtained that the T_w statistic:

$$T_w = n(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10})^t V_{\hat{\boldsymbol{\beta}}_1}^{-1} (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}) \tag{2-39}$$

$$T_w \overset{n\to\infty}{\sim} \sum_{j=1}^{p} c_j \chi_j^2 \tag{2-40}$$

where χ_j^2 are independent random variables that follow a chi-square distribution with one-freedom degree, and the coefficients, c_j, are the eigenvalues of the matrix $\boldsymbol{Q}_0^{-1}\boldsymbol{Q}_1$ defined in equations (2-34) and (2-35), respectively. Then, the distribution of $W = \sum_{j=1}^{p} c_j \chi_j^2$ is found based on Moschopoulos and Canada (1984), as it is known that the moment-generating function of the linear combination is:

$$M_W(t) = E\left(e^{t\sum_{j=1}^{p} c_j \chi_j^2}\right) = \prod_{i=1}^{p} (1 - 2c_j t)^{-\frac{1}{2}} \tag{2-41}$$

In Mathai (1982), it is shown that $\left|\frac{1-\frac{c_1}{c_i}}{1-2c_1 t}\right| < 1$, and the following property:

$$1 - 2c_i t = (1 - 2c_1 t)\left(\frac{c_i}{c_1}\right)\left(1 - \frac{1 - \frac{c_1}{c_i}}{1 - 2c_1 t}\right) \qquad i = 2, \ldots, p$$

By using a series of powers, from the equation (2-41), it is obtained:

$$(1 - 2c_i t)^{-\frac{1}{2}} = \left(\frac{c_1}{c_i}\right)^{-\frac{1}{2}} \left(\sqrt{\frac{1}{1 - 2c_1 t}} + \sum_{r=1}^{\infty} \frac{\prod_{k=0}^{r-1}\left(\frac{1}{2} + k\right)}{r!} \left(1 - \frac{c_1}{c_r}\right)^r (1 - 2c_1 t)^{-(r+\frac{1}{2})}\right)$$

with convergence radius given by $t < \text{mín}(\frac{1}{2c_i})$. Thus, the following quantities are defined:

$$b_i = \left(\frac{c_1}{c_i}\right)^{\frac{1}{2}} \qquad \text{and} \qquad A(c_i, r) = \prod_{k=0}^{r-1} \left(\frac{1}{2} + k\right) \frac{\left(1 - \frac{c_1}{c_i}\right)^r}{r!}$$

Therefore, from the equation (2-41) and the previous series, it is obtained:

$$M_W(t) = \left(\sum_{j=0}^{p} b_i\right) \sum_{j=0}^{\infty} a_j (1 - 2c_1 t)^{-\left(\frac{2p+j}{2}\right)} \tag{2-42}$$

The coefficients a_j satisfy the recursion given in equation (2-43):

$$\prod_{i=2}^{p} \left[\sum_{i=0}^{\infty} A(c_i, r) x^{-r}\right] = \sum_{j=0}^{\infty} a_j x^{-j} \tag{2-43}$$

From equation (2-43), the following property is obtained:

$$a_j = A_j^{(p)}, \qquad A_j^{(i)} = \sum_{k=0}^{\infty} A_k^{(i-1)} A(c_i, j-k)$$

with $i = 2, \ldots, p$, $j = 0, 1, \ldots$ and $r = 0, 1, \ldots$ and $A_r^{(2)} = A(c_2, r)$. The term $(1-2c_1t)^{-(\frac{2p+j}{2})}$ in equation (2-42) is the moment-generating function of a random variable with gamma distribution, that is:

$$g_j(y) = \frac{y^{\frac{p}{2}+j-1} e^{-\frac{y}{2c_1}}}{(2c_1)^{\frac{p}{2}+j} \Gamma(\frac{p}{2}+j)}$$

Therefore, by properties of the moment generator function, the W distribution function is:

$$\begin{aligned} F_W(w) = P(W \leq w) &= \left(\prod_{i=1}^{p} b_i\right) \sum_{j=0}^{\infty} \left(a_j \int_0^w g_j(y) dy\right) \\ &= \left(\prod_{i=1}^{p} b_i\right) \sum_{j=0}^{\infty} \left(a_j \int_0^w \frac{y^{\frac{1}{2}+j-1} e^{-\frac{y}{2c_1}}}{(2c_1)^{\frac{1}{2}+j} \Gamma(\frac{1}{2}+j)} dy\right) \end{aligned} \tag{2-44}$$

□

With the previous theorem, it is possible to build a deviation analysis table of the Poisson crossover design, with sources of variation and their significance on the response. For example, for sequence effects, the hypothesis is defined as follows:

$$H_0 : \boldsymbol{\beta}_s = \mathbf{0} \qquad \text{vs} \qquad H_1 : \boldsymbol{\beta}_s \neq 0 \tag{2-45}$$

and the statistics test for this hypothesis is given by

$$T_{ws} = n \left(\hat{\boldsymbol{\beta}}_s^{\,t} V_{\hat{\boldsymbol{\beta}}_s}^{-1} \hat{\boldsymbol{\beta}}_s\right) \tag{2-46}$$

Based on Theorem 2.2.1, T_{ws} follows a distribution as defined in equation (2-33). When null hypothesis is true, it implies that the values of T_{ws} are close to zero, it is possible to construct a p-value for the test defined in (2-45), as follows:

$$p\text{-valor}_{H_{0s}} = P(W > T_{ws}) = 1 - \left(\prod_{i=1}^{p} b_i\right) \sum_{j=0}^{\infty} \left(a_j \int_0^{T_{ws}} \frac{y^{\frac{1}{2}+j-1} e^{-\frac{y}{2c_1}}}{(2c_1)^{\frac{1}{2}+j}\Gamma(\frac{1}{2}+j)} dy \right) \tag{2-47}$$

There are several algorithms to calculate the p-value given in equation (2-47), these are those described by Imhof (1961), Davies (1980), Farebrother (1984) and Liu et al. (2009). By using the CompQuadForm package from R project, built by Duchesne and Micheaux (2010) for the R Core Team (2017), it is possible to calculate the p-values with the algorithms presented above. In the same way, it is possible to build statistics test for other hypotheses sets, and to build, with these specifications, a deviation analysis table for all the effects present in the crossover design. Table **2-4** shows the deviations analysis, similar to an ANOVA table.

Table 2-4.: Deviation analysis

Source	T_w-value	p-value
Sequences	$n\left(\hat{\boldsymbol{\beta}}_s^{\,t} V_{\hat{\boldsymbol{\beta}}_s}^{-1} \hat{\boldsymbol{\beta}}_s\right)$	$P(W > T_{ws})$
Treatments	$n\left(\hat{\boldsymbol{\beta}}_\tau^{\,t} V_{\hat{\boldsymbol{\beta}}_\tau}^{-1} \hat{\boldsymbol{\beta}}_\tau\right)$	$P(W > T_{w\tau})$
Periods	$n\left(\hat{\boldsymbol{\beta}}_\pi^{\,t} V_{\hat{\boldsymbol{\beta}}_\pi}^{-1} \hat{\boldsymbol{\beta}}_\pi\right)$	$P(W > T_{w\pi})$
Order 1 carry-over	$n\left(\hat{\boldsymbol{\beta}}_{\theta_1}^{\,t} V_{\hat{\boldsymbol{\beta}}_{\theta_1}}^{-1} \hat{\boldsymbol{\beta}}_{\theta_1}\right)$	$P(W > T_{w\theta_1})$
⋮	⋮	⋮
Order $T-1$ carry-over	$n\left(\hat{\boldsymbol{\beta}}_{\theta_{T-1}}^{t} V_{\hat{\boldsymbol{\beta}}_{\theta_{T-1}}}^{-1} \hat{\boldsymbol{\beta}}_{\theta_{T-1}}\right)$	$P(W > T_{w\theta_{T-1}})$

The construction of GEE for a crossover design with Poisson response provides a versatile and useful tool for analysis. In addition to a deviation analysis table, there are estimators of the design parameters and their variances. This table lets make inferences about the variation causes of the design and works in the same way as an ANOVA table.

2.3. Simulation study

Two simulation studies were carried out: One to evaluate the sample sizes necessary to achieve convergence in Results 1 and Results 2, and another to evaluate the convergence and performance of the result of Theorem 2.2.1.

2.3.1. Simulation for Results 1 and 2

A simulation study with a 2×2 crossover design is proposed to evaluate the minimum sample sizes so that the estimate in Result 1 is close to the normal distribution. With this simulation it is also evaluated the coverage of the intervals constructed by means of the Result 2. Number of replications from 2 to 100 were evaluated; that is, the number of experimental

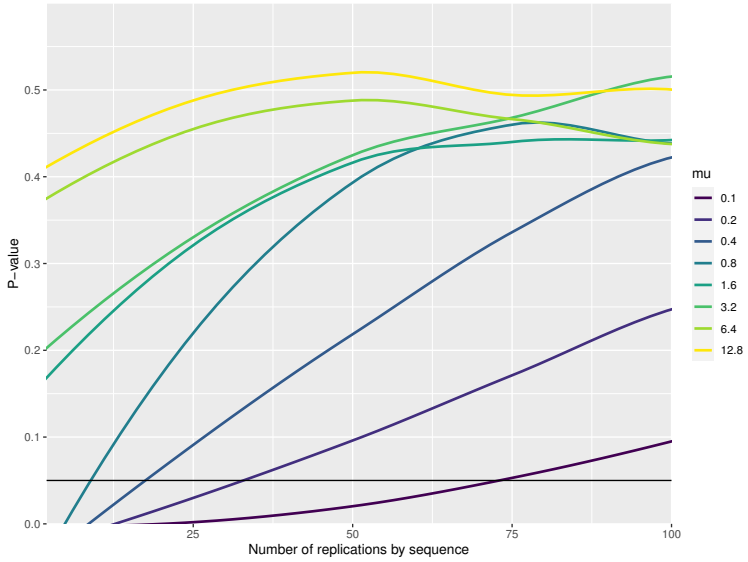

Figure 2-1.: P_value of Shapiro-Wilk normality test for the mean $\overline{\boldsymbol{Y}}$ in a crossover Poisson design with true values of μ_{ij} in $(0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4, 12.8)$. The lines represent the mean value of the 1000 simulations

units that receive each treatment sequence. It is further assumed that each sequence has the same number of replicates. Different values of μ_{ij} were also evaluated and in each one 1000 simulations of the mean are performed. After this, the Shapiro-Wilk test is performed and the p-value is obtained, which is shown in Figure **2-1**.

It is observed that the larger value of μ_{ij}, the fewer replicates are necessary to achieve a close approximation to normal. For values greater than 0.5, less than 20 replications are necessary. Special care must be taken if the response variable has too many zeros, because the approximation needs too many replications per sequence.

For the treatment effect, its value is $\tau = -1.0$ with a number of replications equal to 2 and the other parameters are obtained as a random number from a standard normal distribution and a crossover design is simulated, with which the 95 % confidence interval is obtained for the value of $\tau = -1.0$ and it is observed whether this value is contained or not. This process

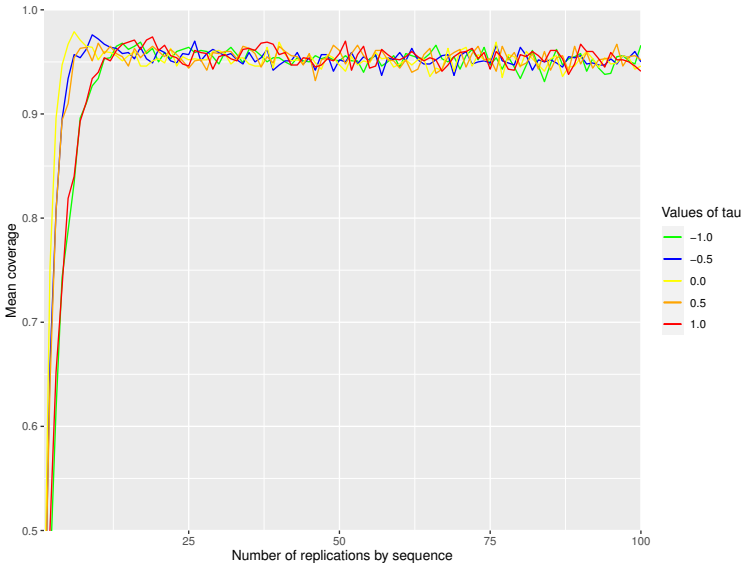

Figure 2-2.: Percentage coverage of 1000 confidence intervals obtained using the results of Theorem 2 for different values of the treatment effect

is repeated 1000 times and the coverage percentage is obtained. This procedure is performed for a number of replications between 2 and 100 and values of τ in $(-1.0, -0.5, 0.0, 0.5, 1.0)$. The results can be seen in Figure **2-2**.

The same simulation exercise is performed for the carry-over effect and the period effect, and the results are shown in figures **2-3** and **2-4**. In the simulation exercises, it is observed that for the treatment carry-over effect, the coverage stabilizes after 2 replicates per sequence. For the period effect after 14 replicates per sequence and for the carry-over effect after 6 replicates, therefore, it is correct to state that at least 4 replicates are needed to guarantee that the methodology with only logarithms achieves a good estimation of the parameters.

2.3.2. Simulation for the Theorem 2.2.1

Comparison between GEE and normal models

A simulation study is conducted in order to compare the behaviour of the estimators and their usual analysis.
In this study data from 2 x 2 and switch back crossover designs with Poisson response and different scenarios with presence of carry-over effects and without them were simulated, using the PoisNor package (Amatya and Demirtas, 2017). A normal model with linear log link with and without carry-over effect is adjusted, also a model like that of section 2.1 and with GEE Poisson response like that of section 2.2 with carry-over effect, the true value

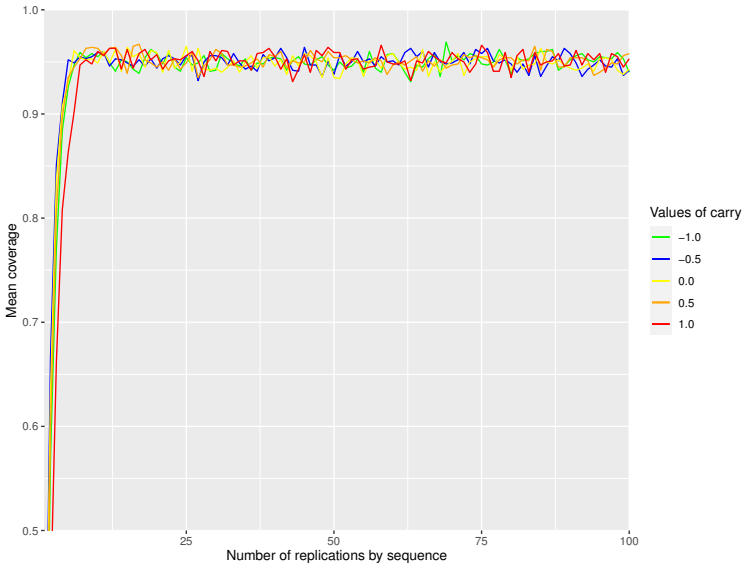

Figure 2-3.: Percentage coverage of 1000 confidence intervals obtained using the results of Theorem 2 for different values of the carry-over effect

Sequence	First Period	Second Period	Third Period
1	A	B	A
2	B	A	B

Table 2-5.: Swicth back crossover design

(TV) is presented, the mean parameter estimated by the model (PE), the standardized bias $SB = \left|\frac{E(\hat{\theta})-\theta}{SD(\hat{\theta})} \times 100\,\%\right|$, the root mean square error $RMSE = \sqrt{E(\hat{\theta}-\theta)^2}$ and the coverage of the 95 % confidence interval (CR) for the three models with 10000 simulated data in each case, the results for a 2×2 crossover design are presented in Table **2-6** and for switch back crossover (Its structure is presented in the Table **2-5**) design are presented in Table **2-7**, in this Table θ_1 is the first order carry-over and θ_2 is the second order carry-over.

If the coverages in Table **2-6** are observed, the three models offer a coverage close to 92 % when there are no carry-over effect. But when carry-over effects are present, the Normal model does not offer an unbiased estimate of the treatment effect, while the two models proposed here work well, since their coverage is close to 93 %. When the structure of the crossover design is more complex, in Table **2-7** it is observed that the estimation of the treatment effect has a coverage close to zero in the model with normal distribution, low with the model of logarithms of cell mean. It was adequate in the case of the GEE. The standardized error is very high for the estimation of the treatment effect in the models that assume a normal response without carry-over effect. In the Poisson (2.1) model when the

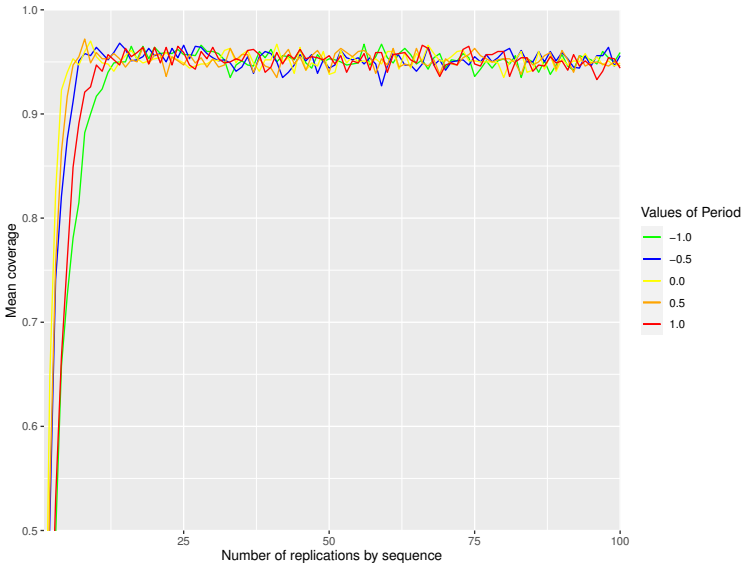

Figure 2-4.: Percentage coverage of 1000 confidence intervals obtained using the results of Theorem 2 for different values of the period effect

$\mu = 1.5,\ \tau = 0,\ \pi = 0,\ \theta = 0, n_1 = 10, n_2 = 10$

TV	Normal				Poisson 2.2.1				Poisson 2.2.2			
	PE	SB	RMSE	CR	PE	SB	RMSE	CR	PE	SB	RMSE	CR
1.50	1.49	11.6	0.08	92.8	1.49	15.1	0.08	94.2	1.49	15.1	0.08	92.2
0.00	0.00	0.62	0.08	92.0	0.01	1.05	0.11	93.9	-0.01	1.05	0.11	92.6
0.00	-0.01	0.05	0.08	91.9	-0.01	0.02	0.07	94.2	-0.01	0.01	0.08	91.8
0.00	NA	NA	NA	NA	0.02	0.78	0.15	94.7	0.02	0.72	0.15	92.4

$\mu = 1.5,\ \tau = 0.5,\ \pi = 0.8,\ \theta = -0.5, n_1 = 10, n_2 = 10$

TV	Normal				Poisson 2.2.1				Poisson 2.2.2			
	PE	SB	RMSE	CR	PE	SB	RMSE	CR	PE	SB	RMSE	CR
1.50	1.56	63.1	0.11	91.9	1.48	17.9	0.09	93.8	1.48	17.9	0.09	92.8
0.5	0.03	663.6	0.48	0.0	0.51	8.18	0.18	94.6	0.51	8.18	0.18	92.3
0.8	0.74	69.05	0.11	87.9	0.81	13.3	0.09	93.9	0.82	13.2	0.09	92.6
-0.5	NA	NA	NA	NA	-0.52	7.46	0.19	96.7	-0.51	7.46	0.19	92.3

$\mu = 1.5,\ \tau = 0.7,\ \pi = 0.6,\ \theta = 0.1, n_1 = 20, n_2 = 30$

TV	Normal				Poisson 2.2.1				Poisson 2.2.2			
	PE	SB	RMSE	CR	PE	SB	RMSE	CR	PE	SB	RMSE	CR
1.50	1.45	79.9	0.08	87.1	1.49	12.2	0.06	94.9	1.49	12.2	0.06	93.8
0.7	0.79	185.6	0.11	52.3	0.70	6.22	0.11	94.2	0.71	6.22	0.11	93.9
0.6	0.65	102.2	0.07	81.4	0.60	6.45	0.06	93.9	0.60	6.45	0.63	93.8
0.1	NA	NA	NA	NA	0.09	3.98	0.124	94.1	0.10	3.98	0.12	93.7

Table 2-6.: Simulation study for 2×2 crossover design

$\mu = 1.5$, $\tau = 0.5$, $\pi_1 = 0.3$, $\pi_2 = 0.4$, $\theta_1 = 0$, $\theta_2 = 0$, $n_1 = 10, n_2 = 10$

TV	Normal				Poisson 2.2.1				Poisson 2.2.2			
	PE	SB	RMSE	CR	PE	SB	RMSE	CR	PE	SB	RMSE	CR
1.50	1.49	13.8	0.07	91.4	1.48	19.3	0.07	96.2	1.49	19.3	0.07	92.5
0.50	0.490	10.5	0.07	92.4	0.51	6.56	0.21	93.9	0.51	9.23	0.13	90.7
0.30	0.30	3.53	0.09	91.3	0.31	6.39	0.11	96.7	0.31	6.39	0.10	91.1
0.40	0.41	1.68	0.08	92.3	0.40	4.00	0.09	97.4	0.40	4.00	0.09	92.7
0.00	NA	NA	NA	NA	0.01	3.44	0.15	95.2	-0.01	3.51	0.17	90.5
0.00	NA	NA	NA	NA	0.00	0.47	0.15	96.5	0.00	0.47	0.16	91.5

$\mu = 1.5$, $\tau = 0.5$, $\pi_1 = 0.3$, $\pi_2 = 0.4$, $\theta_1 = 0.3$, $\theta_2 = 0$, $n_1 = 10, n_2 = 10$

TV	Normal				Poisson 2.2.1				Poisson 2.2.2			
	PE	SB	RMSE	CR	PE	SB	RMSE	CR	PE	SB	RMSE	CR
1.50	1.43	84.6	0.11	84.1	1.48	24.5	0.07	94.2	1.49	24.6	0.07	91.8
0.50	0.77	361	0.28	2.41	0.52	7.77	0.21	93.9	0.51	6.895	0.13	92.8
0.30	0.48	185	0.22	48.8	0.30	0.42	0.10	94.2	0.30	4.25	0.11	93.4
0.40	0.49	122	0.12	71.8	0.41	8.25	0.09	94.7	0.41	8.25	0.09	92.6
0.30	NA	NA	NA	NA	0.30	3.83	0.62	94.2	0.29	1.82	0.16	93.9
0.00	NA	NA	NA	NA	0.01	4.3	0.16	94.7	0.01	4.27	0.17	92.7

$\mu = 1.5$, $\tau = 0.5$, $\pi_1 = 0.3$, $\pi_2 = 0.4$, $\theta_1 = 0.3$, $\theta_2 = -0.2$, $n_1 = 10, n_2 = 10$

TV	Normal				Poisson 2.2.1				Poisson 2.2.2			
	PE	SB	RMSE	CR	PE	SB	RMSE	CR	PE	SB	RMSE	CR
1.50	1.44	70.3	0.09	89.5	1.48	21.9	0.08	94.2	1.48	21.9	0.08	91.6
0.50	0.83	434.6	0.33	0.3	0.71	90.7	0.31	43.9	0.52	13.7	0.14	93.8
0.30	0.548	254	0.26	22.4	0.31	5.67	0.11	94.2	0.31	5.66	0.11	91.6
0.40	0.43	39.8	0.08	88.9	0.40	1.85	0.10	94.7	0.40	1.85	0.10	91.6
0.30	NA	NA	NA	NA	0.38	439	0.71	94.2	0.29	3.97	0.17	92.9
-0.20	NA	NA	NA	NA	-0.19	217	0.43	94.7	-0.19	4.98	0.18	91.2

Table 2-7.: Simulation study for switch back crossover design

carry-over effect is different from zero, the standardized bias is larger than in the GEE model for the treatment effect.

Asymptotic distribution and power of the proposed test

To test the power of the test statistic proposed in the theorem 2.2.1, the values of τ were tested for the treatment effect between -2 and 2. For each value of τ, 1000 simulations were run for a numbers of replications of five experimental units per sequence. Period and carry-over effects were randomly selected between -1 and 1 in each simulation in order to determine whether the test statistic is robust to the treatment effect. Then both, the test statistic T_w defined in equation (2-32) and the percentage of times the null hypothesis $H_0 : \tau = 0$ was rejected using the distribution defined in equation (2-34) were calculated. The previous exercise was repeated with numbers of replications of 10, 20, 50 and 100.

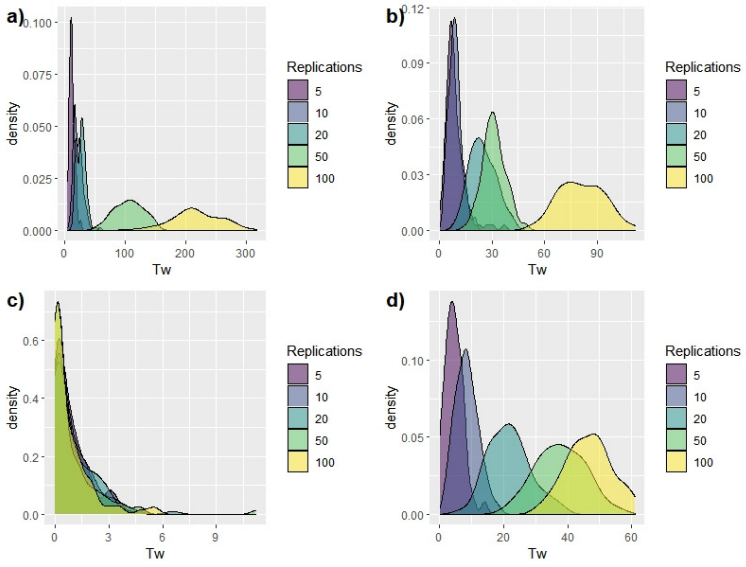

Figure 2-5.: Empirical distribution of the statistic T_w under the null hypothesis $H_0 : \tau = 0$ for different values of number of replications per sequence. The values of the treatment effect (τ) were: a) -2, b) -1, c) 0 and d) 1.

Figure **2-5** shows the distribution of the test statistic, T_w, for different values of τ, but keeping the null hypothesis as $H_0 : \tau = 0$. In the case that the true value of τ is zero, the distribution is very similar regardless of the number of replications, regardless of the number of replications per sequence. That is, if there is no treatment effect, H_0 will not be rejected with high confidence regardless of the number of replications. As soon as the value of τ moves away from zero, the test statistic starts to takes higher values, which are favored by the number of replications. That is, if there is a treatment effect, H_0 will be rejected more easily if each sequence has more replications.

The power results of the test are shown in Figure **2-6** and coincide with the discussion of the distribution of T_w. The power increases as the treatment effect moves away from 0 (the null hypothesis), and furthermore, its growth is faster with large values of the number of replications per sequence. The power takes values close to 0.05 when $\tau = 0$, which is the level of significance used in the hypothesis test regardless of the number of replications.

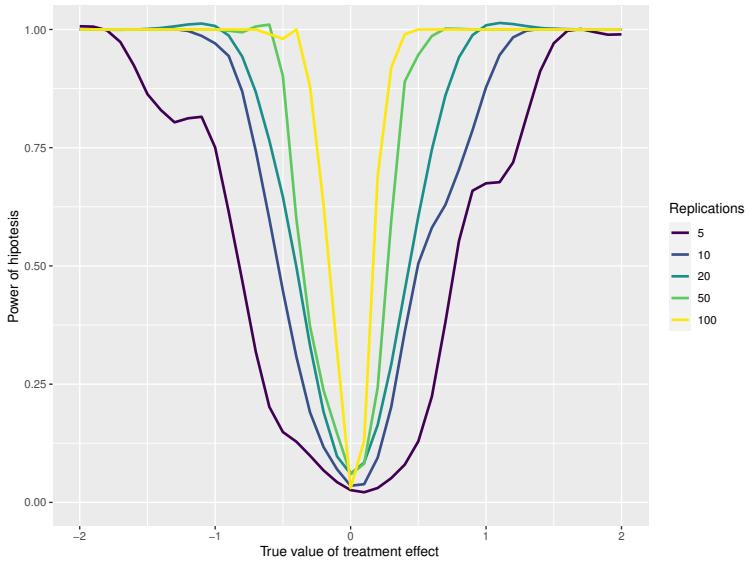

Figure 2-6.: Power of the test statistic for different values of the treatment effect τ under H_0; $\tau = 0$ and different number of replications per sequence.

2.4. Application

2.4.1. Analysis of the student data

Chard et al. (2019) conducted a pilot study to investigate the impact of providing supplementary drinking water on the cognitive performance of pupils of two school grades (5 and 6) in water-scarce schools in rural Mali. 47 students were assigned to take the control treatment (normal conditions) on the first day and receive the treatment (controlled hydration) on the second day. 60 received the treatments in reverse (Hydration the first day and control the second day). One part of this test assesses visual attention. Two nearly identical pictures were presented side-by-side. Students were given one minute to circle differences between the two images and the number of correct differences identified.
Since school grade can be taken as a covariate, three different models are fitted: 1) the one where only the effects of period, treatment and carry-over are taken into account, 2) GEE without the effect of school grade, and 3) GEE with effect of school grade. It is necessary to compare them by using the QIC, defined in equation (2-31). Table **2-8** shows the QIC values for the three previously fitted models.

Loglinear analysis

Table **2-9** shows the means for each sequence in each period, together to the variance (in parentheses). Note that in all cases, the variance is lower than the mean, i.e, under-dispersion

Model	QIC
Poisson (Section 2.2.1)	461.35
GEE (without school grade)	459.56
GEE (with school grade)	455.49

Table 2-8.: QIC for three models adjusted for the number of correct differences identified by student

Sequence	Period 1	Period 2
Placebo/Hydration	1.851 (1.782)	3.085 (1.601)
Hydration/Placebo	1.667 (1.141)	2.400 (2.210)

Table 2-9.: Means of correct differences identified by student (variances in parenthesis)

with an estimated dispersion parameter of $\hat{\phi} = 0.7718$. The estimated correlation matrix is given by:

$$\widehat{Cor}(\overline{Y}) = \hat{\mathbf{R}} = \begin{pmatrix} 1.0000 & 0.3423 & 0.0000 & 0.0000 \\ 0.3423 & 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 & 0.5123 \\ 0.0000 & 0.0000 & 0.5123 & 1.0000 \end{pmatrix}$$

The analysis of data is carried out and shows in Table **2-10** to exemplify the method exposed in Section 2.2.1. The carry-over effect and period effect are significative, however, a direct effect of the treatment is not observed.

Hypothesis	Parameter	Estimate	Standar error	p-value
$H_0 : \theta = 0$	carry-over	0.1780	0.0946	0.0299
$H_0 : \tau = 0$	Treatment	-0.0524	0.0644	0.2077
$H_0 : \pi = 0$	Period	0.2189	0.0473	<0.001

Table 2-10.: Loglinear analysis

Analysis through GEE with school grade

By applying the methodology of Section 2.2.2, the analysis with GEE for the number of correct differences identified by student is carry out. The model proposed for this design has effect of periods, school grade, treatments and first-order carry-overs.

The given statistic, T_w, is calculated applying the equation (2-32) and the p-values of the associated hypothesis described in equation (2-47). Table **2-11** shows the deviation analysis

Source	T_w	p-value
Treatment	1.083	0.2980
Period	46.5353	<0.001
carry-over	4.8281	0.0279
Grade	3.5865	0.0003

Table 2-11.: Deviance analysis

for this data set, and the estimate of the effects vector is given by:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\tau} \\ \hat{\pi} \\ \hat{\theta} \\ \hat{SG_6} \end{pmatrix} = \begin{pmatrix} -0.136 \\ 0.4377 \\ 0.2091 \\ 0.3227 \end{pmatrix}$$

Note that the difference between the conclusions obtained in Table **2-10** and Table **2-11** is since the fact that the GEE model the correlation between individuals, generating a better adjustment to the data set because the QIC is lower.
In the GEE analysis, a larger significant effect of period is observed, which together with its estimated value gives evidence of a higher response in the second application period. The next effect is the grade effect, which implies a better performance of students in grade 6 over those in grade 5. The carry-over effect is significant and positive, implying that hydration leaves a positive effect on the placebo in the following period (i.e., the placebo leaves a negative effect on hydration in the following period). The treatment effect is not significant in the application period, but due to the presence of carry-over, it is significant over the entire design structure.

3. A correlation structure for the analysis of Gaussian and non-Gaussian responses in crossover experimental designs with repeated measures

The content of this chapter was published as a research article in the journal Statistical Papers (Cruz et al., 2023b). [1]

3.1. Introduction

Experimental designs are a very useful tool to analyze the effects of treatments applied to a set of experimental units (Hinkelmann and Kempthorne, 2005). In this kind of studies, it is frequent that experimental units are observed at a unique time point, this is known as a transversal experiment; notwithstanding, sometimes experimental units are observed several times during the study, keeping the same treatment, giving rise to longitudinal studies (Davis, 2002). There are situations in which an experimental unit receives all treatments, each one in a different period, this induces a setup where a sequence of treatments is applied to each unit. This kind of designs is known as crossover design (Jones and Kenward (2015), Jankar et al. (2020), Patterson (1951), Zhang et al. (2012)).
In the scope of crossover designs, published results focus on the case of a normally distributed (Jones and Kenward, 2015) or binary (two possible outputs, namely success or failure) response variable. The latter case has been treated by means of generalized linear models for binary data (Ratkowsky et al. (1992), Curtin (2017), Li et al. (2019)).
(Jones and Kenward, 2015, pg 204) described a crossover experiment with three treatments to control arterial pressure: treatment A is a placebo, treatments B and C are 20 and 40 mg doses of a test drug. Thus, there were six three-period sequences: ABC, ACB, BCA, BAC, CAB, and CBA, each one of them was applied to two individuals lasting six weeks each and the measurement was made in the middle of each period. Due to ethical reasons, there was

[1] https://link.springer.com/article/10.1007/s00362-022-01391-z

no washout period between the treatments. In each period, 10 consecutive measurements of diastolic arterial pressure were taken: 30 and 15 minutes before, and 15, 30, 45, 60, 75, 90, 120 and 240 minutes after the administration of the treatment, as shown in Table **3-1**.

Sequence		Period 1	Period 2	Period 3
(1)	Ind 1	10 measurements	10 measurements	10 measurements
	Ind 2	10 measurements	10 measurements	10 measurements
	⋮	⋮	⋮	⋮
(6)	Ind 11	10 measurements	10 measurements	10 measurements
	Ind 12	10 measurements	10 measurements	10 measurements

Table 3-1.: Structure of the blood pressure crossover design

A second experiment carried out by researchers of the Department of Animal Sciences of Universidad Nacional de Colombia, was focused on inferring the effects of two diets, A (grass) and B (a mixture of grass and paper recycling waste) on dairy cattle performance. Eight cows split into two groups of four received the diets in such a way that the first group was fed diet A from day 1 to day 42, and diet B from day 43 to day 84, while the second group was fed diet B during the first period and diet A during the second one. Measurements of milk protein yield were taken at 1, 14, 28, 42, 56, 70 and 84 days, further details on this experiment can be found in Jaime (2019); the design structure is shown in Table **3-2**.

Sequence		Period 1	Period 2
(1) AB	Ind 1	3 measurements	3 measurements
	⋮	⋮	⋮
	Ind 4	3 measurements	3 measurements
(1) BA	Ind 5	3 measurements	3 measurements
	⋮	⋮	⋮
	Ind 8	3 measurements	3 measurements

Table 3-2.: Structure of the crossover design in cows

A thrid experiment is described in (Jankar and Mandal, 2021), in the work environment experiment, there were a total of $n = 288$ participants. These participants were divided into four groups: $G1$, $G2$, $G3$ and $G4$ with each group having an equal number of (72) individual participants. Periods were named Wave1, Wave2, Wave3 and Wave4, where each Wave had a duration of 2 weeks. The four treatments involved in this experiment are office designs named as A (Activity-Based), B (Open Plan), C (Team Offices) and D (Zoned Open Plan). The structure of the crossover design is given in Table **3-3**.

In this experiment, sensors were used to detect desktop occupancy. These sensors contained infrared cameras and had functionality to establish a region of interest within the video frame. They produced a count of the number of people seen within the area of interest once every 5 minutes Pitchforth et al. (2020). This variable generated a repeated measurement within each period of the binomial type, and was grouped for each hour of work (from 09:00 to 17:00) for which, there were 9 measurements of the number of times that they were occupied the job site and those that they were not occupied by sensor in each period.

Sequence	Period 1	Period 2	Period 3	Period 4
G1	B	A	D	C
G2	C	D	A	B
G3	D	B	C	A
G4	A	C	B	D

Table 3-3.: Structure of the crossover design in work environment experiment

To analyze this sort of designs, Basu and Santra (2010), Josephy et al. (2015), Hao et al. (2015), Lui (2015), Rosenkranz (2015), Grayling et al. (2018), Madeyski and Kitchenham (2018) and Kitchenham et al. (2018) used mixed effects models for crossover designs with Gaussian response and a single observation per period including additive carry-over effects. In Biabani et al. (2018) were presented a review on crossover designs in neuroscience, all papers considered Gaussian responses and did not account for carry-over effects due to the presence of a washout period, even when it lasted a few days. On the other hand, Oh et al. (2003), Curtin (2017) and Li et al. (2018) used generalized linear models for crossover designs of two periods and two sequences of two treatments with a continuous (normal or gamma) or binary response variable and each experimental unit observed once per period. Generalized estimating equations (GEE) was used by Jankar et al. (2020) to estimate model parameters in crossover experiments. A Bayesian formulation of generalized linear models where the response was the survival time with a single observation per period was presented in Shkedy et al. (2005) and Liu and Li (2016).
Moreover, Dubois et al. (2011), Diaz et al. (2013) and Forbes et al. (2015) used Gaussian mixed models to analyze records from crossover designs with repeated measures using the area under curve as a strategy to obtain a single observation per period and one experimental unit, and they did not account for carry-over effects.
In all aforementioned approaches, it is assumed that the crossover design considers a washout period between treatments and that the response variable of each individual is observed once per period. These assumptions are not fulfilled in the experiments described above because of two reasons: i) in patients with arterial hypertension the treatment cannot be stopped, while in the case of dairy cattle, cows must be fed every day; moreover, in both studies, the

placebo is part of the treatment design, so considering it as a washout period would modify the experiment, ii) in each treatment, experimental units were observed several times per period. Therefore, there is a necessity for developing a consistent methodology to analyze this sort of experiments.
To analyze longitudinal data, generalized linear mixed models (GLMM) or GEE can be used. While the GLMM explicitly models within-subject correlation using random effects, the GEE implicitly explains such correlations using sandwich variance estimates Liang and Zeger (1986). Therefore, the GEE estimators yield a more solid inference about the fixed effects compared to the GLMM that depend on the distributional assumptions about the random effects Zhang et al. (2012). Since the main goal of a crossover experiment is to obtain a robust inference about treatment effects, GEE models are more suitable than GLMM models.
In this study, we develop a methodology to analyze data from crossover designs with repeated measures using GEE and considering two correlation structures, between and within periods, which are combined via a Kronecker product to yield the overall correlation matrix. These approach was found to improve the estimation of parameters of interest with respect to methods based on GEE that consider a single correlation structure for all the observations of a subject; in addition, this method accounts for carry-over effects, hence, it does not need a washout period. It is important to note that the washout period is relevant in crossover designs except in strongly balanced designs, because in these designs the treatment effect is not aliased with the carry-over effects Wang and Chinchilli (2021).

In the second section, we present some background and define the methodology. In the third section, we propose a method to estimate the correlation matrix, and provide theoretical results that support our estimation and modelling approaches. In the fourth section, we present a simulation study showing some advantages of our model as compared to the model with a single correlation structure for each period. Lastly, in the fifth section, the methodology is applied to real data from the aforementioned arterial pressure, occupation and dairy cattle experiments.

3.2. Crossover designs with repeated measures

A crossover design has the following components (Patterson, 1951): i) *Sequences* which are each one of the distinct treatment combinations to be sequentially and randomly applied to the experimental units, ii) *Treatments*, which are randomly applied to each experimental unit within each sequence, iii) *Periods*, the time intervals in which the treatments that make up a sequence are applied, usually, each period is same for all sequences and, consequently, the number of periods is equal to the number of elements of each sequence, iv) *Experimental*

unit, the subjects or elements that receive the treatments, the total number of experimental units in the study is denoted by n.

Another important feature of the crossover design is the existence of carry-over effects, defined in Vegas et al. (2016) as follows: persistency of the effect of a treatment over those that will be applied later, i.e., the treatment is applied in a given period, but its effect still affects the response in later periods when other treatments are applied, this residual effect is known as carry-over effect. When the effect persists after one period, it is known as first order carry-over effect, and when it persists after two periods, it is called a second order carry-over effect and so on (Patterson, 1951).

In a crossover design with S sequences of length (the number of elements comprising each sequence) P, Y_{ijk} is defined as the kth respose of the ith experimental unit at jth period. Let n_{ij} be the number of observations of the ith experimental unit during the jth period. Then, vector $\boldsymbol{Y}_{ij}$ is defined as:

$$\boldsymbol{Y}_{ij} = \left(Y_{ij1}, \ldots, Y_{ijn_{ij}}\right)^T \tag{3-1}$$

Also, vector $\boldsymbol{Y}_i$ is defined as:

$$\boldsymbol{Y}_i = \left(\boldsymbol{Y}_{i1}, \ldots, \boldsymbol{Y}_{iP}\right)^T \tag{3-2}$$

which has dimension $\sum_{j=1}^{P} n_{ij}$, where P is the number of periods.

According to these definitions, the ideas presented when discussing the arterial pressure and dairy cattle experiments, and assuming that Y_{ijk} has a distribution in the exponential family, we propose the following model based on GEE:

$$\begin{aligned} &E(Y_{ijk}) = \mu_{ijk}, \qquad i = 1, \ldots, n,\ j = 1, \ldots, P,\ k = 1, \ldots, n_{ij},\ L = \max_{ij}\{n_{ij}\} \\ &g(\mu_{ijk}) = \boldsymbol{x}_{ijk}^T\boldsymbol{\beta} = \mu + \gamma_j + \tau_{d[i,j]} + \theta_{d[i,j-1]} + \cdots + \theta_{d[i,j-P+1]} \end{aligned} \tag{3-3}$$

where $g(\cdot)$ is the link function related to the exponential family, $\boldsymbol{x}_{ijk}$ is the vector of the design matrix corresponding to kth response from the ith experimental unit at the jth period, $\boldsymbol{\beta}$ is the vectors of fixed effects, μ is the overall mean, α_i is the effect of the ith sequence, γ_j is the effect of the jth period, $\tau_{d[i,j]}$ is the effect of treatment d applied in the period j to ith experimental unit ($d = 1, \ldots, S$; where S is the number of treatments), $\theta_{(1)[i,j-1]}$ is the first order carry-over effect of d treatment, $\theta_{d[i,j-u]}$ is the carry over effect of order u of d treatment ($u = 1, \ldots, P$) and L is the maximum number of observations of an experimental unit in each period.

The carry-over effects are considered in the three experiments discussed above because there was not a washout period. Due to the fact that observations of the same experimental unit are correlated, parameter estimation is carried out using GEE (Liang and Zeger, 1986). To this end, the following system of $q = dim(\boldsymbol{\beta})$ equations has to be solved:

$$\boldsymbol{U}(\boldsymbol{\beta}) = \left[\left\{ \sum_{i=1}^{n_i} \boldsymbol{x}_{mi}^T \boldsymbol{D} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \eta} \right) [\boldsymbol{V}(\boldsymbol{\mu}_i)]^{-1} \left(\frac{\boldsymbol{y}_i - \boldsymbol{\mu}_i}{\phi} \right) \right\}_{m=1,\ldots,q} \right]_{q \times 1} \tag{3-4}$$

where $\boldsymbol{\mu}_i = (\mu_{i11}, \ldots, \mu_{iPn_{iP}})^T$, $\boldsymbol{y}_i = (y_{i11}, \ldots, y_{iPn_{iP}})^T$, $\boldsymbol{x}_{mi}$ is the mth column of the design matrix of the ith experimental unit, ϕ is the dispersion parameter, $n_i = \sum_{j=1}^{P} n_{ij}$ is the total number of observations of the ith experimental unit, and the covariance component $\boldsymbol{V}(\boldsymbol{\mu}_i)$ is defined as:

$$\boldsymbol{V}(\boldsymbol{\mu}_i) = \left[\boldsymbol{D}(V(\mu_{ijk})^{\frac{1}{2}}) \boldsymbol{R}(\boldsymbol{\alpha}) \boldsymbol{D}(V(\mu_{ijk})^{\frac{1}{2}}) \right]_{P \times P} \tag{3-5}$$

where $V(\mu_{ijk})$ is the variance function corresponding to the exponential family, $\boldsymbol{D}(\cdot)$ is a diagonal matrix, for example, in the case that Y_{ijk} is binary, $V(\mu_{ijk}) = \mu_{ijk}(1 - \mu_{ijk})$ and $\boldsymbol{D}(V(\mu_{ijk})$ would be of the form:

$$\boldsymbol{D}(V(\mu_{ijk})) = \begin{pmatrix} \mu_{i11}(1-\mu_{i11}) & 0 & \cdots & 0 \\ 0 & \mu_{i12}(1-\mu_{i12}) & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_{iPL}(1-\mu_{iPL}) \end{pmatrix}_{PL \times PL}$$

and $R(\boldsymbol{\alpha})$ is the correlation matrix related to the covariance matrix $\boldsymbol{\Sigma}_i = Var(\boldsymbol{Y}_i)$ as follows:

$$(\boldsymbol{\Sigma}_i)_{PL \times PL} = \left(\boldsymbol{D}(Var(Y_{ijk})^{\frac{1}{2}}) \boldsymbol{R}(\boldsymbol{\alpha}) \boldsymbol{D}(Var(Y_{ijk})^{\frac{1}{2}}) \right)_{PL \times PL}.$$

3.3. Kronecker correlation matrix

As in a crossover design, the measurement periods are homogeneous, which is plausible to assume that the correlation within periods is similar. In addition, it is assumed that the correlation between the measurements of one period and those of another period are proportional to the correlation within the periods. The advantages of the use of such correlation structure, which corresponds to a Kronecker product, over the unstructured covariance matrix have been reported by several authors (Boik (1991), Srivastava et al. (2008), Roy and Khattree (2005), Leorato and Mezzetti (2016) and Krzyśko and Skorzybut (2009)). So, we propose a correlation structure of the form:

$$\boldsymbol{R}(\boldsymbol{\alpha}) = \boldsymbol{\Psi} \otimes \boldsymbol{R}_1(\boldsymbol{\alpha}_1) \tag{3-6}$$

where $\boldsymbol{R}_1(\boldsymbol{\alpha}_1)$ is the within-period correlation matrix, $\boldsymbol{\Psi}$ is the between-period correlation matrix, and $\otimes$ represents the Kronecker product (Harville, 1997). For example, if we have

three periods ($P = 3$), four observations within each period ($L = 4$) and an exchangeable correlation matrix within period, a possible configuration of the arrays could be:

$$\boldsymbol{R}_1(\boldsymbol{\alpha}_1) = \begin{pmatrix} 1 & \alpha_1 & \alpha_1 & \alpha_1 \\ \alpha_1 & 1 & \alpha_1 & \alpha_1 \\ \alpha_1 & \alpha_1 & 1 & \alpha_1 \\ \alpha_1 & \alpha_1 & \alpha_1 & 1 \end{pmatrix}, \quad \boldsymbol{\Psi} = \begin{pmatrix} 1 & \psi_{12} & \psi_{13} \\ \psi_{12} & 1 & \psi_{23} \\ \psi_{13} & \psi_{23} & 1 \end{pmatrix}$$

with $|\psi_{jj'}| < 1$, then $\boldsymbol{R}(\boldsymbol{\alpha})$ in Equation (3-6) would be:

$$\boldsymbol{R}(\boldsymbol{\alpha}) = \begin{pmatrix} \boldsymbol{R}_1(\boldsymbol{\alpha}_1) & \psi_{12}\boldsymbol{R}_1(\boldsymbol{\alpha}_1) & \psi_{13}\boldsymbol{R}_1(\boldsymbol{\alpha}_1) \\ \psi_{12}\boldsymbol{R}_1(\boldsymbol{\alpha}_1) & \boldsymbol{R}_1(\boldsymbol{\alpha}_1) & \psi_{23}\boldsymbol{R}_1(\boldsymbol{\alpha}_1) \\ \psi_{13}\boldsymbol{R}_1(\boldsymbol{\alpha}_1) & \psi_{23}\boldsymbol{R}_1(\boldsymbol{\alpha}_1) & \boldsymbol{R}_1(\boldsymbol{\alpha}_1) \end{pmatrix}$$

This parameterization guarantees a smaller number of parameters to be estimated within the correlation. To estimate this matrix, we propose the following modified GEE to estimate $\boldsymbol{\beta}$: first, the system of equations (3-4) must be solved and second, the following estimating equation must be solved for $\boldsymbol{\alpha}_1$:

$$\boldsymbol{U}_2(\boldsymbol{\alpha}_1) = \sum_{i=1}^{n} \left(\frac{\partial \boldsymbol{\varepsilon}_i}{\partial \boldsymbol{\alpha}_1} \right)^T \boldsymbol{H}_i^{-1} \left(\boldsymbol{W}_i - \boldsymbol{\varepsilon}_i \right) \tag{3-7}$$

where $\boldsymbol{H}_i = \boldsymbol{D}(V(r_{ijk}))_{q\times q}$ is a diagonal matrix, $\boldsymbol{\varepsilon}_i = E(\boldsymbol{W}_i)_{q\times 1}$,
$\boldsymbol{W}_i = (r_{i11}r_{i12},\, r_{i11}r_{i13}, \ldots, r_{iP(L-1)}r_{iPL})^T_{q\times 1}$, and r_{ijk} is the ijkth observed Pearson residual defined as:

$$r_{ijk} = \frac{Y_{ijk} - \hat{\mu}_{ijk}}{\hat{\phi}\sqrt{V(\hat{\mu}_{ijk})}} \tag{3-8}$$

and $q = \binom{P}{2}$. For the third step, we propose the following estimator for $\boldsymbol{\Psi} = (\psi_{jj'})_{P\times P}$:

$$\hat{\psi}_{jj'} = \frac{1}{n} \sum_{i=1}^{n} tr\left(\boldsymbol{R}_1(\hat{\boldsymbol{\alpha}}_1)^{-1} (\boldsymbol{r}_{(j)i} - \bar{\boldsymbol{r}}_{(j)})(\boldsymbol{r}_{(j')i} - \bar{\boldsymbol{r}}_{(j')})^T \right) \tag{3-9}$$

where $\boldsymbol{r}_{(j)i}$ is the vector of Pearson residuals for the ith experimental unit at the jth period, and $\bar{\boldsymbol{r}}_{(j)}$ is the average (over i), $j \neq j'$, with $j = 1, \ldots, P$. The estimation of the correlation structure is carried out in two steps: first, $\boldsymbol{R}(\boldsymbol{\alpha}_1)$ is estimated with the residuals grouped by experimental unit with Equation (3-7); then, with the residuals grouped by period and the experimental unit, the $\boldsymbol{\Psi}$ matrix is estimated with Equation (3-9), but discounting the correlation already captured by $\boldsymbol{R}(\boldsymbol{\alpha}_1)$.

Theorem 3.3.1. *The estimator of* $\boldsymbol{R}(\boldsymbol{\alpha})$ *given by* $\hat{\boldsymbol{\Psi}} \otimes \boldsymbol{R}_1(\hat{\boldsymbol{\alpha}}_1)$ *is asymptotically unbiased and consistent.*

Proof. First, asymptotic properties of Pearson's residuals concerning expectation, variance, and residuals are explored. Define the theoretical Pearson residuals of response of the kth of the ith experimental unit at the jth period as:

$$R_{ijk} = \frac{Y_{ijk} - \mu_{ijk}}{\phi\sqrt{V(\mu_{ijk}))}} \tag{3-10}$$

and adapting the results of Cox and Snell (1968) it is true that:

$$\begin{aligned} E(R_{ijk}) &= \sum_{l=1}^{p} B(\hat{\beta}_l) E(H_l^{(ijk)}) \\ &\quad - \sum_{l,s=1}^{p} \mathcal{K}^{ls} E\left(H_l^{(ijk)} U_s^{(ijk)} + \frac{1}{2} H_{ls}^{(ijk)}\right) + O(n^{-1}) \end{aligned} \tag{3-11}$$

$$\begin{aligned} Var(R_{ijk}) &= 1 + 2\sum_{l=1}^{p} B(\hat{\beta}_l) E(R_{ijk} H_l^{(ijk)}) \\ &\quad - \sum_{l,s=1}^{p} \mathcal{K}^{ls} E\left(2 R_{ijk} H_l^{(ijk)} U_s^{(ijk)}\right) - \\ &\quad \sum_{l,s=1}^{p} \mathcal{K}^{ls} E\left(H_l^{(ijk)} H_s^{(ijk)} R_{ijk} H_{ls}^{(ijk)}\right) + O(n^{-1}) \end{aligned} \tag{3-12}$$

where

$$H_l^{(ijk)} = \frac{\partial R_{ijk}}{\partial \beta_l}, \; H_{ls}^{(ijk)} = \frac{\partial^2 R_{ijk}}{\partial \beta_l \partial \beta_s}$$

and $\mathcal{K}^{ls}$ is the element at position (l, s) of the inverse of the Fisher information matrix, and according to Cordeiro and McCullagh (1991), the bias of $\hat{\boldsymbol{\beta}}$ ($B(\hat{\boldsymbol{\beta}})$) is given by:

$$B(\hat{\boldsymbol{\beta}}) = -\frac{1}{2\phi}(\boldsymbol{X^T W X}) \boldsymbol{X}^T \boldsymbol{D}(z_{ijk}) \boldsymbol{F} \mathbf{1} \tag{3-13}$$

$$\boldsymbol{W} = \boldsymbol{D}(V(Y_{ijk})^{\frac{1}{2}}) \boldsymbol{R}(\boldsymbol{\alpha}) \boldsymbol{D}(V(Y_{ijk})^{\frac{1}{2}}) \boldsymbol{D}\left(\frac{\partial \boldsymbol{\mu}_i}{\partial \eta}\right) \tag{3-14}$$

$$\boldsymbol{F} = \boldsymbol{D}\left(V(\mu_{ijk})^{-1}\left(\frac{\partial \mu_{ijk}}{\partial \eta_{ijk}}\right)\left(\frac{\partial^2 \mu_{ijk}}{\partial \eta_{ijk}^2}\right)\right)$$

where $\mathbf{1}$ is a vector of appropriate size whose entries are all equal to 1, $\boldsymbol{D}(z_{ijk})$ is a diagonal matrix with elements on the diagonal given by the variance of the estimated linear predictors, i.e., the diagonal of the matrix

$$\boldsymbol{z} = Var(\hat{\eta}_{111}, \ldots, \hat{\eta}_{nPL}) \tag{3-15}$$

i.e., $z_{ijk} = Var(\hat{\eta}_{ijk})$ and $\boldsymbol{X}$ is the design matrix of the parametric effects described in Equation (3-3). Now, computing the expected values by taking into account the properties of the exponential family, we obtain:

$$E\left(H_l^{(ijk)}\right) = -\sqrt{\phi V(\mu_{ijk})}\left(\frac{\partial \mu_{ijk}}{\partial \eta_{ijk}}\right) x_{l(ijk)} \tag{3-16}$$

$$E\left(H_{ls}^{(ijk)}\right) = \left[2V(\mu_{ijk})^{-\frac{3}{2}}\left(\frac{\partial V(\mu_{ijk})}{\partial \mu_{ijk}}\right)\left(\frac{\partial \mu_{ijk}}{\partial \eta_{ijk}}\right)^2 - 2V(\mu_{ijk})^{-\frac{1}{2}}\left(\frac{\partial^2 \mu_{ijk}}{\partial \eta_{ijk}^2}\right)\right] \times \frac{1}{2}\sqrt{\phi} x_{l(ijk)} x_{s(ijk)} \tag{3-17}$$

$$E\left(H_l^{(ijk)} U_s^{(ijk)}\right) = -\frac{1}{2}\phi^{\frac{1}{2}} V(\mu_{ijk})^{-\frac{3}{2}}\left(\frac{\partial V(\mu_{ijk})}{\partial \mu_{ijk}}\right)\left(\frac{\partial \mu_{ijk}}{\partial \eta_{ijk}}\right)^2 \times x_{l(ijk)} x_{s(ijk)} \tag{3-18}$$

$$E\left(2R_{ijk} H_l^{(ijk)} U_s^{(ijk)}\right) = -V(\mu_{ijk})^{-2}\left(\frac{\partial V(\mu_{ijk})}{\partial \mu_{ijk}}\right)^2\left(\frac{\partial \mu_{ijk}}{\partial \eta_{ijk}}\right)^2 \times x_{l(ijk)} x_{s(ijk)} - 2\phi\left(\frac{\partial V(\mu_{ijk})}{\partial \mu_{ijk}}\right)^{-1}\left(\frac{\partial \mu_{ijk}}{\partial \eta_{ijk}}\right)^2 x_{l(ijk)} x_{s(ijk)} \tag{3-19}$$

$$E\left(H_l^{(ijk)} H_s^{(ijk)}\right) = \left[\phi + \frac{\left(\frac{\partial V(\mu_{ijk})}{\partial \mu_{ijk}}\right)}{4V(\mu_{ijk})}\right] w_{ijk} x_{l(ijk)} x_{s(ijk)} \tag{3-20}$$

$$E\left(R_{ijk} H_{ls}^{(ijk)}\right) = \frac{1}{2}\phi V(\mu_{ijk})^{-\frac{1}{2}}\left(\frac{\partial \mu_{ijk}}{\partial \eta_{ijk}}\right) \tag{3-21}$$

where w_{ijk} is the element ijk of the diagonal of the matrix $\boldsymbol{W}$ defined in Equation (3-14). From Equation (3-16) and the bias of $\boldsymbol{\beta}$ given in Equation (3-13), it follows that:

$$\sum_{l=1}^{p} B(\hat{\beta}_l) E\left(H_l^{(ijk)}\right) = -\phi^{\frac{1}{2}} V(\mu_{ijk})^{-\frac{1}{2}}\left(\frac{\partial \mu_{ijk}}{\partial \eta_{ijk}}\right) \boldsymbol{e}_{ijk} \boldsymbol{X} B(\hat{\boldsymbol{\beta}}) \tag{3-22}$$

where $\boldsymbol{e}_{ijk}$ is a vector of zeros with a 1's at the ijkth position. From equations (3-17) and (3-18), we get:

$$E\left(H_l^{(ijk)} U_s^{(ijk)} + \frac{1}{2} H_{ls}^{(ijk)}\right) = -\frac{1}{2}\phi^{\frac{1}{2}} V(\mu_{ijk})^{-\frac{1}{2}}\left(\frac{\partial^2 \mu_{ijk}}{\partial \eta_{ijk}^2}\right)^2 x_{l(ijk)} x_{s(ijk)}$$

$$\sum_{l,s=1}^{p} \mathcal{K}^{ls} E(H_l^{(ijk)} U_s^{(ijk)} + \frac{1}{2} H_{ls}^{(ijk)}) = -\frac{1}{2}\phi^{\frac{1}{2}} V(\mu_{ijk})^{-\frac{1}{2}}\left(\frac{\partial^2 \mu_{ijk}}{\partial \eta_{ijk}^2}\right)^2 x_{l(ijk)} x_{s(ijk)} \tag{3-23}$$

and therefore, from equations (3-22), (3-23) and (3-11):

$$E(R_{111}, R_{112}, \ldots, R_{nPL}) = \frac{-1}{2\sqrt{\phi}}(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{J}\boldsymbol{z} \tag{3-24}$$

where

$$\boldsymbol{H} = \boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X})^{\frac{1}{2}}\boldsymbol{X}^{T}\boldsymbol{W}^{\frac{1}{2}}$$
$$\boldsymbol{J} = \boldsymbol{D}\left(V(Y_{ijk})\right)\boldsymbol{D}\left(\frac{\partial \boldsymbol{\mu}_i^2}{\partial^2 \eta}\right)$$

from equations (3-19), (3-20) and (3-21):

$$-\sum_{l,s=1}^{p}\mathcal{K}^{ls}E\left(2R_{ijk}H_l^{(ijk)}U_s^{(ijk)} + H_l^{(ijk)}H_s^{(ijk)}R_{ijk}H_{ls}^{(ijk)}\right)$$
$$= \left[-\phi w_{ijk} - \frac{\left(\frac{\partial V(\mu_{ijk})}{\partial \mu_{ijk}}\right)\left(\frac{\partial^2 \mu_{ijk}}{\partial \eta_{ijk}^2}\right)}{2V(\mu_{ijk})} - \frac{1}{2}w_{ijk}\left(\frac{\partial^2 V(\mu_{ijk})}{\partial \mu_{ijk}^2}\right)\right]\frac{z_{ijk}}{\phi} \tag{3-25}$$

and from equations (3-13) and (3-21), it follows that:

$$2\sum_{l=1}^{p}B(\hat{\beta}_l)E(R_{ijk}H_{ls}^{(ijk)})$$
$$= \frac{1}{2\phi}\frac{\left(\frac{\partial V(\mu_{ijk})}{\partial \mu_{ijk}}\right)\left(\frac{\partial^2 \mu_{ijk}}{\partial \eta_{ijk}^2}\right)}{V(\mu_{ijk})}\boldsymbol{e}_{ijk}\boldsymbol{Z}\boldsymbol{D}(z_{ii})\boldsymbol{D}\left(V(\mu_{ijk})^{-1}\left(\frac{\partial^2 \mu_{ijk}}{\partial \eta_{ijk}^2}\right)\left(\frac{\partial \mu_{ijk}}{\partial \eta_{ijk}}\right)\right)\mathbf{1} \tag{3-26}$$

Therefore, from the results (3-25) and (3-26), we have that:

$$[Var(R_{111}), Var(R_{112}), \ldots, Var(R_{nPL})] = \mathbf{1} + \frac{1}{2\phi}\left(\boldsymbol{Q}\boldsymbol{H}\boldsymbol{J} - \boldsymbol{M}\right)\boldsymbol{z} \tag{3-27}$$

where $\mathbf{1}$ is a vector of ones and

$$\boldsymbol{Q} = \boldsymbol{D}\left(V(Y_{ijk})\right)^{\frac{1}{2}}\boldsymbol{D}\left(\frac{\partial V(Y_{ijk})}{\partial \mu}\right)$$
$$\boldsymbol{M} = \boldsymbol{D}\left(V(Y_{ijk})\right)^{\frac{1}{2}}\boldsymbol{D}\left(\frac{\partial V(Y_{ijk})^2}{\partial^2 \mu} + 2\phi\boldsymbol{D}(\boldsymbol{W}) + \frac{\frac{\partial V(Y_{ijk})}{\partial \mu}}{V(Y_{ijk})}\right)$$

By theorem 2 of Liang and Zeger (1986), it is known that the GEE estimators of η_{ijk} is consistent and unbiased, i.e.,

$$\boldsymbol{Z} \xrightarrow[n\to\infty]{p} \mathbf{0} \tag{3-28}$$

Thus from (3-24) and (3-27), we find that:

$$E(R_{ijk}) = O(n^{-1}) \tag{3-29}$$
$$Var(R_{ijk}) = 1 + O(n^{-1}) \tag{3-30}$$

and furthermore, by Section 3 of Cordeiro (2004) and equations (3-29) and (3-30), it follows that:

$$R_{ijk} \xrightarrow[n\to\infty]{d} N(0,1) \tag{3-31}$$

Let

$$\boldsymbol{\Gamma} = \{\gamma_{ij}\}_{n\times n} \tag{3-32}$$

be a matrix whose first column is $\frac{\mathbf{1}}{\sqrt{n}}$ and the following columns are:

$$g_{i-1} = \left(\frac{1}{\sqrt{(i-1)i}}, \ldots, \frac{1}{\sqrt{(i-1)i}}, -\frac{i-1}{\sqrt{(i-1)i}}, 0, \ldots, 0\right), \qquad i = 2, \ldots, n \tag{3-33}$$

$$\boldsymbol{\Gamma}^T = \left(\frac{\mathbf{1}}{\sqrt{n}} \;\vdots\; \boldsymbol{G}\right)^T = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & \cdots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{3}{\sqrt{12}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{(n(n-1))}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & -\frac{n-1}{\sqrt{n(n-1)}} \end{pmatrix}$$

where

$$\boldsymbol{G}^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & \cdots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{3}{\sqrt{12}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{(n(n-1))}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & -\frac{n-1}{\sqrt{n(n-1)}} \end{pmatrix}$$

then the matrix $\boldsymbol{\Gamma}$ is a Helmert matrix (Lancaster, 1965) and therefore:

$$\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T = \boldsymbol{I}_n, \qquad \mathbf{1}_{n-1}^T\boldsymbol{G} = 0, \qquad \boldsymbol{G}^T\boldsymbol{G} = \boldsymbol{I}_{n-1} - \frac{1}{n-1}\mathbf{1}_{n-1}\mathbf{1}_{n-1}^T \tag{3-34}$$

If r_{ijk} is defined as the estimated Pearson residual of the ith experimental unit in the jth period and the kth observation, i.e.,

$$r_{ijk} = \hat{R}_{ijk} = \frac{Y_{ijk} - \hat{\mu}_{ijk}}{\hat{\phi}\sqrt{V(\hat{\mu}_{ijk})}}$$

and the matrix $\boldsymbol{r}_i$ of residuals of the ith individual, where the first row has the L Pearson residuals defined in Equation (3-8) corresponding to the first period, and the second row of the L corresponding to the second period and so on until completing a matrix with P rows and L columns, i.e.:

$$\boldsymbol{r}_i = \begin{pmatrix} r_{i11} & r_{i21} & \cdots & r_{i1L} \\ r_{i21} & r_{i22} & \cdots & r_{i2L} \\ \vdots & \vdots & \ddots & \vdots \\ r_{iP1} & r_{i2P} & \cdots & r_{iPL} \end{pmatrix} \tag{3-35}$$

By Equation (3-29) and the correlation assumption given in Equation (3-6), it is true that:

$$\begin{aligned} &E(\boldsymbol{r}_i) = \boldsymbol{0}_{P\times L} \\ &Corr\left(Vec(\boldsymbol{r}_i)\right) = \boldsymbol{\Psi} \otimes \boldsymbol{R}_1(\boldsymbol{\alpha}_1) \\ &Corr\left(Vec(\boldsymbol{r}_i), Vec(\boldsymbol{r}_{i'})\right) = \boldsymbol{0}_{PL\times PL} \qquad i \neq i' \end{aligned} \tag{3-36}$$

And defining $\boldsymbol{R}$ as:

$$\boldsymbol{R} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n)_{P\times nL}$$

and since $\boldsymbol{\Gamma}$ is orthogonal, then $\boldsymbol{\Gamma} \otimes \boldsymbol{I}_L$ is also orthogonal. Thus (Srivastava et al., 2008):

$$\boldsymbol{R}(\boldsymbol{\Gamma} \otimes \boldsymbol{I}_L) = \left(\sqrt{n}\bar{\boldsymbol{r}} \;\vdots\; \boldsymbol{R}(\boldsymbol{G} \otimes \boldsymbol{I}_L)\right) \tag{3-37}$$

and according to Equation (3-36), we get:

$$\begin{aligned} \boldsymbol{R}\left(\boldsymbol{I}_n \otimes \boldsymbol{\Psi}^{-1}\right)\boldsymbol{R}^T &= (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n)\left(\boldsymbol{I}_n \otimes \boldsymbol{\Psi}^{-1}\right)(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n)^T \\ &= n\bar{\boldsymbol{r}} \otimes \boldsymbol{\Psi}^{-1}\boldsymbol{R} + \boldsymbol{R}(\boldsymbol{G} \otimes \boldsymbol{I}_L)\left(\boldsymbol{I}_n \otimes \boldsymbol{\Psi}^{-1}\right)(\boldsymbol{G}^T \otimes \boldsymbol{I}_L)\boldsymbol{R}^T \\ &= n\bar{\boldsymbol{r}} \otimes \boldsymbol{\Psi}^{-1}\boldsymbol{R} + \boldsymbol{Z}(\boldsymbol{G}^T \otimes \boldsymbol{I}_L)\boldsymbol{Z}^T \end{aligned}$$

where $\boldsymbol{Z}$ is:

$$\boldsymbol{Z}_{P\times(n-1)L} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_{(n-1)}) = \boldsymbol{R}(\boldsymbol{G} \otimes \boldsymbol{I}_L) = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n)(\boldsymbol{G} \otimes \boldsymbol{I}_L) \tag{3-38}$$

with $\bar{\boldsymbol{r}}$ is the matrix of the average residuals defined in Equation (3-35) for each period, that is,

$$\bar{\boldsymbol{r}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{r}_i = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} r_{i11} & r_{i21} & \cdots & r_{i1L} \\ r_{i21} & r_{i22} & \cdots & r_{i2L} \\ \vdots & \vdots & \ddots & \vdots \\ r_{iP1} & r_{i2P} & \cdots & r_{iPL} \end{pmatrix}$$

and

$$\boldsymbol{Z}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}}r_{111} - \frac{1}{\sqrt{2}}r_{211} & \cdots & \frac{1}{\sqrt{2}}r_{11L} - \frac{1}{\sqrt{2}}r_{21L} \\ \frac{1}{\sqrt{2}}r_{121} - \frac{1}{\sqrt{2}}r_{221} & \cdots & \frac{1}{\sqrt{2}}r_{12L} - \frac{1}{\sqrt{2}}r_{22L} \\ \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{2}}r_{1P1} - \frac{1}{\sqrt{2}}r_{2P1} & \cdots & \frac{1}{\sqrt{2}}r_{1PL} - \frac{1}{\sqrt{2}}r_{2PL} \end{pmatrix}_{P\times L}$$

$$\boldsymbol{Z}_2 = \begin{pmatrix} \frac{1}{\sqrt{6}}\sum_{i=1}^{2} r_{i11} - \frac{2}{\sqrt{6}}r_{311} & \cdots & \frac{1}{\sqrt{6}}\sum_{i=1}^{2} r_{i1L} - \frac{2}{\sqrt{6}}r_{31L} \\ \frac{1}{\sqrt{6}}\sum_{i=1}^{2} r_{i21} - \frac{2}{\sqrt{6}}r_{321} & \cdots & \frac{1}{\sqrt{6}}\sum_{i=1}^{2} r_{i2L} - \frac{2}{\sqrt{6}}r_{32L} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{6}}\sum_{i=1}^{2} r_{iP1} - \frac{2}{\sqrt{6}}r_{3P1} & \cdots & \frac{1}{\sqrt{6}}\sum_{i=1}^{2} r_{iPL} - \frac{2}{\sqrt{6}}r_{3PL} \end{pmatrix}_{P\times L}$$

$$\vdots$$

$$\boldsymbol{Z}_{(n-1)} = \begin{pmatrix} \frac{1}{\sqrt{n(n-1)}}\sum_{i=1}^{n-1} r_{i11} - \frac{n-1}{\sqrt{n(n-1)}}r_{(n-1)11} & \cdots & \frac{1}{\sqrt{n(n-1)}}\sum_{i=1}^{n-1} r_{i1L} - \frac{n-1}{\sqrt{n(n-1)}}r_{(n-1)1L} \\ \frac{1}{\sqrt{n(n-1)}}\sum_{i=1}^{n-1} r_{i21} - \frac{n-1}{\sqrt{n(n-1)}}r_{(n-1)21} & \cdots & \frac{1}{\sqrt{n(n-1)}}\sum_{i=1}^{n-1} r_{i2L} - \frac{n-1}{\sqrt{n(n-1)}}r_{(n-1)2L} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n(n-1)}}\sum_{i=1}^{n-1} r_{iP1} - \frac{n-1}{\sqrt{n(n-1)}}r_{(n-1)P1} & \cdots & \frac{1}{\sqrt{n(n-1)}}\sum_{i=1}^{n-1} r_{iPL} - \frac{n-1}{\sqrt{n(n-1)}}r_{(n-1)PL} \end{pmatrix}_{P\times L}$$

Now by the properties of the Pearson residuals, we have that:

$$E(\boldsymbol{Z}_1) = E(\boldsymbol{Z}_2) = \cdots = E(\boldsymbol{Z}_{n-1}) = \boldsymbol{0}_{P\times L}$$

and by the properties given in Equation (3-34) and because we assume that the experimental units are independent, that is, $Corr(r_{ijk}, r_{i'j'k'}) = 0$, for all $i \neq i'$ and that Equation (3-3) is

true, then:

$$Corr(r_{ijk}, r_{i'j'k'}) = 0 \qquad \forall i \neq i'$$
$$Corr(r_{ijk}, r_{ij'k'}) = Corr(r_{i'jk}, r_{i'j'k'}) = Corr(r_{i'jk}, r_{i'j'k'}), \qquad \forall i \neq i'$$

$$Corr\left(\frac{1}{\sqrt{2}}r_{111} - \frac{1}{\sqrt{2}}r_{211}, \frac{1}{\sqrt{2}}r_{121} - \frac{1}{\sqrt{2}}r_{221}\right) = \frac{1}{2}Corr(r_{111}, r_{121}) + \frac{1}{2}Cov(r_{211}, r_{221})$$
$$= Corr(r_{111}, r_{121}) = Corr(r_{111}, r_{121})$$

$$Corr\left(\frac{1}{\sqrt{n(n-1)}}\sum_{i=1}^{n-1} r_{i11} - \frac{n-1}{\sqrt{n(n-1)}}r_{(n-1)11}, \right.$$
$$\left.\frac{1}{\sqrt{n(n-1)}}\sum_{i=1}^{n-1} r_{i21} - \frac{n-1}{\sqrt{n(n-1)}}r_{(n-1)21}\right)$$
$$= Corr(r_{111}, r_{121}) = Corr(r_{111}, r_{121})$$

furthermore,

$$Var(Vec(\boldsymbol{Z}_1)) = Var\left\{\begin{pmatrix} \frac{1}{\sqrt{2}}r_{111} - \frac{1}{\sqrt{2}}r_{211} \\ \frac{1}{\sqrt{2}}r_{121} - \frac{1}{\sqrt{2}}r_{221} \\ \vdots \\ \frac{1}{\sqrt{2}}r_{1P1} - \frac{1}{\sqrt{2}}r_{2P1} \\ \frac{1}{\sqrt{2}}r_{112} - \frac{1}{\sqrt{2}}r_{212} \\ \vdots \\ \frac{1}{\sqrt{2}}r_{1PL} - \frac{1}{\sqrt{2}}r_{2PL} \end{pmatrix}_{PL\times 1}\right\}$$
$$= \begin{pmatrix} 1 & Corr(r_{111}, r_{121}) & \cdots & Corr(r_{111}, r_{1P1}) \\ Corr(r_{111}, r_{121}) & 1 & \cdots & Corr(r_{121}, r_{1P1}) \\ \vdots & \vdots & \ddots & \vdots \\ Corr(r_{111}, r_{1P1}) & Corr(r_{121}, r_{1P1}) & \cdots & 1 \end{pmatrix}_{PL\times PL}$$
$$= \boldsymbol{\Psi} \otimes \boldsymbol{R}_1(\boldsymbol{\alpha}_1))$$

$$Var(Vec(\boldsymbol{Z}_1)) = Var(Vec(\boldsymbol{Z}_2)) = \cdots = Var(Vec(\boldsymbol{Z}_{(n-1)})) = \boldsymbol{\Psi} \otimes \boldsymbol{R}_1(\boldsymbol{\alpha}_1)) \tag{3-39}$$

By the central limit theorem, we get that:

$$Vec(\bar{\boldsymbol{r}}) \xrightarrow[n\to\infty]{d} N_{LP}(\boldsymbol{0}, \boldsymbol{\Psi} \otimes \boldsymbol{R}_1(\boldsymbol{\alpha}_1)) \tag{3-40}$$

By equations (3-31) and (3-39) it follows that:

$$Vec(\boldsymbol{Z}_j) \xrightarrow[n\to\infty]{d} N_{LP}(\boldsymbol{0}, \boldsymbol{\Psi} \otimes \boldsymbol{R}_1(\boldsymbol{\alpha}_1)) \tag{3-41}$$

and by equations (3-34), (3-40) y (3-41) that:

$$Cov(Vec(\bar{\boldsymbol{r}}), Vec(\boldsymbol{Z}_j)) = \boldsymbol{0}$$
$$Cov(Vec(\boldsymbol{Z}_i), Vec(\boldsymbol{Z}_j)) = \boldsymbol{0}$$

and partitioning $\boldsymbol{Z}_1$ as follows:

$$\begin{aligned}
\boldsymbol{Z}_1 &= \begin{pmatrix} \frac{1}{\sqrt{2}}r_{111} - \frac{1}{\sqrt{2}}r_{211} & \cdots & \frac{1}{\sqrt{2}}r_{11L} - \frac{1}{\sqrt{2}}r_{21L} \\ \frac{1}{\sqrt{2}}r_{121} - \frac{1}{\sqrt{2}}r_{221} & \cdots & \frac{1}{\sqrt{2}}r_{12L} - \frac{1}{\sqrt{2}}r_{22L} \\ \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{2}}r_{1P1} - \frac{1}{\sqrt{2}}r_{2P1} & \cdots & \frac{1}{\sqrt{2}}r_{1PL} - \frac{1}{\sqrt{2}}r_{2PL} \end{pmatrix}_{P\times L} \\
&= \begin{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}}r_{111} - \frac{1}{\sqrt{2}}r_{211} \\ \frac{1}{\sqrt{2}}r_{121} - \frac{1}{\sqrt{2}}r_{221} \\ \vdots \\ \frac{1}{\sqrt{2}}r_{1P1} - \frac{1}{\sqrt{2}}r_{2P1} \end{pmatrix}_{P\times 1} & \cdots & \begin{pmatrix} \frac{1}{\sqrt{2}}r_{11L} - \frac{1}{\sqrt{2}}r_{21L} \\ \frac{1}{\sqrt{2}}r_{1PL} - \frac{1}{\sqrt{2}}r_{2PL} \\ \vdots \\ \frac{1}{\sqrt{2}}r_{1PL} - \frac{1}{\sqrt{2}}r_{2PL} \end{pmatrix}_{P\times 1} \end{pmatrix} \\
&= (\boldsymbol{z}_{(1)1}, \ldots, \boldsymbol{z}_{(L)1})
\end{aligned}$$

Then, it is obtained that

$$E(\boldsymbol{z}_{(j)1}, \boldsymbol{z}_{(j)1}^T) = \Psi_{jj}\boldsymbol{R}_1(\boldsymbol{\alpha}_1) \qquad E(\boldsymbol{Z}_1, \boldsymbol{Z}_1^T) = (trace(\boldsymbol{\Psi}))\boldsymbol{R}_1(\boldsymbol{\alpha}_1)$$

Similarly, it is found that:

$$\begin{aligned}
&E(\boldsymbol{z}_{(j)1}, \boldsymbol{z}_{(j)1}^T) = \Psi_{jj}\boldsymbol{R}_1(\boldsymbol{\alpha}_1) \\
&E(\boldsymbol{z}_{(j)2}, \boldsymbol{z}_{(j)2}^T) = \Psi_{jj}\boldsymbol{R}_1(\boldsymbol{\alpha}_1) \\
&\vdots \\
&E(\boldsymbol{z}_{(j)(n-1)}, \boldsymbol{z}_{(j)(n-1)}^T) = \Psi_{jj}\boldsymbol{R}_1(\boldsymbol{\alpha}_1)
\end{aligned} \tag{3-42}$$

$$\begin{aligned}
&E(\boldsymbol{Z}_1, \boldsymbol{Z}_1^T) = (trace(\boldsymbol{\Psi}))\boldsymbol{R}_1(\boldsymbol{\alpha}_1) \\
&\vdots \\
&E(\boldsymbol{Z}_{(n-1)}, \boldsymbol{Z}_{(n-q)}^T) = (trace(\boldsymbol{\Psi}))\boldsymbol{R}_1(\boldsymbol{\alpha}_1)
\end{aligned} \tag{3-43}$$

Therefore, since $\boldsymbol{\Psi}_{jj} = 1$, $\forall j = 1, \ldots, P$, $trace(\boldsymbol{\Psi}) = 1$ from Equation (3-42), we get:

$$E\left(\boldsymbol{R}_1(\boldsymbol{\alpha}_1)^{\frac{1}{2}}\boldsymbol{z}_{(j')k}\boldsymbol{z}_{(j)k}^T\boldsymbol{R}_1(\boldsymbol{\alpha}_1)^{\frac{1}{2}}\right) = \boldsymbol{\Psi}_{jj'}\boldsymbol{I}_P, \qquad \forall k = 1, \ldots, n-1 \tag{3-44}$$

$$Cov[\boldsymbol{R}_1(\boldsymbol{\alpha}_1)^{\frac{1}{2}}\boldsymbol{z}_{(k)j}\boldsymbol{z}_{(i)j}^T\boldsymbol{R}_1(\boldsymbol{\alpha}_1)^{\frac{1}{2}}, \boldsymbol{R}_1(\boldsymbol{\alpha}_1)^{\frac{1}{2}}\boldsymbol{z}_{(k)j'}\boldsymbol{z}_{(i)j'}^T\boldsymbol{R}_1(\boldsymbol{\alpha}_1)^{\frac{1}{2}}] = \mathbf{0} \tag{3-45}$$

By Theorem 2 in Liang and Zeger (1986), it is known that $\boldsymbol{R}_1(\hat{\boldsymbol{\alpha}}_1)$ is consistent and unbiased for $\boldsymbol{R}_1(\boldsymbol{\alpha}_1)$, by equations (3-44), (3-40) and (3-41), we have that

$$\hat{\psi}_{jj'} = \frac{1}{n}\sum_{i=1}^{n} tr\left(\boldsymbol{R}_1(\hat{\boldsymbol{\alpha}}_1)^{-1}(\boldsymbol{r}_{(j)i} - \bar{\boldsymbol{r}}_{(j)})(\boldsymbol{r}_{(j')i} - \bar{\boldsymbol{r}}_{(j')})^T\right) \tag{3-46}$$

is a consistent and asymptotically unbiased estimator for $\Psi_{jj'}$, which proves the theorem. □

The form of $\boldsymbol{R}_1(\boldsymbol{\alpha}_1)$ in (3-6) is given by correlation structures such as: independence, autoregressive, exchangeable, etc (Davis, 2002).
As the QIC is a measure that works well to select the most suitable correlation matrix in GEE, this measure will be used to compare models with different correlation matrices Pan (2001a). The QIC is defined as:

$$QIC = -2QL(\hat{\mu}; \boldsymbol{I}) + 2trace(\hat{\boldsymbol{\Omega}}_I^{-1}\hat{\boldsymbol{V}}_{\boldsymbol{R}}) \tag{3-47}$$

where $\hat{\mu}_{ijk} = \hat{\eta}_{ijk} = g^{-1}(\boldsymbol{x}_{ijk}\hat{\beta})$ is the estimated expected value of observation Y_{ijk} under the model assuming the correlation matrix R, $\hat{\Omega}_I$ is the estimated covariance matrix of $\boldsymbol{\beta}$ under independence, and $\hat{V}_R$ is the covariance matrix of $\boldsymbol{\beta}$ under the model with correlation matrix $\boldsymbol{R}$ defined as in (3-6).

3.4. Simulation Study

3.4.1. With unstructured between periods

A study on the performance of the estimation method, as well as the capacity of the QIC to select the correlation matrix, was carried out. A crossover design with an extra period was used, which consists of two sequences of three periods (ABA and BAB). Within each observation period, 2, 5, 10 and 15 repetitions of the response variable per individual were analyzed; Regarding the number of individuals per sequence, values of 5, 10 and 100 were used. The effect of period was sampled from a normal standard, as well as the effect of time within each period.
For the treatment effect (difference between B and A), values between -2 and 2 were analyzed in a similar way to the carry-over effect of B on A. Each of the above combinations was simulated for a normal response and a binary response with a theoretical correlation matrix $\boldsymbol{\Psi}$ as follows:

$$\boldsymbol{\Psi} = \begin{pmatrix} 1 & \psi_{12} & \psi_{13} \\ \psi_{21} & 1 & \psi_{23} \\ \psi_{13} & \psi_{23} & 1 \end{pmatrix}, \qquad \psi_{ij} \sim U(-1,1) \text{ and } det(\boldsymbol{\Psi}) > 0$$

As for the matrix $\boldsymbol{R}_1(\boldsymbol{\alpha_1})$, one of two structures was first selected: exchangeable and autoregressive. Then, a value of α_1 was selected from a uniform distribution of -1 to 1.
In each scenario, six models were fitted with the same linear predictor, but different correlation matrices; the models were: 1) Independence, independence between and within periods, it is similar to a GLM, 2) Unstructured, a unstructured correlation matrix within periods, 3) Exchangeable, an exchangeable correlation matrix within periods and independence between periods, 4) Autoregresive, an autoregresive order 1 correlation matrix within periods and independence between periods, 5) Kron-Exchangeable, with kronecker matrix and exchangeable within periods using the methodology proposed in this work and 6) Kron-Autoregresive, one with kronecker matrix and autoregressive within periods using the methodology proposed in this work.
Within each scenario, the estimated value of the treatment effect and associated carry-over,

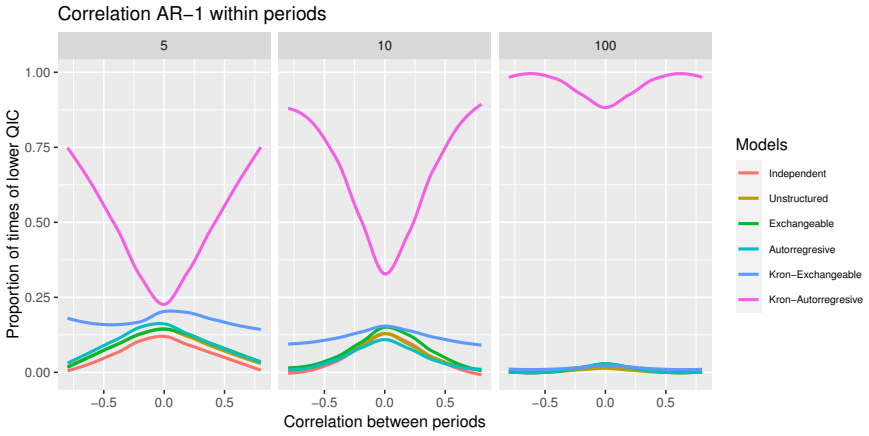

Figure 3-1.: Proportion of times the model obtains the lowest QIC relative to the value of Ψ, the value of the correlation between periods when the correlation within periods is autoregresive. The proportion was calculated over all simulations where the correlation matrix was exchangeable and Kronecker product with $\boldsymbol{\Psi}$. Each subplot represents the number of individuals per sequence.

the root mean square error of each estimate, the QIC, and the p-value of the hypothesis of no treatment effect were calculated. Each scenario was run 1000 times.
In Figures 3-1 and 3-2, the x-axis represents the minimum correlation between periods, i.e., mín$(\psi_{12}, \psi_{1,3}, \psi_{23})$, and the y-axis represents the percentage of times that each of the models obtains the lowest QIC with respect to the other five models. It is observed that the QIC is used to find the most suitable model for the data set, i.e, which has a correlation structure similar to the theoretical one, and as the number of individuals per sequence increases, the percentage of times that the models with Kronecker-exchangeable correlation matrix have the lowest QIC increases (Figure 3-2). On the other hand, the same was observed with the

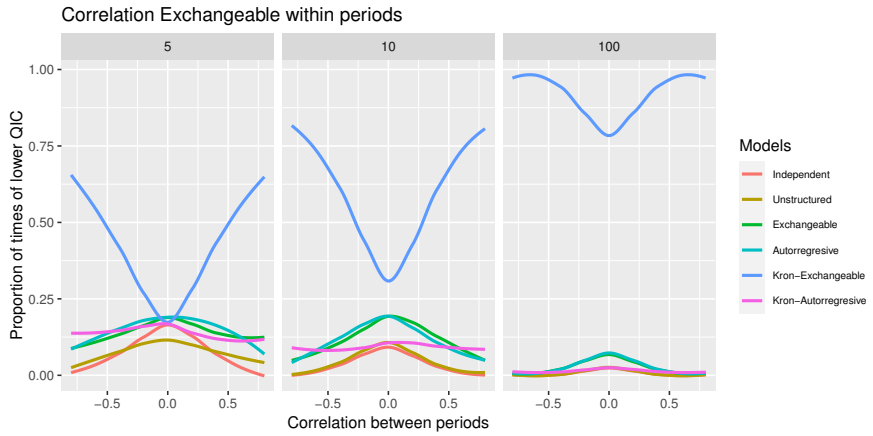

Figure 3-2.: Proportion of times the model obtains the lowest QIC relative to the value of Ψ, the value of the correlation between periods when the correlation within periods is Exchangeable. The proportion was calculated over all simulations where the correlation matrix was exchangeable and Kronecker product with $\boldsymbol{\Psi}$. Each subplot represents the number of individuals per sequence.

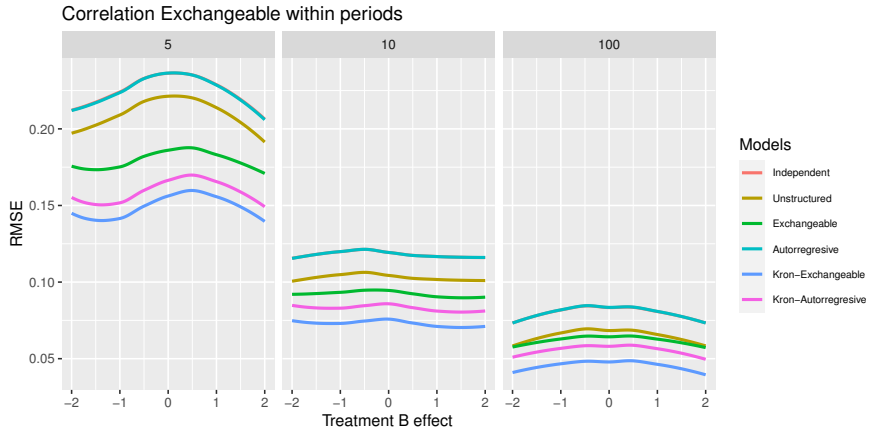

Figure 3-3.: RMSE for each value of the treatment effect. The average over all simulations where the correlation matrix was exchangeable and Kronecker product with $\boldsymbol{\Psi}$ is shown. Each subplot represents the number of individuals per sequence.

Kronecker-autoregresive correlation matrix (Figure **3-1**). This proportion is lower when the correlation between periods is close to 0, which is consistent with what is expected since at this point the matrix $\boldsymbol{\Psi}$ is close to the identity.

Figure **3-3** shows the average RMSE for estimating the treatment effect in each of the models when the matrix defined in Equation (3-6), where $\boldsymbol{R}_1(\boldsymbol{\alpha}_1)$ is an exchangeable correlation matrix. It is observed that as the number of individuals per sequence increases, the RMSE

decreases, which is in accordance with expectations. However, it is also observed that the

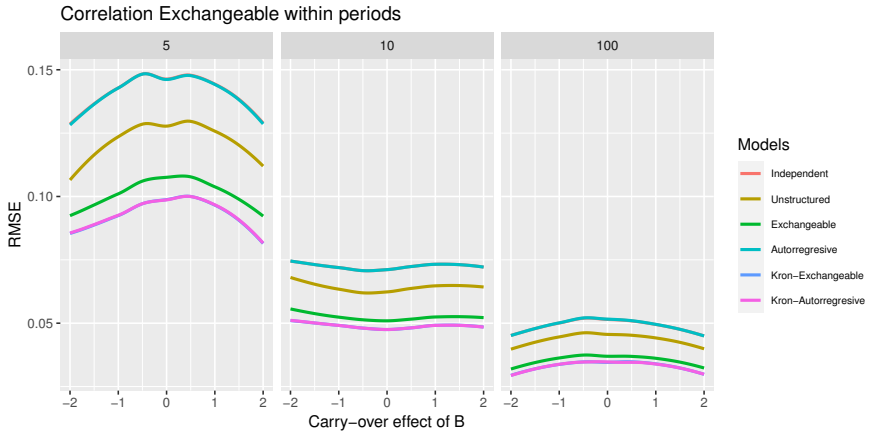

Figure 3-4.: RMSE for each value of the carry-over effect. The average over all simulations where the correlation matrix was exchangeable and Kronecker product with $\mathbf{\Psi}$ is shown. Each subplot represents the number of individuals per sequence.

model with Kronecker exchangeable correlation matrix has a better behavior than the other five models. The Kronecker AR1 model has a higher RMSE than the exchangeable model, but lower than all the others. In all scenarios with the assumption of independence or with unstructured correlation have a higher RMSE. This means that the Kronecker specification works well to capture the treatment effect more accurately than the usual correlation models. In addition, since the average is being calculated over all the scenarios, it is observed that the proposed model is robust to low or high correlation between periods and within periods.

The RMSE for the estimation of the carry-over effect was similar for the two models proposed in this work, but they showed a better performance (as measured by the QIC) than the other models, as can be seen in figures **3-4** and **3-6**.
If Figure **3-5** is observed, the model with the best RMSE for treatment is the one with the autoregressive correlation matrix, this means that the Kronecker specification estimates the treatment effects more accurately than the other models. However, the other model with a Kronecker and exchangeable correlation matrix does not differ too much, that is, both models have a low RMSE compared to the other four models.
It can be seen in Figure **3-7** that the power of the test increases as the value of the treatment effect moves away from zero, but when the effect is null, it approaches 5 %, that is, the six models have a type I error similar to the theoretical one. This trend is maintained when the number of individuals increases to 10 and 100, but the power of the autoregressive Kronecker model is always greater than the other five, although close to the Kronecker exchangeable model, which shows the advantage in estimating the treatment effect of the proposed models.

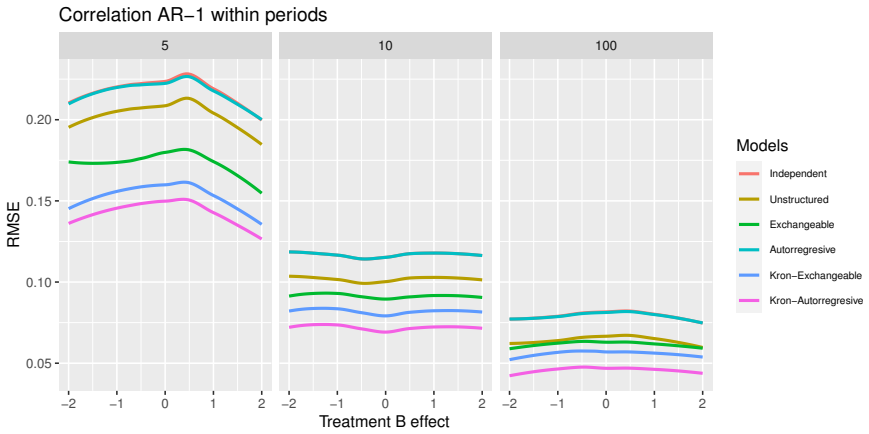

Figure 3-5.: RMSE for each value of the treatment effect. The average over all simulations where the correlation matrix was autoregresive and Kronecker product with $\mathbf{\Psi}$ is shown. Each subplot represents the number of individuals per sequence.

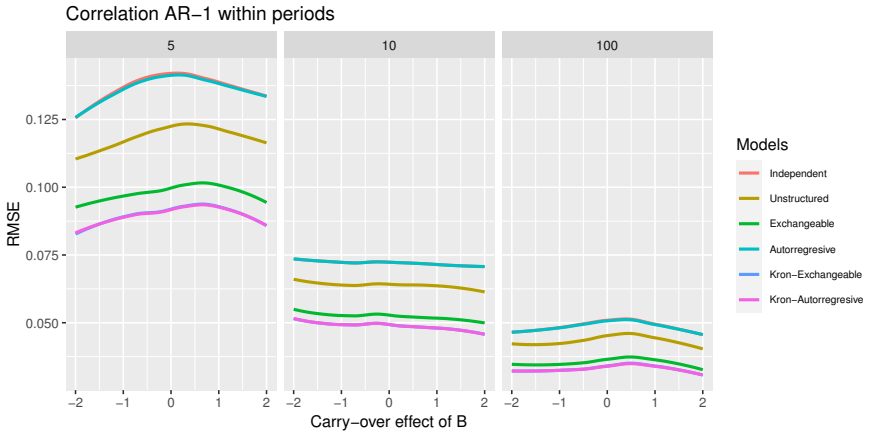

Figure 3-6.: RMSE for each value of the carry-over effect. The average over all simulations where the correlation matrix was autoregresive and Kronecker product with $\mathbf{\Psi}$ is shown. Each subplot represents the number of individuals per sequence.

If instead Figure 3-8 is analyzed, the behavior is similar to that described above, but the model with the best behavior is the Kronecker exchangeable. Furthermore, both the RMSE and the percentage were calculated over all possible combinations of correlation values, demonstrating the robustness of the proposed model to changes in correlation values and possible carry-over effects.

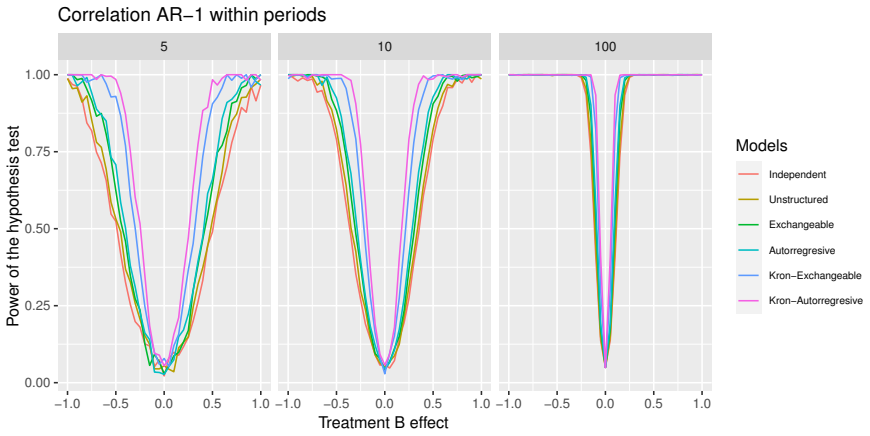

Figure 3-7.: Percentage of times that the null hypothesis $H_0 : \tau_B = 0$ is rejected. The percentage was calculated over all simulations where the correlation matrix was autoregresive and Kronecker product with $\boldsymbol{\Psi}$. Each subplot represents the number of individuals per sequence.

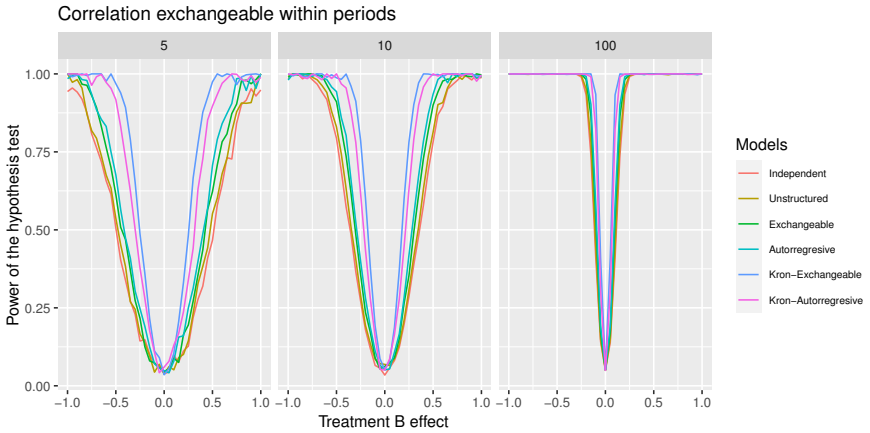

Figure 3-8.: Percentage of times that the null hypothesis $H_0 : \tau_B = 0$ is rejected. The percentage was calculated over all simulations where the correlation matrix was exchangeable and Kronecker product with $\boldsymbol{\Psi}$. Each subplot represents the number of individuals per sequence.

3.4.2. With $\boldsymbol{\Psi} = \boldsymbol{I}_3$

In this subsection, a theoretical correlation matrix $\boldsymbol{\Psi} = \boldsymbol{I}_3$ is defined for the simulation. Similar to the matrix $\boldsymbol{R}_1(\boldsymbol{\alpha_1})$, one of two structures was first selected: exchangeable and autoegressive. Then, a value of α_1 was selected from a uniform distribution of -1 to 1. In Figures **3-9** and **3-10**, the x-axis represents the value α of correlation within periods, and the y-axis represents the percentage of times that each of the models obtains the lowest QIC

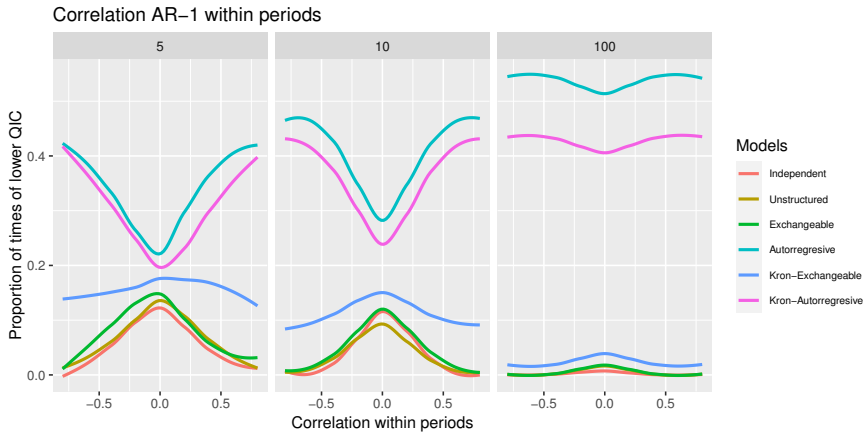

Figure 3-9.: Proportion of times the model obtains the lowest QIC relative to the value of Ψ, the value of the correlation between periods when the correlation within periods is autoregresive. The proportion was calculated over all simulations where the correlation matrix was exchangeable and Kronecker product with $\boldsymbol{\Psi} = \boldsymbol{I}_3$. Each subplot represents the number of individuals per sequence.

with respect to the other five models. Figure **3-9** showed a similar behavior in the model with autoregressive correlation (the same theoretical) and the correlation model with Kronecker product with autoregressive correlation proposed in this work.

As the number of individuals per sequence increases, in Figure **3-10**, the percentage of times that the model with Kronecker and interchangeable correlation matrix has the lowest QIC increases. On the other hand, the same was observed but with a model with Kronecker and an autoregressive correlation matrix.

This proportion is lower when the correlation between periods is close to 0, which is consistent with what is expected since at this point the matrix $\boldsymbol{\Psi} \otimes \boldsymbol{R}_1(\boldsymbol{\alpha_1})$ is close to the identity.

Figure **3-11** shows the average RMSE to estimate the treatment effect in each of the models when $\boldsymbol{R}_1(\boldsymbol{\alpha}_1)$ is an interchangeable correlation matrix. It is observed that as the number of individuals per sequence increases, the RMSE decreases, which is in agreement with what was expected.

However, it was also observed that the Kronecker-interchangeable correlation matrix model performs better than the other five models when there are 5 individuals per sequence, and the Kronecker-autoregresive model and the interchangeable-only model have a higher RMSE than the interchangeable model, but lower than all the others. However, when the number of individuals increases to 10 or 100, the models with the lowest RMSE were the Kronecker-interchangeable and the interchangeable correlation matrices.

All scenarios with the assumption of independence or with unstructured correlation had a higher RMSE. This means that the Kronecker specification works well to capture the

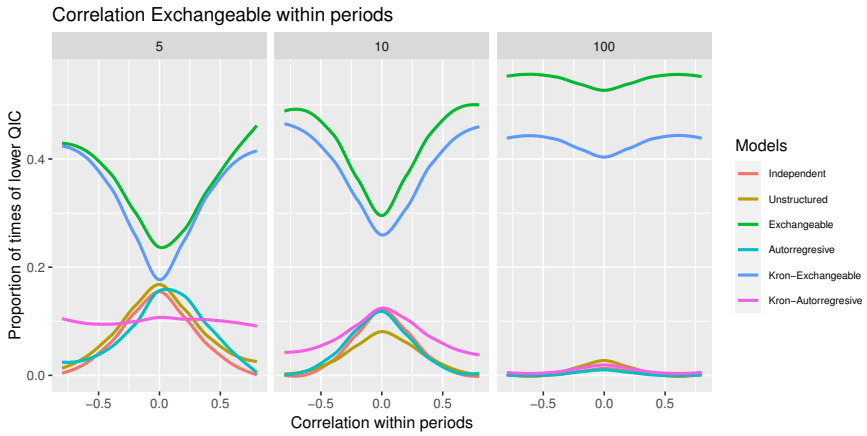

Figure 3-10.: Proportion of times the model obtains the lowest QIC relative to the value of Ψ, the value of the correlation between periods when the correlation within periods is Exchangeable. The proportion was calculated over all simulations where the correlation matrix was exchangeable and Kronecker product with $\boldsymbol{\Psi} = \boldsymbol{I}_3$. Each subplot represents the number of individuals per sequence.

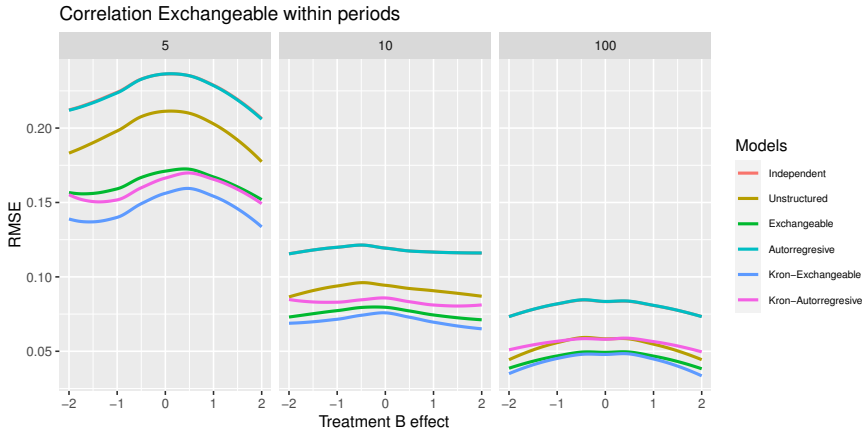

Figure 3-11.: RMSE for each value of the treatment effect. The average over all simulations where the correlation matrix was exchangeable and Kronecker product with $\boldsymbol{\Psi} = \boldsymbol{I}_3$ is shown. Each subplot represents the number of individuals per sequence.

treatment effect more accurately than the usual correlation models. In addition, since the average was being calculated over all the scenarios, it was observed that the proposed model is robust to low or high correlation between periods and within periods.

If Figure **3-12** was observed, the model with the best RMSE for the treatment was the one with the autoregressive correlation matrix, and a behavior similar to the previous one was

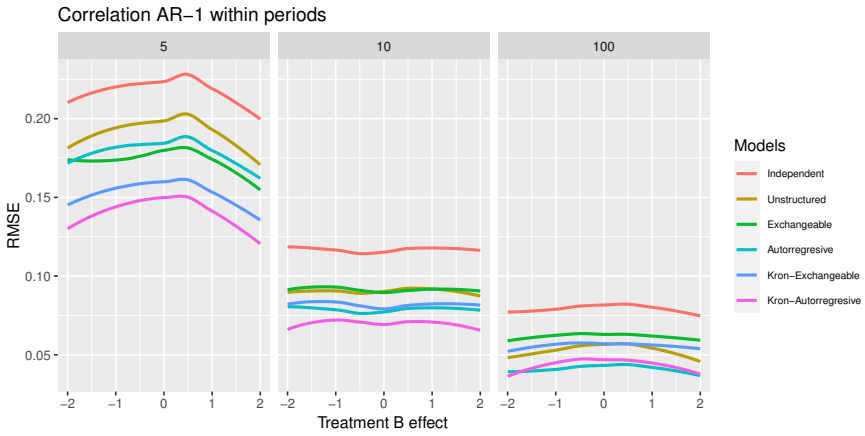

Figure 3-12.: RMSE for each value of the treatment effect. The average over all simulations where the correlation matrix was autoregresive and Kronecker product with $\boldsymbol{\Psi} = \boldsymbol{I}_3$ is shown. Each subplot represents the number of individuals per sequence.

observed but with the autoregressive correlation matrices.
That is, the methodology proposed in this paper gives results similar to the usual ones when the correct correlation matrix is $\boldsymbol{\Psi} = \boldsymbol{I}_P$, and this methodology gives better results than the usual ones when $\boldsymbol{\Psi} \neq \boldsymbol{I}_P$.

3.4.3. Without structured correlation matrix

In this section, it is assumed that the entire correlation matrix is unstructured. That is, $\boldsymbol{R}(\boldsymbol{\alpha})$ has the structure given in Equation (3-48):

$$\boldsymbol{R}(\boldsymbol{\alpha}) = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1\,12} \\ r_{12} & 1 & r_{23} & \cdots & r_{2\,12} \\ r_{13} & r_{23} & 1 & \cdots & r_{3\,12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{1\,12} & r_{2\,12} & r_{3\,12} & \cdots & 1 \end{pmatrix}, \; r_{ij} \sim U(-1,1) \text{ and } det(\boldsymbol{R}(\boldsymbol{\alpha})) > 0 \tag{3-48}$$

In Figure **3-13**, the x-axis represents the minimum value of $\boldsymbol{\alpha}$ in Equation (3-48), and the y-axis represents the percentage of times that each of the models obtains the Lower QIC compared to the other five models. Figure **3-9** shows a similar behavior between the two models proposed in this work, which are selected more times than the other models. As the number of individuals per sequence increases, in Figure **3-13**, the percentage of times that the model with exchangeable Kronecker has the lowest QIC increases.

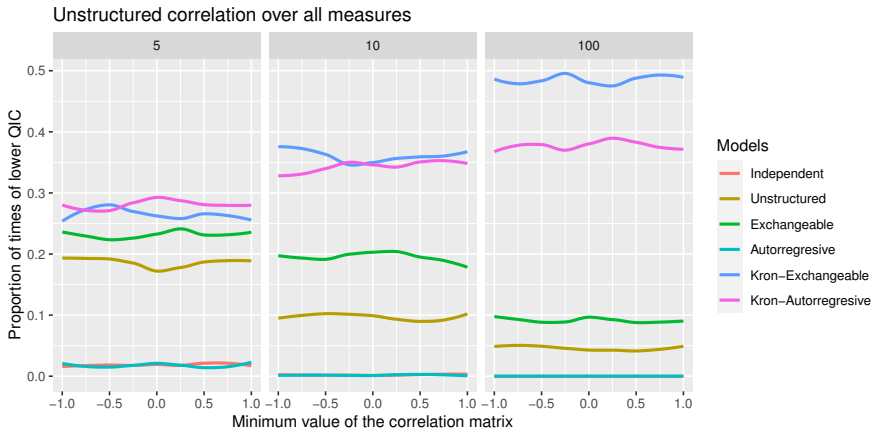

Figure 3-13.: Proportion of times the model obtains the lowest QIC. The proportion was calculated over all simulations where the correlation matrix was given in Equation (3-48). Each subplot represents the number of individuals per sequence.

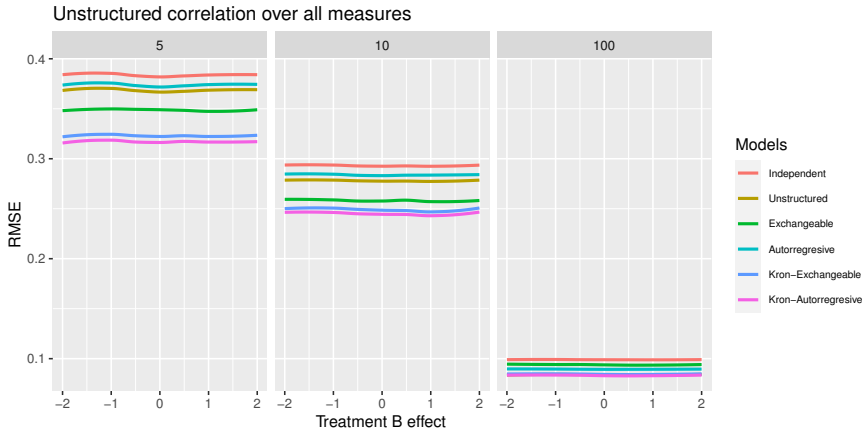

Figure 3-14.: RMSE for each value of the treatment effect. The average over all simulations where the correlation matrix was given in Equation (3-48). Each subplot represents the number of individuals per sequence.

Figure **3-14** shows the average RMSE to estimate the treatment effect in each of the models when $\boldsymbol{R}(\boldsymbol{\alpha})$ is a correlation matrix given in Equation (3-48). It was observed that as the number of individuals per sequence increases, the RMSE decreases, which is in agreement with what was expected.

However, it was also observed that the Kronecker-autoregressive or Kronecker-interchangeable correlation matrix model (the methodology proposed in this work) works better than the other 4 models when there were 5 or 10 individuals per sequence. On the other hand, when

the number of individuals increases to 100, the 6 models had very similar RMSEs, with a slight advantage of the two models proposed here and the unstructured correlation model. In all scenarios with the assumption of independence or with unstructured correlation, the estimator of treatment effect had a higher RMSE. This means that the Kronecker specification works well to capture the treatment effect more accurately than the usual correlation models. In addition, since the average was being calculated over all the scenarios, it was observed that the proposed model is robust to low or high correlation between periods and within periods. That is, the methodology proposed in this paper gives gives better results than the usual ones when $\boldsymbol{R}(\boldsymbol{\alpha})$ was a correlation matrix given in Equation (3-48).

Models	5 individuals	10 individuals	100 individuals
Independent	0.0054	0.0088	0.0325
Unstructured	0.0192	0.0275	1.3306
Exchangeable	0.0100	0.0194	0.0669
Autorregresive	0.0062	0.0094	0.0556
Kron-Exchangeable	0.2692	0.5181	9.0894
Kron-Autorregresive	0.2623	0.5106	8.7969

Table 3-4.: Average execution time of each model according to the number of individuals per sequence.

Finally, the average execution time of each model is presented in Table **3-4**, evidencing that the proposed model is more computationally expensive than the usual models, even when compared to an unstructured correlation matrix. R software R Core Team (2022) and an adaptation of the gee Carey. (2019) and geem McDaniel et al. (2013) libraries were used. In Supplementary File 1, the same simulation exercise was performed, but assuming that $\boldsymbol{\Psi} = \boldsymbol{I}_3$. There, it can be seen that the methodology proposed in this work gave similar results as compared to the usual models when the correct correlation matrix is $\boldsymbol{\Psi} = \boldsymbol{I}_3$, while it gave better results when $\boldsymbol{\Psi} \neq \boldsymbol{I}_3$ or even when $\boldsymbol{R}(\boldsymbol{\alpha})$ was totally unstructured.

3.5. Application

The systolic blood pressure data is presented in Figure **3-15**. The model used to analyze this experiment had a linear predictor considering fixed effects of treatment, period, baseline (the two measurements taken before applying the treatment), first and second order linear and quadratic carry-over effects as a function of time. On the other hand, the model use to analyze data from the second experiment had a linear predictor considering fixed effects of baseline, treatment, period, and first order linear carry-over effects as a function of time (quadratic effects were not considered because there were three observations within each period). For

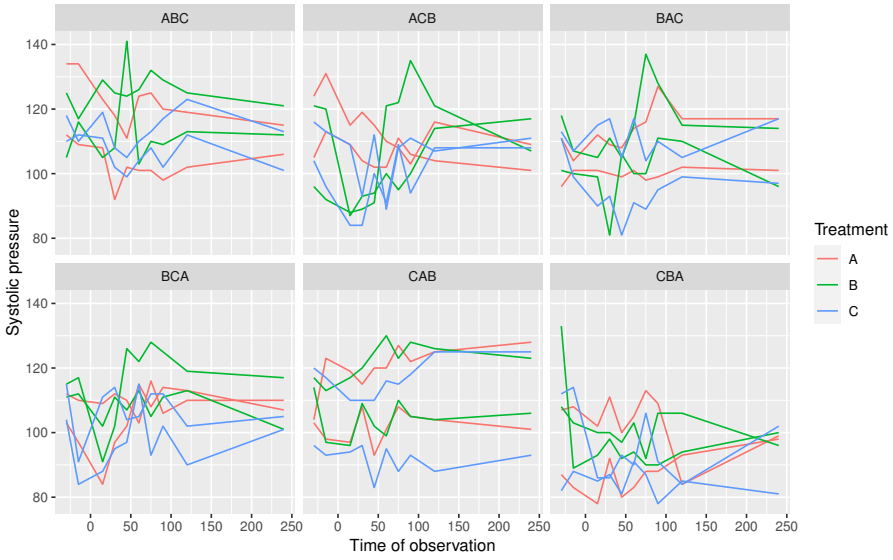

Figure 3-15.: Blood systolic pressure (mmHg), the time of observation is presented in minutes from the application

the third experiment, the logit model used to analyze this experiment had a linear predictor considering fixed effects of treatment, first and second order linear and quadratic carry-over and period effects as a function of time. The structures of the working correlation matrices were: i) independence, ii) first order autoregressive, and iii) fitted individually and were compared via the QIC to determine the one exhibiting the best fit to each dataset. Table 3-5 presents the QIC for each correlation structure in the three datasets.

According to Table 3-5, the Kronocker and exchangeable correlation matrix had the smallest

Matriz $R(\alpha)$	QIC arterial	QIC cows	QIC occupation
$\boldsymbol{I}_{PL}$	46086	20.21	419.98
$\boldsymbol{I}_P \otimes AR(1)_L$	44166	18.06	418.62
$\boldsymbol{I}_P \otimes Exch_L$	45059	17.98	419.32
$AR(1)_{PL}$	44923	20.46	418.72
$Exch_{PL}$	44158	20.21	418.90
$\boldsymbol{\Psi} \otimes AR(1)_L$	43400	14.68	418.61
$\boldsymbol{\Psi} \otimes Exch_L$	43393	15.06	419.33

Table 3-5.: Correlation matrices and the corresponding QIC values

QIC value for the arterial pressure, and the Kronocker and autoregressive correlation matrix for the dairy cattle experiments. Hence, correlation matrices having the Kronecker structure had the best fit in the two datasets. The correlation matrices estimated using equations (3-7)

and (3-9) for the arterial pressure data are, respectively:

$$\hat{\boldsymbol{\Psi}} = \begin{pmatrix} 1.0000 & 0.0537 & 0.4486 \\ 0.0537 & 1.0000 & 0.3756 \\ 0.4486 & 0.3756 & 1.0000 \end{pmatrix} \tag{3-49}$$

$$\hat{\boldsymbol{R}}_1(\boldsymbol{\alpha}) = \begin{pmatrix} 1.0000 & 0.4958 & 0.4958 & \cdots & 0.4958 \\ 0.4958 & 1.0000 & 0.4958 & \cdots & 0.4958 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.4958 & 0.4958 & 0.4958 & \cdots & 1.0000 \end{pmatrix} \tag{3-50}$$

and for the dairy cattle dataset:

$$\hat{\boldsymbol{\Psi}} = \begin{pmatrix} 1.0000 & 0.3650 \\ 0.3650 & 1.0000 \end{pmatrix} \tag{3-51}$$

$$\hat{\boldsymbol{R}}_1(\boldsymbol{\alpha}) = \begin{pmatrix} 1.0000 & 0.3362 & 0.3362^2 \\ 0.3362 & 1.0000 & 0.3362 \\ 0.3362^2 & 0.3362 & 1.0000 \end{pmatrix} \tag{3-52}$$

Regarding the occupancy data, Table **3-5** shows that although the lowest QIC is obtained in the Kronocker and autoregressive matrix, it is very close to considering only the autoregressive within periods, and if we observe the estimate of $\boldsymbol{\Psi}$ for this data set in the following equation, it is noted that it is very close to the identity matrix:

$$\hat{\boldsymbol{\Psi}} = \begin{pmatrix} 1.0000 & -0.0191 & -0.0050 & 0.0405 \\ -0.0191 & 1.0000 & 0.0271 & 0.0588 \\ -0.0050 & 0.0271 & 1.0000 & -0.0145 \\ 0.00405 & 0.0588 & -0.0145 & 1.000 \end{pmatrix} \tag{3-53}$$

$$\hat{\boldsymbol{R}}_1(\boldsymbol{\alpha}) = \begin{pmatrix} 1.0000 & 0.7532 & 0.7532^2 & \cdots & 0.7532^8 \\ 0.7532 & 1.0000 & 0.7532 & \cdots & 0.7532^7 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.7532^8 & 0.7532^7 & 0.7532^6 & \cdots & 1.0000 \end{pmatrix} \tag{3-54}$$

By computing the matrix as the Kronecker product of matrices in equations (3-49) and (3-50), notice the positive correlation between periods 1 and 3, a small positive correlation between periods 1 and 2. On the other hand, the matrix obtained as the Kronecker product of matrices in equations (3-53) and (3-54) had a positive but small correlation between periods 1 and 2.

The matrices with the Kronecker structure are used to estimate the location parameters of the linear model (since these had the smallest QIC) yielding Table **3-6** for the arterial pressure data, Table **3-7** for the dairy cattle data and Table **3-8** for the occupation data. Each

Parameter	Estimate	Robust SE	Robust z	p-value
Intercept	105.1764	3.4255	30.7037	<0.001
Base	9.4286	2.2744	4.1455	<0.001
Time	0.0961	0.0364	2.6385	0.0042
Period 2	5.0793	3.5378	-1.4357	0.0755
Period 3	11.4009	7.4704	-1.5261	0.0635
Treatment B	-1.1014	1.8655	-0.5904	0.2775
Treatment C	-10.9991	3.2589	-3.3751	0.0004
Carry-over B	-4.5981	5.4023	0.8511	0.1973
Carry-over C	-11.5899	5.6351	2.0567	0.0199
Time^2	-0.0003	0.0001	-2.3884	0.0085

Table 3-6.: Estimates obtained by GEE for arterial pressure data

table shows the estimates, their standard errors (computed using the "sandwich" variance estimator (Hardin and Hilbe, 2003)), the z statistic and the p-value corresponding to the null hypothesis $\beta = 0$ Table **3-6** shows significant effects of baseline, treatment C (with respect to

Parameter	Estimate	Robust SE	Robust z	p-value
Intercept	1.9911	0.4579	18.9045	<0.001
Base	0.3707	0.1719	4.6515	0.0310
Time	-0.0086	0.0068	1.6093	0.2046
Period 2	0.2517	0.0565	19.8628	<0.001
Treatment A	-0.1438	0.0494	8.4638	0.0036
Carry-over A	-4.2480	1.4128	-3.0069	0.0026
Time^2	0.0001	0.0001	1.0223	0.3120

Table 3-7.: Estimates obtained by GEE for dairy cattle data

A), time (linear and quadratic), period, and carry-over effect of treatment C. The interaction between carry-over effects and time was not significant, which means that the effect of the previous treatment remains the same during the rest of the experiment. Moreover, Table **3-7** shows significant effects of baseline (milk protein yield before the beginning of the experiment), the treatment in the same application period and a significant carry-over effect of the paper-based diet over the grass-based diet, it follows that the treatment effect is evident in the medium term, i.e., the cows being fed paper took a time to recover after changing to the grass-based diet. There is no significant effect of linear (negative) and quadratic (positive) regression coefficients of time.

For the occupancy data, a logit link with a binomial response was used, having as response

	Estimate	Robust S.E.	Robust z	pvalor
Intercept	-6.7465	0.6062	-11.1288	0.0000
areaB	-0.0673	0.0783	-0.8594	0.1951
areaC	0.2049	0.0729	2.8115	0.0025
areaD	-0.0041	0.0766	-0.0539	0.4785
wave	0.8954	0.3892	2.3010	0.0107
wave^2	-0.1606	0.0643	-2.4964	0.0063
hours	0.7170	0.0325	22.0927	0.0000
hours^2	-0.0269	0.0012	-22.1740	0.0000
Carry-over B	0.2208	0.0916	2.4110	0.0080
Carry-over C	0.1311	0.0949	1.3821	0.0835
Carry-over D	0.1037	0.0946	1.0961	0.1365
Carry-over None	0.6177	0.2477	2.4941	0.0063

Table 3-8.: Estimates obtained by GEE for occupation data

the number of times the chair was occupied and how many failures when the chair was empty. In Table **3-8**, a significant effect of the nest treatment is observed, as well as a significant linear and quadratic effect of the waves, also of the time of measurement and there are significant carry-over effects.

4. Semi-parametric generalized estimating equations for repeated measurements in crossover designs

The content of this chapter was published as a research article in the journal Statistical Methods in Medical Research (Cruz et al., 2023c). [1]

4.1. Introduction

In the context of crossover experimental designs, each experimental unit receives a sequence of treatments, and each treatment is applied over a period of time (Biabani et al., 2018). These designs are very useful in medical experimentation, since they require fewer experimental units to obtain the same results as cross-sectional studies. The disadvantage is given by the appearance of carry-over effects, which are defined as the residual effects that remain in the response of the individual and that are caused by the treatments applied in the previous periods (Madeyski and Kitchenham (2018) and Kitchenham et al. (2018)).

The most recent works on the analysis of crossover designs assume the non-existence of carry-over effects, due to the presence of a washout period between the successive applications of treatments (Curtin, 2017). This assumption is common in works based on classical generalized linear models (Li et al., 2018), Bayesian models (Oh et al., 2003) or generalized estimating equation (GEE) models (Curtin, 2017). However, in some crossover designs the length of the washout period is very short and does not guarantee the elimination of the residual effects of each of the treatments, such as the one presented in Jones and Kenward (2015, page 204). In this design, three treatments for blood pressure control are used and treatment C is a placebo. In this experiment, if there is a carry-over effect of the placebo, it is not cleaned in the washout period. The doctors tried to control the hypertension in patients, and so, treatments can be stopped for a very short time due to the characteristics of the disease.

Furthermore, in the design presented in Jones and Kenward (2015, pag 204), the systolic

[1] https://doi.org/10.1177/09622802231158736

blood pressure is observed ten times within each period: 30 and 15 minutes before the application, and 15, 30, 45, 60, 75, 90 , 120 and 240 minutes after the application, as seen in the Table **4-1**, which generates a repeated measurement for each application period. This type of design is known as a repeated measures crossover design (Dubois et al., 2011). Dubois et al. (2011), Diaz et al. (2013) and Forbes et al. (2015) used Gaussian linear mixed models to study crossover repeated measures designs. However, those studies considered one observation per period by calculating the area under the curve, and they did not include the carry-over effects. Additionally, this modeling does not allow us to observe the temporal behavior of the response variable within the period, nor the presence of carry-over effects that fluctuate over time.

On the other hand, when the response variable of the crossover experiment shown by Kenward and Roger (2009) is analyzed, it is observed that it does not adequately fit a normal distribution. Since the response in this experiment is blood sugar levels, which are skewed and always positive, a gamma distribution seems more suitable for analysis. In both experiments, we have a response that can be assumed to be in the exponential family and the responses of the same experimental unit are correlated. Therefore, in this chapter we propose an extension of the GEE with splines to model the effects of main interest in the design (treatments, period) with a parametric component and the temporary effects through smoothing splines.

This methodology makes it possible to unbiasedly isolate the temporal behavior of the carry-over effect from the period and treatment effects, which is demonstrated theoretically and through a simulation exercise. Subsequently, when applying the methodology in the blood pressure design, a significant carry-over effect of the placebo treatment is obtained, corroborating the importance of taking it into account in the analysis.

Sequence		Period 1	Period 2	Period 3
(1) ABC	Ind 1	10 observations	10 observations	10 observations
	Ind 2	10 observations	10 observations	10 observations
⋮		⋮	⋮	⋮
(6) CBA	Ind 11	10 observations	10 observations	10 observations
	Ind 12	10 observations	10 observations	10 observations

Table 4-1.: Structure of the blood pressure crossover design, with three period, six sequences and ten measurements (that were taken -30, -15, 15, 30, 45, 60, 75, 90, 120 and 240 minutes from application) per period

This chapter is structured as follows: In Section 2, the semiparametric model with GEE is described, and the estimation equations are derived. In section 3, asymptotic consistency and unbiasedness of estimators are established. In Section 4, a simulation study is carried

out to display the advantages of the proposed model over those models often found in the literature and some diagnostics measures for its residuals. In Section 5, an application of to blood pressure data is performed out to illustrate the model properties and to carry out an overall analysis of this dataset. Finally, some conclusions are presented in Section 6.

4.2. Repeated measures crossover design

A crossover design entails five components (Jones and Kenward, 2015): i) sequences which are randomly assigned combinations of the applied treatments on the experimental units, ii) treatments, that are applied to each experimental unit as a part of a sequence in a given time, iii) periods, that represent the application lapse for the treatments which are part of a sequence, v) experimental units, which are the elements on which a treatment is applied. In each sequence, there are n_l experimental units, therefore the total number of experimental units is $n = \sum_{l=1}^{S} n_l$. Further, it is frequent that each period has the same length for all sequences; therefore, the number of observation periods equals the length of each sequence. For the structure of crossover designs, the carry-over effects constitute part of them. Vegas et al. (2016) defined the carry-over as a treatment's effect persistence over those treatments applied later. That is, if a treatment is applied on a given period, then there exists the possibility of a residual or carry-over effect that persists in the following periods when other treatments are applied. When the carry-over effect of a treatment affects the one applied in the next period, it is known as a frist-order carry over effect.
In a crossover design with S sequences of length P, let n_{ij} be the number of observations on the i-th experimental unit and j-th period, then $\boldsymbol{Y}_{ij}$ is a vector defined as:

$$\boldsymbol{Y}_{ij} = \left(Y_{ij1}, \ldots, Y_{ijn_{ij}}\right)^T \tag{4-1}$$

Moreover, we define a vector $\boldsymbol{Y}_i$ vector contains every observation on the i-th experimental unit

$$\boldsymbol{Y}_i = \left(\boldsymbol{Y}_{i1}, \ldots, \boldsymbol{Y}_{iP}\right)^T \tag{4-2}$$

and its size is $\sum_{j=1}^{P} n_{ij}$.
Regarding the use of smoothing functions, Wild and Yee (1996) proposed a kernel smoothing to select explanatory variables in GEE models, Lin and Carroll (2001) derived a semiparametric estimation equation for repeated measures data and presented some asymptotic properties without including the correlation matrix. On the other hand, He et al. (2002) presented a semiparametric model with correlated normal data and explored the properties of symmetric kernels, Stoklosa and Warton (2018) developed a GEE generalization for adaptive multivariate splines through m-estimators, and Yang and Niu (2021) discussed a GEE model with two semiparametric functions for normally distributed responses and some kernel smoothing functions.

Accordingly, GEE will be used because Y_{ijk} (the response variable) has a distribution that belongs to the exponential family, and also a semiparametric model with B-splines for the time and carry-over effects as follows:

$$E(Y_{ijk}) = \mu_{ijk}, \qquad Var(Y_{ijk}) = \phi V(\mu_{ijk})$$

$$g(\mu_{ijk}) = \boldsymbol{x}_{ijk}^T\boldsymbol{\beta} + f(\boldsymbol{Z}_{ijk}) + \sum_{c=1}^{C} f_c(\boldsymbol{Z}_{ijk}) \tag{4-3}$$

$$\boldsymbol{V}(\boldsymbol{\mu}_i) = \left[\boldsymbol{D}(V(\mu_{ijk})^{\frac{1}{2}})\boldsymbol{R}(\boldsymbol{\alpha})\boldsymbol{D}(V(\mu_{ijk})^{\frac{1}{2}})\right]_{\sum_{j=1}^{P} n_{ij} \times \sum_{j=1}^{P} n_{ij}} \tag{4-4}$$

where $g(\cdot)$ is the link function associated to the exponential family, $\boldsymbol{x}_{ijk}$ is the vector of the design matrix associated to the k-th response of the i-th experimental unit in the j-th period, $\boldsymbol{\beta}$ represents the parametric effects, ϕ is the dispersion parameter, $\boldsymbol{Z}_{ijk}$ is the time at which the k-th observation of the i-th experimental unit was measured in the j-th period. For example in the blood pressure data set described in Table 4-1, $\boldsymbol{Z}_{i11} = -30$, $\boldsymbol{Z}_{i12} = -15, \ldots, \boldsymbol{Z}_{i1\,10} = 240$, that is, the minutes of measurement within each period, f is a function describing the time's effect period, f_c is a function describing the previous treatment carry-over effect on the current period (with $f_c(\boldsymbol{Z}_{i1k}) = 0$ because it is assumed that there are no carry-over effects in period 1), C is the total number of carry-over effects, for example in the design described in Table 4-1, $C = 2$, that is, f_1 is the carry-over effect of B and f_2 is the carry-over effect of C, $V(\mu_{ijk})$ is the variance function related to the exponential family, and $R(\boldsymbol{\alpha})$ is the associated correlation matrix.
Let $\{s_1(t), \ldots, s_m(t)\}$ be a basis splines, then the f_1 and f_2 functions can be approximated through the following equations (Yu and Peace, 2012)

$$\hat{f}(t) = \sum_{b=1}^{m} \hat{\alpha}_{1b} s_b(t) \tag{4-5}$$

$$\hat{f}_c(t) = \sum_{b=1}^{m} \hat{\alpha}_{2b} s_b(t) \tag{4-6}$$

where $m = max(n_{ij})$. Adapting the estimation equations given by He et al. (2002), the following generalized estimation equations are proposed for $\alpha_1 = \{\alpha_{11}, \ldots, \alpha_{1m}\}$, $\alpha_2 = \{\alpha_{21}, \ldots, \alpha_{2m}\}$ and $\boldsymbol{\beta}$:

- For the time effect

$$U_1(\boldsymbol{\alpha_1}, t|\boldsymbol{\beta}, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \left\{ diag\left(\frac{\partial \mu_{ijk}}{\partial \boldsymbol{\alpha}_1}\right)\right\}_i \frac{\boldsymbol{V}_{1i}^{-1}}{\phi} \left(\boldsymbol{y}_i - \boldsymbol{\mu}_i \left[\boldsymbol{X}_i\boldsymbol{\beta}, \sum_{b=1}^{m} \alpha_{1b} s_b(t), \hat{f}_c(\boldsymbol{Z}_i)\right]\right) \tag{4-7}$$

where

$$\boldsymbol{V}_{1i} = \left\{ diag\left(V\left(\boldsymbol{\mu}_i \left[\boldsymbol{X}_i\boldsymbol{\beta}, \sum_{b=1}^{m} \alpha_{1b} s_b(t), \hat{f}_c(\boldsymbol{Z}_i) \right] \right) \right) \right\}_i^{\frac{1}{2}} \times \boldsymbol{R}(\boldsymbol{\alpha}) \times$$
$$\left\{ diag\left(V\left(\boldsymbol{\mu}_i \left[\boldsymbol{X}_i\boldsymbol{\beta}, \sum_{b=1}^{m} \alpha_{1b} s_b(t), \hat{f}_c(\boldsymbol{Z}_i) \right] \right) \right) \right\}_i^{\frac{1}{2}}, \qquad i = 1, \ldots, n,$$

$\left\{ diag\left(\frac{\partial \mu_{ijk}}{\partial \boldsymbol{\alpha}_1} \right) \right\}_i$ is a diagonal matrix with elements on the diagonal given by

$$\left\{ \frac{\partial \mu_{i11}}{\partial \boldsymbol{\alpha}_1}, \frac{\partial \mu_{i12}}{\partial \boldsymbol{\alpha}_1}, \ldots, \frac{\partial \mu_{i1n_{i1}}}{\partial \boldsymbol{\alpha}_1}, \ldots, \frac{\partial \mu_{iPn_{iP}}}{\partial \boldsymbol{\alpha}_1} \right\}$$

and $V(\cdot)$ is the variance function of the exponential family applied to each of the i-th individual's expected values.

- For the carry-over effects

$$U_2(\boldsymbol{\alpha_2}, t|\boldsymbol{\beta}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}) =$$
$$\sum_{i=1}^{n} \left\{ diag\left(\frac{\partial \mu_{ijk}}{\partial \boldsymbol{\alpha}_2} \right) \right\}_i \frac{\boldsymbol{V}_{2i}^{-1}}{\phi} \left(\boldsymbol{y}_i - \boldsymbol{\mu}_i \left(\boldsymbol{X}_i\boldsymbol{\beta}, \sum_{b=1}^{m} \alpha_{2b} s_b(t), \hat{f}(\boldsymbol{Z}_{1i}) \right) \right) \tag{4-8}$$

where

$$\boldsymbol{V}_{2i} = \left\{ diag\left(V\left(\boldsymbol{\mu}_i \left[\boldsymbol{X}_i\boldsymbol{\beta}, \sum_{b=1}^{m} \alpha_{2b} s_b(t), \hat{f}(\boldsymbol{Z}_{1i}) \right] \right) \right) \right\}_i^{\frac{1}{2}} \times \boldsymbol{R}(\boldsymbol{\alpha}) \times$$
$$\left\{ diag\left(V\left(\boldsymbol{\mu}_i \left[\boldsymbol{X}_i\boldsymbol{\beta}, \sum_{b=1}^{m} \alpha_{2b} s_b(t), \hat{f}(\boldsymbol{Z}_i) \right] \right) \right) \right\}_i^{\frac{1}{2}}, \qquad i = 1, \ldots, n.$$

- For the fixed effects, that is, treatment, sequence, period or other covariates

$$U_3(\boldsymbol{\beta}|\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}) =$$
$$\sum_{i=1}^{n} \left\{ diag\left(\frac{\partial \mu_{ijk}}{\partial \boldsymbol{\beta}} \right) \right\}_i \frac{\boldsymbol{V}_{3i}^{-1}}{\phi} \left(\boldsymbol{y}_i - \boldsymbol{\mu}_i \left[\boldsymbol{X}_i\boldsymbol{\beta}, \hat{f}(\boldsymbol{Z}_i, \hat{f}_c(\boldsymbol{Z}_i) \right] \right) \tag{4-9}$$

where

$$\boldsymbol{V}_{3i} = \left\{ diag\left(V\left(\boldsymbol{\mu}_i \left[\boldsymbol{X}_i\boldsymbol{\beta}, \hat{f}(\boldsymbol{Z}_i), \hat{f}_c(\boldsymbol{Z}_i) \right] \right) \right) \right\}_i^{\frac{1}{2}} \times \boldsymbol{R}(\boldsymbol{\alpha}) \times$$
$$\left\{ diag\left(V\left(\boldsymbol{\mu}_i \left[\boldsymbol{X}_i\boldsymbol{\beta}, \hat{f}(\boldsymbol{Z}_i), \hat{f}_c(\boldsymbol{Z}_i) \right] \right) \right) \right\}_i^{\frac{1}{2}}, \qquad i = 1, \ldots, n.$$

- For the correlation matrix

$$U_4(\boldsymbol{\alpha}|\boldsymbol{\beta},\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2)=\sum_{i=1}^{S}\sum_{k=1}^{n_i}\left(\frac{\partial\boldsymbol{\varepsilon}_{ik}}{\partial\boldsymbol{\alpha}}\right)^T\boldsymbol{F}_{ik}^{-1}\left(\boldsymbol{W}_{ik}-\boldsymbol{\varepsilon}_{ik}\right) \tag{4-10}$$

  where $\boldsymbol{F}_{ik}=\boldsymbol{D}(V(r_{ijk}))_{q\times q}$ is a diagonal matrix, $\boldsymbol{\varepsilon}_{ik}=E(\boldsymbol{W}_{ik})_{q\times 1}$ and $\boldsymbol{W}_{ik}=(r_{i1k}r_{i2k}, r_{i1k}r_{i3k},\ldots,r_{i(T-1)k}r_{iTk})_{q\times 1}^T$, r_{ijk} is the ijk-th Pearson residual and $q=\binom{T}{2}$.

- For the estimation of ϕ, the following equation is used:

$$\hat{\phi}=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{P}\sum_{k=1}^{n_{ij}}r_{ijk}^2$$

To get the estimators of $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, the following steps are performed:

1. Set initial values $\boldsymbol{\alpha}_2^{(0)}$, $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\alpha}^{(0)}$.
2. Find the value $\boldsymbol{\alpha}_1^{(1)}$ that solves the equation
$$U_1(\boldsymbol{\alpha_1},t|\boldsymbol{\beta}^{(0)},\boldsymbol{\alpha}_2^{(0)},\boldsymbol{\alpha}^{(0)})=0.$$
3. Find the value $\boldsymbol{\alpha}_2^{(1)}$ that solves the equation
$$U_2(\boldsymbol{\alpha_2},t|\boldsymbol{\beta}^{(0)},\boldsymbol{\alpha}_1^{(1)},\boldsymbol{\alpha}^{(0)})=0.$$
4. Find the value $\boldsymbol{\beta}^{(1)}$ that solves the equation
$$U_3(\boldsymbol{\beta}|\boldsymbol{\alpha}_1^{(1)},\boldsymbol{\alpha}_2^{(1)},\boldsymbol{\alpha}^{(0)})=0.$$
5. Find the value $\boldsymbol{\alpha}^{(1)}$ that solves the equation
$$U_4(\boldsymbol{\alpha}|\boldsymbol{\beta}^{(1)},\boldsymbol{\alpha}_1^{(1)},\boldsymbol{\alpha}_2^{(1)},\boldsymbol{\alpha}^{(0)})=0.$$
6. Repeat steps (2) to (5) until convergence.

To solve Equation (4-9), the Fisher scoring algorithm is used, that is, in the w-th step, the estimator of $\boldsymbol{\beta}$ is given by:

$$\begin{aligned}\boldsymbol{\beta}^{(w+1)}&=\boldsymbol{\beta}^{(w)}-\left[E\left\{U_3'(\boldsymbol{\beta}^{(w)}|\boldsymbol{\alpha}_1^{(w)},\boldsymbol{\alpha}_2^{(w)},\boldsymbol{\alpha}^{(w)})\right\}\right]^{-1}U_3(\boldsymbol{\beta}^{(w)}|\boldsymbol{\alpha}_1^{(w)},\boldsymbol{\alpha}_2^{(w)},\boldsymbol{\alpha}^{(w)})\\&=\boldsymbol{\beta}^{(w)}-\left\{\sum_{i=1}^{n}\boldsymbol{X}_i^T\boldsymbol{W}_i^{(w)}\boldsymbol{X}_i\right\}^{-1}\left\{\sum_{i=1}^{n}\boldsymbol{X}_i^T\boldsymbol{W}_i^{(w)}\left(\boldsymbol{N}_i^{(w)}\right)^{-1}\boldsymbol{u}_i^{(w)}\right\}\end{aligned} \tag{4-11}$$

where

$$E\left\{U_3'(\boldsymbol{\beta}^{(w)}|\boldsymbol{\alpha}_1^{(w)},\boldsymbol{\alpha}_2^{(w)},\boldsymbol{\alpha}^{(w)})\right\} = E\left\{\frac{\partial U_3(\boldsymbol{\beta}^{(w)}|\boldsymbol{\alpha}_1^{(w)},\boldsymbol{\alpha}_2^{(w)},\boldsymbol{\alpha}^{(w)})}{\partial\boldsymbol{\beta}^{(w)}}\right\} = \sum_{i=1}^{n}\boldsymbol{X}_i^T\boldsymbol{W}_i^{(w)}\boldsymbol{X}_i,$$

$$\boldsymbol{W}_i^{(w)} = \left\{diag\left(\frac{\partial\mu_{ijk}^{(w)}}{\partial\boldsymbol{\beta}^{(w)}}\right)\right\}_i \boldsymbol{V}_{3i}^{-1}\left\{diag\left(\frac{\partial\mu_{ijk}^{(w)}}{\partial\boldsymbol{\beta}^{(w)}}\right)\right\}_i^{-1},$$

$$\boldsymbol{N}_i^{(w)} = \left\{diag\left(\frac{\partial\mu_{ijk}^{(w)}}{\partial\boldsymbol{\beta}^{(w)}}\right)\right\}_i,$$

$$\boldsymbol{u}_i^{(w)} = \left(\boldsymbol{y}_i - \boldsymbol{\mu}_i^{(w)}\left[\boldsymbol{X}_i\boldsymbol{\beta}^{(w)}, \hat{f}(\boldsymbol{Z}_i^{(w)}, \hat{f}_c(\boldsymbol{Z}_i^{(w)})\right]\right).$$

Carrying out procedure similar to Tsuyuguchi et al. (2020), Equation (4-11) can be written as:

$$\boldsymbol{\beta}^{(w+1)} = \left\{\sum_{i=1}^{n}\boldsymbol{X}_i^T\boldsymbol{W}_i^{(w)}\boldsymbol{X}_i\right\}^{-1}\left\{\sum_{i=1}^{n}\boldsymbol{X}_i^T\boldsymbol{W}_i^{(w)}\boldsymbol{z}_i^{(w)}\right\} \tag{4-12}$$

where

$$\boldsymbol{z}_i^{(w)} = \boldsymbol{X}_i\boldsymbol{\beta}^{(w)} - \boldsymbol{N}_i^{-1(m)}\boldsymbol{u}_i^{(w)}, \qquad i = 1,\ldots,n \tag{4-13}$$

Therefore, $\hat{\boldsymbol{\beta}}$ is obtained analogously to a weighted least squares solution on the transformed response variable $\boldsymbol{z}_i$, where the effects of the variables associated to time and the carry-over effect have been removed. With these considerations, the asymptotic theory of estimators is developed using the following theorem:

Theorem 4.2.1. *Under the assumption that the r-th derivative of f_1 and f_2 is bounded for some $r \geq 2$ and that the number of knots $m = m_n \to \infty$, but $\frac{m}{n} \to 0$ then $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \xrightarrow{n\to\infty} \mathbf{0}$. Also, if $m = O\left(n^{\frac{1}{(2r+1)}}\right)$ then:*

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\sum_{b=1}^{m}\hat{\alpha}_{1b}s_b(\boldsymbol{Z}_{1ijk}) - f_1(\boldsymbol{Z}_{1ijk})\right\}^2 = O\left(n^{-\frac{2r}{(2r+1)}}\right)$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\sum_{b=1}^{m}\hat{\alpha}_{1b}s_b(\boldsymbol{Z}_{2ijk}) - f_2(\boldsymbol{Z}_{2ijk})\right\}^2 = O\left(n^{-\frac{2r}{(2r+1)}}\right)$$

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})$$

where

$$\boldsymbol{A} = \sum_{i=1}^{n}\boldsymbol{N}_i\boldsymbol{V}_{3i}^{-1}\boldsymbol{N}_i^T \tag{4-14}$$

$$\boldsymbol{B} = \sum_{i=1}^{n}\boldsymbol{N}_i\boldsymbol{V}_{3i}^{-1}\left(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i\right)\left(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i\right)^T\boldsymbol{V}_{3i}^{-1}\boldsymbol{N}_i^T \tag{4-15}$$

Proof. Let

$$\boldsymbol{\theta}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2) = \begin{pmatrix} \boldsymbol{B}_n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ \sqrt{n}\boldsymbol{H}_{1n}(\hat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1) + \sqrt{n}\boldsymbol{H}_{1n}\boldsymbol{W}_1^T\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ \sqrt{n}\boldsymbol{H}_{2n}(\hat{\boldsymbol{\alpha}}_2 - \boldsymbol{\alpha}_2) + \sqrt{n}\boldsymbol{H}_{2n}\boldsymbol{W}_2^T\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \end{pmatrix} \tag{4-16}$$

with $\boldsymbol{B}$ as in Equation (4-15), and

$$\begin{aligned} \boldsymbol{W}_1 &= (\boldsymbol{\pi}_{11}, \ldots, \boldsymbol{\pi}_{1n}), \ \boldsymbol{\pi}_{1i} = (\pi(\boldsymbol{Z}_{1i11}), \ldots, \pi(\boldsymbol{Z}_{1ijn_{ij}})) \\ \boldsymbol{W}_2 &= (\boldsymbol{\pi}_{21}, \ldots, \boldsymbol{\pi}_{2n}), \ \boldsymbol{\pi}_{2i} = (\pi(\boldsymbol{Z}_{2i11}), \ldots, \pi(\boldsymbol{Z}_{2ijn_{ij}})) \\ \boldsymbol{H}_{1n} &= n\boldsymbol{W}_1^T\boldsymbol{W}_1, \ \boldsymbol{H}_{2n} = n\boldsymbol{W}_2^T\boldsymbol{W}_2 \\ \boldsymbol{X} &= (\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_n^T), \ \pi(t) = \{s_1(t), \ldots, s_m(t)\} \end{aligned} \tag{4-17}$$

Following ideas presented in Speckman (1988) to guarantee that both $\boldsymbol{X}$ and $\boldsymbol{Z}_{ij}$ have finite second moments, we assume that there exists a random variables, δ_{ijk}, with $E(\delta_{ijk}) = 0$ and $Var(\delta_{ijk}) \leq \infty$ and continuous functions $g_1, \ldots, g_m$ such that:

$$x_{ijkl} = g_l(\boldsymbol{Z}_{ijk}) + \delta_{ijkl} \qquad 1 \leq i \leq n, \ 1 \leq j \leq P, \ 1 \leq k \leq n_{ij}, \ 1 \leq l \leq dim\,(\boldsymbol{\beta}) \tag{4-18}$$

These functions allow modeling the possible relationship between the vector of variables associated to the parametric effects and the measurement times within each period. Let $\boldsymbol{X}_{ij} = (x_{ij1}, \ldots, x_{ijn_{ij}})$ be the parametric effects design matrix, then the following properties hold:

i) The succession $\{n_{ij}\}$ is bounded for all $1 \leq i \leq n$ and $1 \leq j \leq P$, that is:

$$max(n_{ij}) < \infty$$

ii) Since Y_{ijk} is a random variable that belongs to the exponential family and due to the definition of the generalized estimation equations in Liang and Zeger (1986), and by Lemmma 5.3 given in Lehmann and Casella (2006, pag 116), then

$$E(\boldsymbol{u}_{1i}) = E\left(\boldsymbol{y}_i - \boldsymbol{\mu}_i\left[\boldsymbol{X}_i\boldsymbol{\beta}, \sum_{b=1}^{m}\alpha_{1b}s_b(t), \hat{f}_2(\boldsymbol{Z}_{2i})\right]\right) = 0$$

Therefore, the expected value of (4-7) is:

$$\begin{aligned} E(U_1(\boldsymbol{\epsilon}_i, t)) &= E\left\{\sum_{i=1}^{n}\left\{diag\left(\frac{\partial\mu_{ijk}}{\partial\boldsymbol{\alpha}_1}\right)\right\}_i \frac{\boldsymbol{V}_{1i}^{-1}}{\phi}\left(\boldsymbol{y}_i - \boldsymbol{\mu}_i\left[\boldsymbol{X}_i\boldsymbol{\beta}, \sum_{b=1}^{m}\alpha_{1b}s_b(t), \hat{f}_2(\boldsymbol{Z}_{2i})\right]\right)\right\} \\ &= \sum_{i=1}^{n}\left\{diag\left(\frac{\partial\mu_{ijk}}{\partial\boldsymbol{\alpha}_1}\right)\right\}_i \frac{\boldsymbol{V}_{1i}^{-1}}{\phi} E\left(\boldsymbol{y}_i - \boldsymbol{\mu}_i\left[\boldsymbol{X}_i\boldsymbol{\beta}, \sum_{b=1}^{m}\alpha_{1b}s_b(t), \hat{f}_2(\boldsymbol{Z}_{2i})\right]\right) \\ &= 0 \, \forall t \in \mathbb{R} \end{aligned}$$

Analogous results are obtained for equations (4-8), (4-9) and (4-10). As $E(Y_{ijk}^2) < \infty$ and the density function satisfies the regularity conditions then, by Theorems 1 and 2 of Pan (2001b) and by theorem 2.6 of Lehmann and Casella (2006, pg 440), for Equation (4-7), it follows that:

$$0 < E\left(U_1(\boldsymbol{\epsilon}_i, t)U_1(\boldsymbol{\epsilon}_i, t)^T\right) < \infty$$

Similarly, for equations (4-8), (4-9) and (4-10), the following results are obtained:

$$\begin{aligned} E(U_2(\boldsymbol{\epsilon}_i, t)) = 0, \qquad & 0 < E\left(U_2(\boldsymbol{\epsilon}_i, t)U_2(\boldsymbol{\epsilon}_i, t)^T\right) < \infty, \ \forall t \in \mathbb{R}, \\ E(U_3(\boldsymbol{\epsilon}_i, t)) = 0, \qquad & 0 < E\left(U_3(\boldsymbol{\epsilon}_i, t)U_3(\boldsymbol{\epsilon}_i, t)^T\right) < \infty, \ \forall t \in \mathbb{R}, \\ E(U_4(\boldsymbol{\epsilon}_i)) = 0, \qquad & 0 < E\left(U_4(\boldsymbol{\epsilon}_i)U_4(\boldsymbol{\epsilon}_i)^T\right) < \infty, \end{aligned}$$

where $\boldsymbol{\epsilon}_i = (\epsilon_{i11}, \ldots, \epsilon_{in_{iP}}) = \boldsymbol{z}_i - \hat{\boldsymbol{z}}_i$, with $\boldsymbol{z}_i$ defined in Equation (4-13).

iii) According to theorem 2.6 of Lehmann and Casella (2006, pg 441), there exist $\{b_{1ijk}\}$, $\{b_{2ijk}\}$ and $\{b_{3ijk}\}$ with $0 < \text{ínf}_{ijk}(b_{ijk}) \leq \sup_{ijk}(b_{ijk}) < \infty$, $0 < \text{ínf}_{ijk}(b_{ijk}) \leq \sup_{ijk}(b_{ijk}) < \infty$ and $0 < \text{ínf}_{ijk}(b_{ijk}) \leq \sup_{ijk}(b_{ijk}) < \infty$ such that when $s \to 0$, the following properties hold, respectively::

$$\begin{aligned} \sup_{ijk} |E(U_1(\boldsymbol{\epsilon}_{ijk} + s, t)) - b_{1ijk}s| &= O(s^2), \ \forall t \in \mathbb{R}, \\ \sup_{ijk} |E(U_2(\boldsymbol{\epsilon}_{ijk} + s, t)) - b_{2ijk}s| &= O(s^2), \ \forall t \in \mathbb{R}, \\ \sup_{ijk} |E(U_3(\boldsymbol{\epsilon}_{ijk} + s)) - b_{3ijk}s| &= O(s^2), \end{aligned} \tag{4-19}$$

Also, when $s \to 0$, exist constants $c > 0$ and, $C < \infty$ such that:

$$\begin{aligned} \sup_{ijk} \left\{E(U_1(\boldsymbol{\epsilon}_{ijk} + s, t) - U_1(\boldsymbol{\epsilon}_{ijk}, t))^2\right\} &\leq C|s|, \ \forall t \in \mathbb{R} \\ \sup_{ijk} \left\{E(U_2(\boldsymbol{\epsilon}_{ijk} + s, t) - U_2(\boldsymbol{\epsilon}_{ijk}, t))^2\right\} &\leq C|s|, \ \forall t \in \mathbb{R} \\ \sup_{ijk} \left\{E(U_3(\boldsymbol{\epsilon}_{ijk} + s) - U_1(\boldsymbol{\epsilon}_{ijk}))^2\right\} &\leq C|s| \end{aligned}$$

Furthermore, $|U_1(\nu+\eta, t) - U_1(\nu, t) - U_1(\eta, t)| \leq c$, $|U_2(\nu+\eta, t) - U_2(\nu, t) - U_2(\eta, t)| \leq c$ and $|U_3(\nu + \eta) - U_3(\nu) - U_3(\eta)| \leq c$ for any $|\eta| \leq s$ and $\nu, t \in \mathbb{R}$.

- Let $\boldsymbol{\Delta}_n$ be a diagonal matrix with elements δ_{ijkl} defined in Equation (4-18) and let $\boldsymbol{\Lambda}_n$ be a diagonal matrix with elements b_{3ijk} defined in Equation (4-19), then by definition of the random variables δ_{ijkl} it follows that:

$$E(\boldsymbol{\Delta}_n) = 0 \text{ and } \sup_n \left\{\frac{1}{n}E(||\boldsymbol{\Delta}_n||^2)\right\} < \infty$$

Since, $\boldsymbol{\Gamma}_n$ is a block diagonal matrix with elements $A_i = E\left(U_3(\boldsymbol{\epsilon}_i)U_3(\boldsymbol{\epsilon}_i)^T\right)$, then by the theorem 1 in Pan (2001b),

$$\frac{1}{n}\boldsymbol{\Delta}_n^T\boldsymbol{\Gamma}_n\boldsymbol{\Delta}_n \xrightarrow{p} \boldsymbol{B} \tag{4-20}$$

$$\frac{1}{n}\boldsymbol{\Delta}_n^T\boldsymbol{\Lambda}_n\boldsymbol{\Delta}_n \xrightarrow{p} \boldsymbol{A} \tag{4-21}$$

- The matrices $\boldsymbol{H}_{1n}$ and $\boldsymbol{H}_{2n}$ defined in Equation (4-17) are symmetric and positive definite, so they have square root (Bunch and Hopcroft, 1974). Let $\boldsymbol{H}_{1n}^{\frac{1}{2}}$ and $\boldsymbol{H}_{2n}^{\frac{1}{2}}$ be those square roots, respectively. Also, as a kernel spline forms a linearly independent basis, and by Theorem 21.5.1 of Harville (1997, pg 537), $\boldsymbol{H}_{1n}^{\frac{1}{2}}$ and $\boldsymbol{H}_{2n}^{\frac{1}{2}}$ are non-singular matrices and their eigenvalues are bounded between zero and infinity.

With the previous results and using Theorem 1 and Theorem 2 of He et al. (2002), the following results are obtained:

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\sum_{b=1}^{m}\hat{\alpha}_{1b}s_b(\boldsymbol{Z}_{1ijk}) - f_1(\boldsymbol{Z}_{1ijk})\right\}^2 = O\left(n^{-\frac{2r}{(2r+1)}}\right) \tag{4-22}$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\sum_{b=1}^{m}\hat{\alpha}_{1b}s_b(\boldsymbol{Z}_{2ijk}) - f_2(\boldsymbol{Z}_{2ijk})\right\}^2 = O\left(n^{-\frac{2r}{(2r+1)}}\right) \tag{4-23}$$

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}) \tag{4-24}$$

where the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are defined in equations (4-14) and (4-15), respectively. □

4.3. Model diagnostics

4.3.1. Selection

In order to compare the fit of the proposed model model (equations (4-3) and (4-4)) with the fit of conventional models, then the quasi-likelihood criterion (QIC) defined by Pan (2001a) is used:

$$QIC = -2QL(\hat{\boldsymbol{\mu}}; \boldsymbol{I}) + 2trace(\hat{\boldsymbol{\Omega}}_I^{-1}\hat{\boldsymbol{V}}_{\boldsymbol{R}}) \tag{4-25}$$

where $\hat{\mu} = \hat{\eta} = g^{-1}(x\hat{\beta})$ is the estimated expected value for the observation with the model assuming the correlation matrix $\boldsymbol{R}$ and the estimates are obtained using Equation (4-12), $\hat{\boldsymbol{\Omega}}_I$ is the estimated variance matrix for vector $\boldsymbol{\beta}$ under a correlation matrix $\boldsymbol{R}(\boldsymbol{\alpha}) = \mathbb{I}_{n_i}$, and $\hat{\boldsymbol{V}}_{\boldsymbol{R}}$ is the variance matrix estimated for the vector $\boldsymbol{\beta}$ assuming the correlation matrix $\boldsymbol{R}(\boldsymbol{\alpha})$ as in Equation (4-15). After fitting several models, the model with lowest QIC is selected because it is the one featuring the best balance between goodness of fit and complexity.

4.3.2. Residuals

Due to the similarity of the proposed estimator of $\boldsymbol{\beta}$ and the weighted least squares, with weigthts given by matrix $\boldsymbol{W}_i$, Pearson standardized residuals are proposed to assess model validity:

$$r_{ijk} = \frac{\boldsymbol{e}_{ijk}^T \hat{\boldsymbol{W}}_i^{\frac{1}{2}} (\hat{\boldsymbol{z}}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}})}{\sqrt{1 - \hat{h}_{ijk}}} \tag{4-26}$$

where $\boldsymbol{e}_{ijk}$ is a vector of n_{ij} zeros, except at position k, $\hat{\boldsymbol{W}}_i^{\frac{1}{2}}$ is the square root of the matrix $\hat{\boldsymbol{W}}_i$ and h_{ijk} is the ijk element of the diagonal of the projection matrix $\mathcal{H}$, which is:

$$\mathcal{H} = diag\left\{\boldsymbol{H}_1, \cdots, \boldsymbol{H}_n\right\} \tag{4-27}$$

with

$$\boldsymbol{H}_i = \boldsymbol{W}_i^{\frac{1}{2}} \boldsymbol{X}_i (\boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i)^{-1} \boldsymbol{X}_i^T \boldsymbol{W}_i \tag{4-28}$$

According to Tsuyuguchi et al. (2020), the residuals defined in Equation (4-26) are asymptotically normal with zero mean and standard deviation close to 1. Therefore, these can be used to validate the fitted model and the conditional distribution assumption of Y.

4.4. Simulation study

For the simulation study, the crossover design with extra period (Jones and Kenward, 2015) defined in Table **4-2** will be used, so, assume the following model:

$$g(\mu_{ijk}) = \boldsymbol{x}_{ijk}^T \boldsymbol{\beta} + c_1 cos(t_{jk}) + c_2 sen(t_{jk}) \delta_{jk}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$, $\beta_0 \sim N(0,1)$, $\beta_1 \in (0.5, 1, 2)$, $\beta_2 = \beta_3 = 3$, $c_1 \sim N(0,1)$, $c_2 \sim N(0,1)$, $i = 1, \ldots, n$, $j = 1, \ldots, 3$ $k = 1, \ldots, 3$, $n = 2, \ldots, 100$ and $Y_{ijk}|(\boldsymbol{\beta}, c_1, c_2) \sim Poisson(\mu_{ijk})$.

The number of individuals per sequence is varied from 2 to 50. β_0 is the mean, β_1 is the difference between treatments A and B, β_2 and β_3 are the effects of period 2 and 3, respectively. The time effect is modeled by $c_1 cos(t_{jk})$ and the carry-over effect of treatment A on B is modeled by $c_2 sin(t_{jk})$. The δ_{jk} is equal to 0 in period 1 ($j = 1$) and, it is equal to 1 if in the previous period, the individual received treatment A, that is, it is a first-order carry-over effect (Patterson, 1951).

Each scenario is simulated 1000 times with an autoregressive correlation matrix of order 1, and for each component of $\boldsymbol{\beta}$, the following goodness of fit measures are obtained: 1) The

Sequence	Period 1	Period 2	Period 3
ABA	15 observations	15 observations	15 observations
BAB	15 observations	15 observations	15 observations

Table 4-2.: Crossover design with extra period

root mean square error $RMSE = \sqrt{(\hat{\beta}_i - \beta_i)^2}$ and 2) the percentage of times the hypothesis $H_0: \beta_1 = 0.5$, $H_0: \beta_1 = 1$ and $H_0: \beta_1 = 2$ are not rejected at 95 % confidence bands. Further, for each scenario, the following three models are fitted: 1) A model defined by Equation (4-3) which is denoted by GEE-S, 2) a GEE model where the effect of time is linear which is denoted by GEE-1, and 3) a GEE model with quadratic time effect which is denoted by GEE-2.

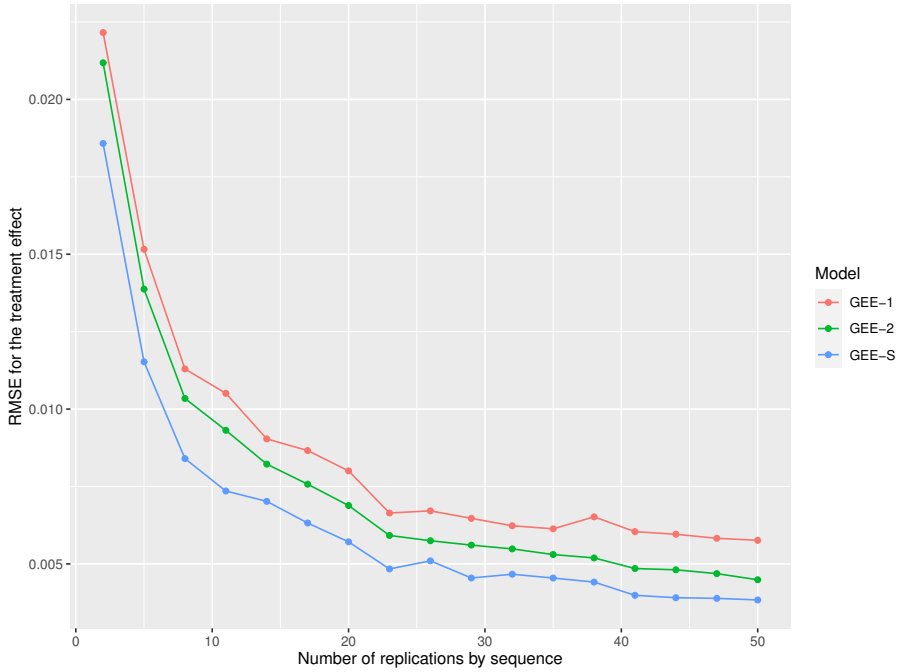

Figure 4-1.: RMSE for the estimate of $\beta_1 = 0.5$ in each of the models

Table 4-4 shows the coverage obtained for three different values of the treatment effect (β_1) using 95 % confidence intervals. A coverage close to 95 % is observed for replicate sizes greater than 8 in the model with Splines; the other two models show coverage equal to 100 %, that is, these models over-estimate the variance of the estimated treatment effects. This is confirmed in Figure 4-1, where it is shown that although the RMSE of each model decreases as the number of replicas increases, it does so faster in the Splines model.
On the other hand, Table 4-3 shows the coverage obtained for the period effect (β_2 and β_3)

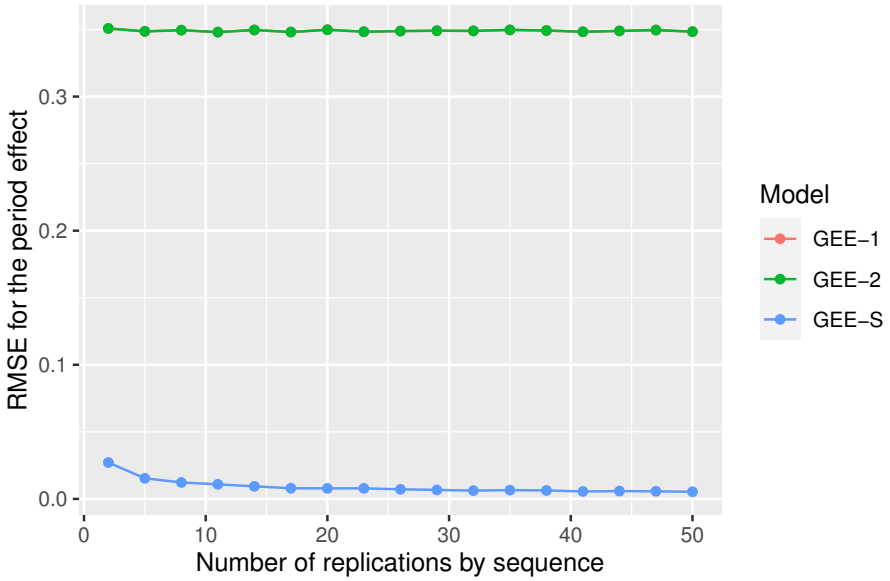

Figure 4-2.: RMSE for the estimate of $\beta_2 = 3$ in each of the models

using 95 % confidence intervals. The models without splines have a coverage of zero in all the scenarios because the effect of the period is confused with the time effect in the estimation equation, which is also observed in Figure 4-2. The confidence intervals obtained from splines have a coverage close to 95 % when there were 11 experimental units per sequence, see Figure 4-2. Results from this simulation suggest that when time effects are not linear or quadratic, models that do not use splines overestimate the variance of the treatment effects, and these fail to estimate unbiasedly the period effects. This leads to erroneous conclusions about the effectiveness of the sequences and treatments.

	$H_0: \beta_2 = 3$			$H_0: \beta_3 = 3$		
n	GEE-S	GEE-1	GEE2	GEE-S	GEE-1	GEE2
2	0.79	0.44	0.44	0.80	0.50	0.51
5	0.84	0.00	0.00	0.89	0.00	0.00
8	0.94	0.00	0.00	0.90	0.00	0.00
11	0.94	0.00	0.00	0.92	0.00	0.00
14	0.92	0.00	0.00	0.92	0.00	0.00
17	0.94	0.00	0.00	0.94	0.00	0.00
20	0.94	0.00	0.00	0.95	0.00	0.00
23	0.94	0.00	0.00	0.95	0.00	0.00
26	0.94	0.00	0.00	0.95	0.00	0.00
29	0.92	0.00	0.00	0.94	0.00	0.00
32	0.95	0.00	0.00	0.95	0.00	0.00
35	0.92	0.00	0.00	0.92	0.00	0.00
38	0.92	0.00	0.00	0.92	0.00	0.00
41	0.95	0.00	0.00	0.92	0.00	0.00
44	0.93	0.00	0.00	0.94	0.00	0.00
47	0.93	0.00	0.00	0.92	0.00	0.00
50	0.94	0.00	0.00	0.90	0.00	0.00

Table 4-3.: Proportion of times that hypothesis H_0 is not rejected for some values of β_2 and β_3 (the period effects)

4.5. Application

Two studies are presented below where the model proposed in Equation (4-3) is used. In both the software R Core Team (2022) is used through adaptation of the package **geeM** built by McDaniel et al. (2013).

4.5.1. Systolic pressure data

Jones and Kenward (2015, pg 204) described the following crossover design: 3 treatments for blood pressure control were used; treatment A consisted of a 20 mg dose of a test drug, treatment B was a 40 mg dose of the same drug, and treatment C was a placebo. 6 sequences of three periods (ABC, ACB, BCA, BAC, CAB, CBA) were organized and each one was applied to two individuals. In each application period, 10 successive measurements of systolic blood pressure were made: 30 and 15 minutes before the application, and 15, 30, 45, 60, 75, 90, 120 and 240 minutes after the application, as shown in Table **4-1**, and the profile is shown in Figure **4-3**.

Figure **4-4**.a) shows the smoothed function corresponding to the effect of time on blood

	$H_0: \beta_1 = 0.5$			$H_0: \beta_1 = 1$			$H_0: \beta_1 = 2$		
n	GEE-S	GEE-1	GEE2	GEE-S	GEE-1	GEE2	GEE-S	GEE-1	GEE2
2	0.88	1.00	1.00	0.82	1.00	1.00	0.85	1.00	1.00
5	0.90	1.00	1.00	0.92	1.00	1.00	0.88	1.00	1.00
8	0.93	1.00	1.00	0.96	1.00	1.00	0.93	1.00	1.00
11	0.92	1.00	1.00	0.96	1.00	1.00	0.92	1.00	1.00
14	0.95	1.00	1.00	0.94	1.00	1.00	0.96	1.00	1.00
17	0.94	1.00	1.00	0.94	1.00	1.00	0.94	1.00	1.00
20	0.95	1.00	1.00	0.94	1.00	1.00	0.94	1.00	1.00
23	0.95	1.00	1.00	0.96	1.00	1.00	0.96	1.00	1.00
26	0.96	1.00	1.00	0.96	1.00	1.00	0.97	1.00	1.00
29	0.96	1.00	1.00	0.97	1.00	1.00	0.93	1.00	1.00
32	0.95	1.00	1.00	0.94	1.00	1.00	0.97	1.00	1.00
35	0.95	1.00	1.00	0.94	1.00	1.00	0.92	1.00	1.00
38	0.96	1.00	1.00	0.93	1.00	1.00	0.94	1.00	1.00
41	0.94	1.00	1.00	0.94	1.00	1.00	0.98	1.00	1.00
44	0.95	1.00	1.00	0.94	1.00	1.00	0.93	1.00	1.00
47	0.97	1.00	1.00	0.95	1.00	1.00	0.93	1.00	1.00
50	0.95	1.00	1.00	0.94	1.00	1.00	0.97	1.00	1.00

Table 4-4.: Proportion of times that hypothesis H_0 is not rejected for some values of β_1 (the treatment effect)

pressure; it is based on the moments of measurement for the design in Table **4-1**. That is, 30 and 15 minutes before the application, and 15, 30, 45, 60, 75, 90, 120 and 240 minutes after the application. Additionally, Figure **4-4**.a) shows the average function and its 95 % confidence bands obtained through cross-validation. A wide drop in pressure is observed from the time that the patient expects to receive treatment, then it rises a little and remains stable. This behavior is widely studied in medical settings, see Stergiou et al. (1998) and Fanelli et al. (2021).

The carry-over effects of treatment A and treatment B are observed in Figures **4-4**.b) and **4-4**.c), respectively. The value for the carry-over effect of A is positive and increases over time, which implies that having applied placebo in a previous period will generate higher blood pressure values in the following application period. For treatment B (medium dose), the carry-over effect is close to zero; therefore, it is a negligible effect for the next treatment. Table **4-5** shows the parametric effects, their standard error and the Wald statistic built from matrices (4-14) and (4-15). It is worth highlighting the positive effect of the baseline, i.e., people have the highest blood pressure before starting the study. The periods are not significant i.e., the conditions were similar across the study, and these had a significant effect on pressure of treatment C on pressure reduction.

Finally, Figure **4-4**.d) shows the confidence bands for the quantiles of the standardized

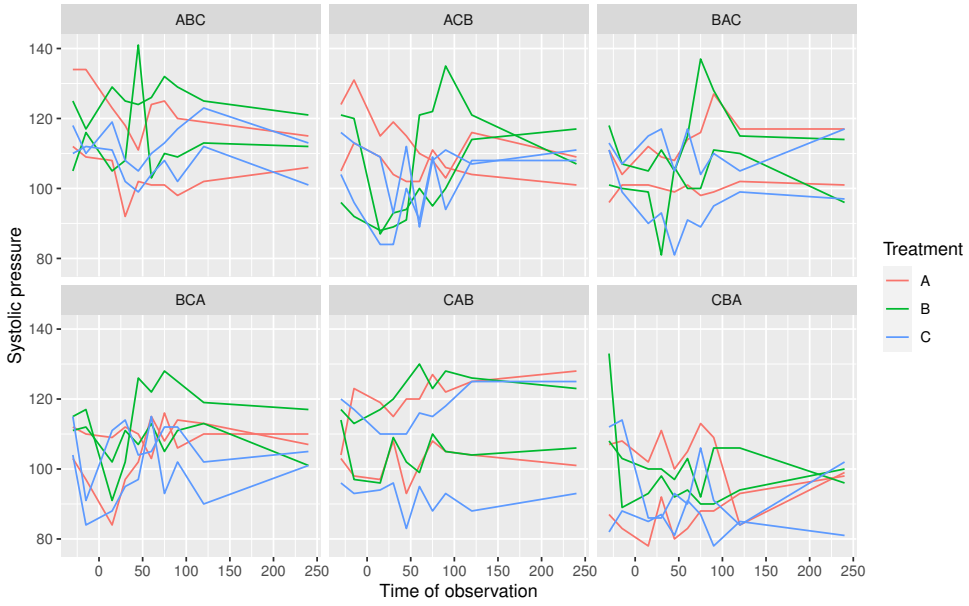

Figure 4-3.: Blood systolic pressure (mmHg) observed through time (minutes)

residuals defined in (4-26) compared to a standard normal distribution. These residuals seem to fit the normal distribution assumption.

4.5.2. Blood Sugar levels in Rabbits

Kenward and Roger (2009) described the following crossover experiment: two treatments for the control of diabetes A and B were used, two sequences of four periods (ABAB, BABA) were organized and each one was applied to twelve female rabbits; each period lasted one week. In each period, at the middle of the week, five successive measurements of the blood sugar level were taken: 0, 1.5, 3, 4.5 and 6 hours after the application, as shown in Table **4-6** and Figure **4-5**.

Assuming that the distribution of blood sugar levels is normal, a first analysis was run. When making the normal probability plot of the standardized residuals defined in (4-26), it is observed that they do not fit correctly to a standard normal distribution, as seen in Figure **4-6** and the assumption is rejected. A gamma distribution is then explored, with log-linear linkage. Figure **4-7** c) shows the confidence bands for the quantiles of the standardized residuals defined in (4-26) against a standard normal distribution, concluding that the gamma distribution assumption is adequate.

To compare the distributions, three models are made to analyze the response variable: i) Under the assumption of normality, ii) Under the assumption of gamma distribution and inverse link, and iii) under the assumption of gamma distribution and a loglinear link. In

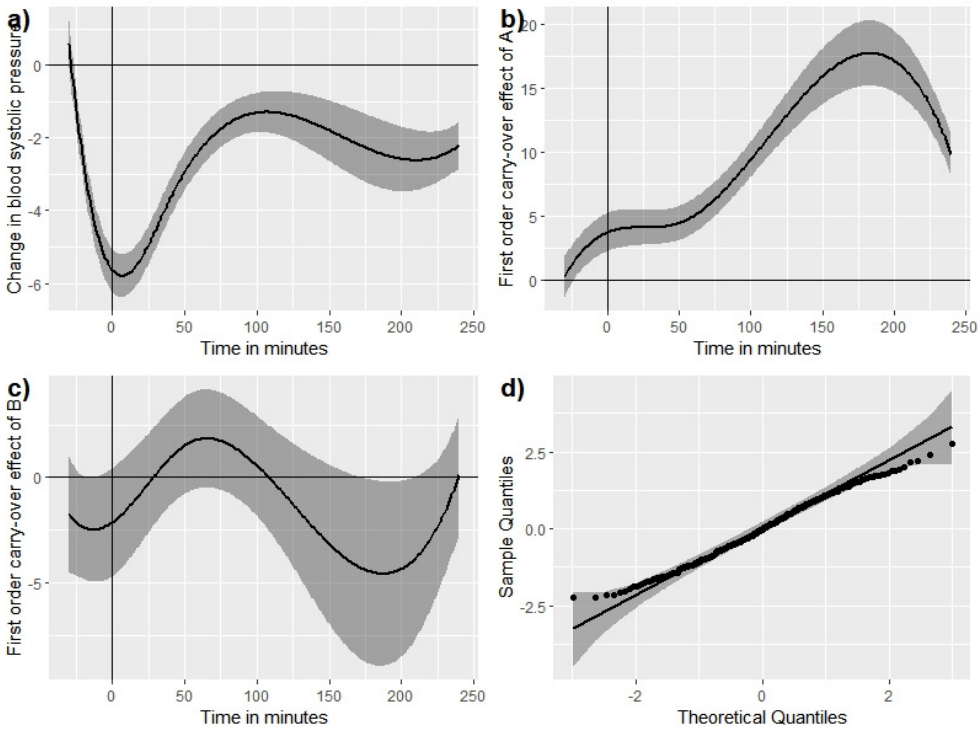

Figure 4-4.: a) Changes of blood systolic pressure through time using splines, b) First order carry-over effect of the treatment A through time using splines, c) First order carry-over effect of the treatment B through time using splines, and d) Residuals normal probability plot. All figures present 95 % confidence intervals and are based on the crossover design of Table 4-1

each one, the QIC is calculated and shows in Table 4-7. Therefore, all the analysis is carried out with the gamma distribution assumption and the loglinear model given by:

$$\ln(\mu_{ijk}) = \boldsymbol{x}_{ijk}^T\boldsymbol{\beta} + f(\boldsymbol{Z}_{ijk}) + f_1(\boldsymbol{Z}_{ijk})$$

where $\boldsymbol{x}_{ijk}$ is the matrix that contains the values of period and treatment, $\boldsymbol{Z}_{ijk}$ contains the measurement time instants within each period, f is the function that describes the temporary effect and f_1 is the function that describes the carry-over effect of A on B.
The spline-smoothed function for the time effect on the blood sugar level of female rabbits is shown in Figure 4-7a), it is based on the moments of measurement for the design of Table 4-6; the average functions and their 95 % confidence bands, estimated through cross validation are presented. A marked decrease in levels is observed until 2 hours, and then an increase until hour 6. In Jones and Kenward (2015, pag 237) it was stated that there was an effect of the hours, but its form is not explained. However, the proposed model permits to describe this effect. Figure 4-7.b) shows the carry-over effects of treatment A over treatment

	Estimate	Std.err	Wald	Pr(> |W|)
Intercept	109.35	3.17	1192.44	0.00
BaseLine	3.85	1.47	6.82	0.01
Period 2	-0.88	3.49	0.06	0.80
Period 3	-3.22	3.60	0.80	0.37
Treatment B	0.70	3.04	0.05	0.82
Treatment C	-5.95	3.60	2.73	0.01

Table 4-5.: Analysis of blood systolic pressure data using GEE-Splines

Sequence		Period 1	Period 2	Period 3	Period 4
(1)	Ind 1	5 observations	5 observations	5 observations	5 observations
	⋮	⋮	⋮	⋮	⋮
	Ind 11	5 observations	5 observations	5 observations	5 observations
(2)	Ind 12	5 observations	5 observations	5 observations	5 observations
	⋮	⋮	⋮	⋮	⋮
	Ind 22	5 observations	5 observations	5 observations	5 observations

Table 4-6.: Structure of the crossover design of blood sugar levels in rabbits

B. This increases over time, but it is close to 0, which implies that having applied A in a previous period will not significantly affect the next application period. The parametric effects, their standard error and the Wald statistic constructed with matrices (4-14) and (4-15) are presented in Table **4-8**. It is noteworthy that there are no significant effects of treatment, similar to that obtained by Jones and Kenward (2015) and Kenward and Roger (2009), but there is an positive effect of period four. This behavior was not analyzed in previous studies and can be seen as increased insulin resistance by blood cells; similar to behaviors reported by Ning et al. (2015) and Da Silva et al. (2020).

In the modelling of blood pressure, the normal distribution presents a better performance than the gamma distribution, while in blood sugar levels, the gamma distribution achieves a better fit than the normal distribution. Choosing the most suitable distribution is desirable, because the standard errors of each estimator of the parametric effects of the model are smaller. In addition, the Pearson residuals show a behavior that conforms to a standard normal in both cases, guaranteeing that the model specification is adequate.

These applications show the importance of the semiparametric approach proposed in this chapter to model time and carry-over effects in crossover designs with repeated measures

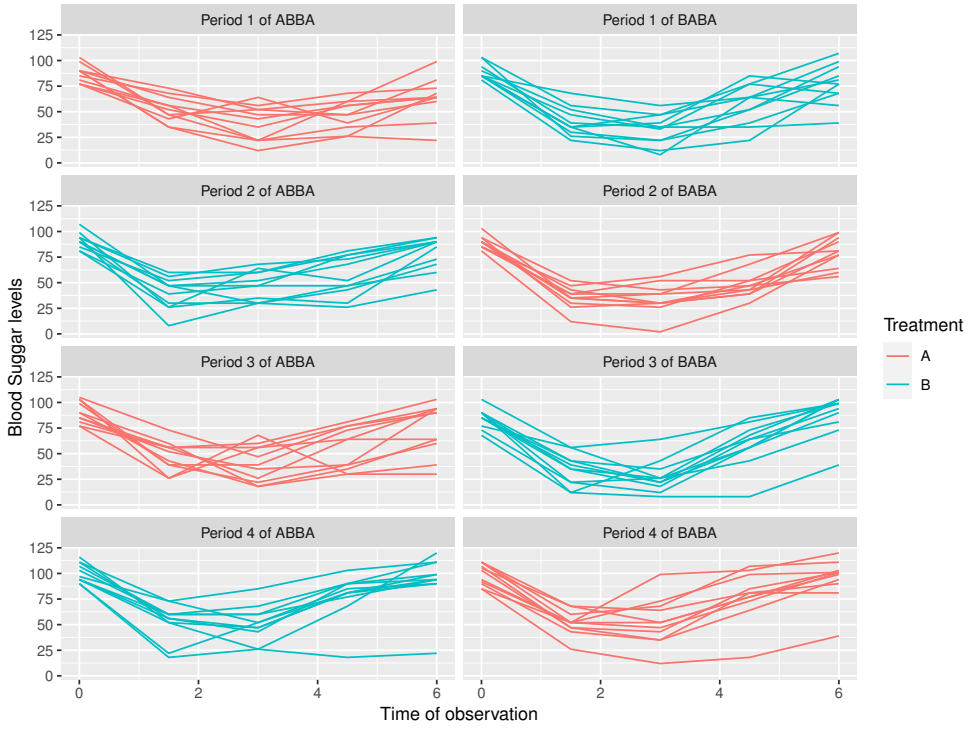

Figure 4-5.: Blood sugar levels (mg/dL) in rabbits through time (hours)

within periods. Also, it complements the simulation study, where the efficiency of the proposed model over conventional models was verified.

Model	QIC
Normal	3927.6
Gamma Inverse	3732.22
Gamma Log	3728.5

Table 4-7.: QIC for the three fitted models to the response of blood sugar levels in rabbits

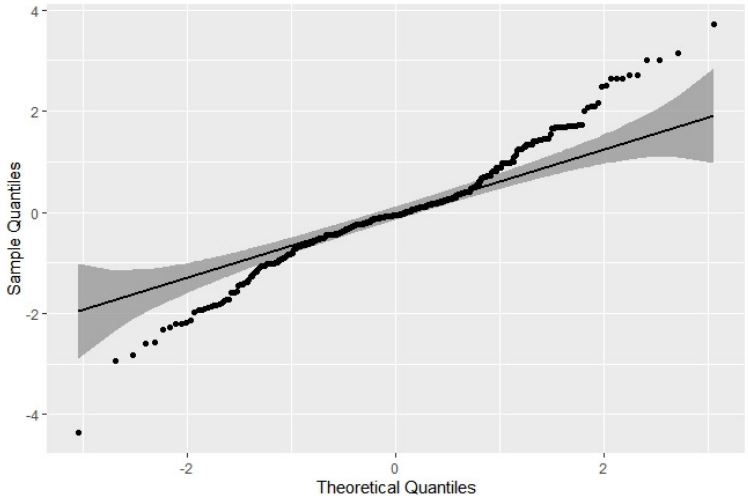

Figure 4-6.: Residuals normal probability plot with 95 % confidence intervals of the residuals obtained assuming a Gaussian distribution of the response in the crossover design of the Table **4-6**

	Estimate	Std.err	Wald	$\Pr(>\|W\|)$
Intercept	4.47	0.05	9734.05	0.00
Period 2	0.01	0.06	0.00	0.96
Period 3	-0.01	0.06	0.01	0.94
Period 4	0.25	0.06	16.68	0.00
Treatment B	-0.02	0.04	0.28	0.60

Table 4-8.: Analysis of blood sugar levels data using GEE-Splines

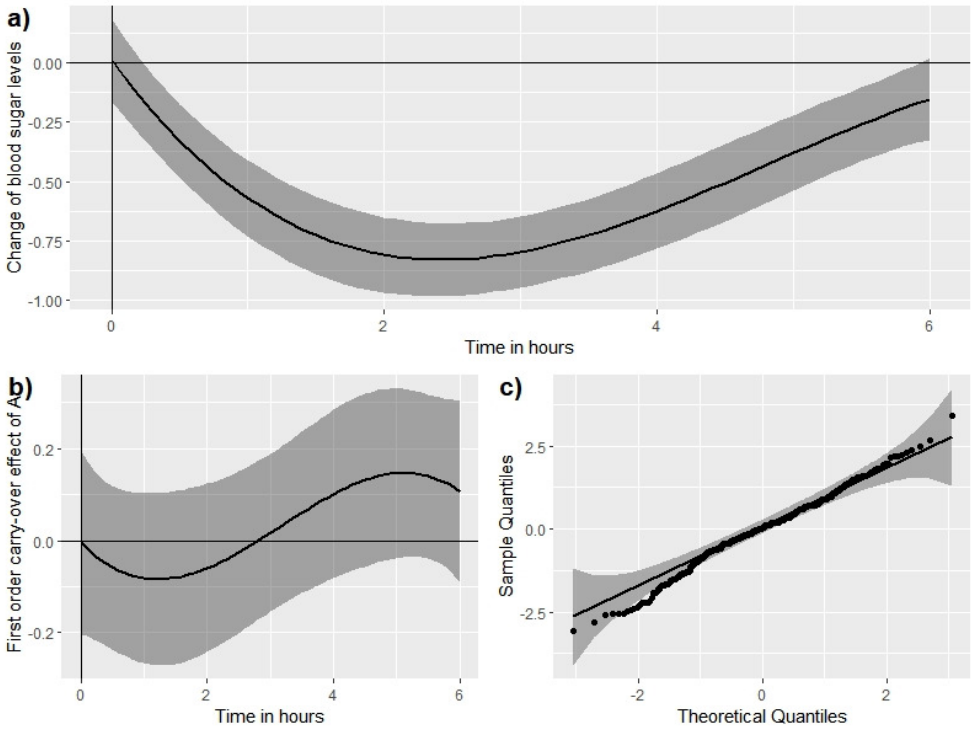

Figure 4-7.: a) Change of blood sugar levels through time using splines, b) First order carry-over effect of the treatment A through time using splines, and c) Residuals normal probability plot. All figures present 95 % confidence intervals and are based on the crossover design of the Table **4-6** with linear log link for the mean of an assumed gamma distribution

5. Conclusions

This chapter presents the conclusions obtained after carrying out this work. Conclusions are presented for chapters 2, 3 and 4, since in these the results of the thesis were built. In the second chapter, when the response variable has a Poisson distribution, we build two ways of approaching a crossover study when their experimental units are observed once per period. The first one by using the log-linear link and the second one by applying GEE. When the response variable presents subdispersion or overdispersion, both methodologies are functional for the analysis. The building of a deviation analysis table, that allows performing a satisfactory inference for a Poisson crossover design, is the principal contribution of this chapter.

Our methodology includes variation causes that are necessary for the analysis, such as sequences, periods, treatments, carry-over of any order and blocking covariates. In the simulation, a behavior similar to normal analysis is observed without the presence of carry-over, but when these effects were presented, the normal model is biased while the models of this job are of better performance. However, the model with GEE has an interval with coverage close to 95 % when the number of individuals increases, In addition, with GEE, regardless of the presence or absence of carry-over effects, unbiased estimates of the important effects of design are always obtained.

In the application exercise, in chapter two, it is noted that with the simple analysis of logarithms, estimates of the important effects are obtained, but the covariate cannot be included, which reduces the effectiveness of the design. With the GEE, more accurate conclusions are obtained. This methodology allows the inclusion of designs with more sequences or periods and even covariates. In the case of circular designs, GEE could also be used to estimate the different effects present in the design.

Defining the correlation matrix structure is a highly relevant decision when using GEE, a proper specification of correlation structures in longitudinal data analysis improves estimation efficiency, leading to more reliable statistical inferences. Hence, in the third chapter, we develop a family of correlation structures that combine some of the classical structures used in longitudinal data analysis with an additional matrix that allows more flexibility in the overall correlation structure by adding a few parameters. Moreover, we provide an explicit estimation method of these parameters which features some sound statistical properties.

The theoretical results supporting some asymptotic properties of the proposed estimators

are illustrated through the simulation study where a gain in goodness of fit was observed across the different simulation scenarios. The QIC was able to select the correct model, which had the proposed correlation structure. In addition, the confidence intervals built from the GEE showed a coverage that matched the nominal value. The estimation by intervals for each of the parameters presents coverage close to 95 %, which shows a correct theoretical specification of each univariate interval.

As to the real data analysis, in chapter three, the results for the arterial pressure data showed the importance of accounting for carryover effects as they are useful for correctly estimating and interpreting the treatment affects across the time. If not included in the model, these residual effects may induce confusion problems. On the other hand, in the dairy cattle data, significant carryover effects were detected as well. Regarding the estimates of correlation matrices in both datasets, the QIC selected the proposed correlation matrix. Thus, the theoretical and empirical results from real and simulated data analyses suggest that the proposed methodology is promising and may be applied to perform better inferences from data obtained under crossover designs without washout periods and a repeated measures structure.

In the fourth chapter, a semiparametric methodology is built. The proposed methodology provides highly desirable properties of the resulting estimators. It allows doing asymptotic inference and better to model temporal carry-over behaviors that would be intractable in parametric scenarios, as in the case of blood pressure data, where these effects do not present the typical polynomial effects. In addition, detecting these carry-over effects of the placebo allows estimating treatment effects with greater precision and unbiasedness, which is basically the objective of any crossover design. In the insulin data in rabbits, the behavior of the estimated effect of time is similar to a quadratic form, which shows that this methodological proposal encompasses the classical parametric temporal models with linear or cubic polynomials. In addition, the GEE allow modeling a large number of response variables, not only normal or continuous, but also counts or proportions of successes. In the simulation, the inferential gain is evidenced in terms of coverage and control of the type I and II error of the hypothesis tests associated with the parameters of interest; that is, treatment and period effects when the temporal behavior is sinusoidal. While linear or quadratic models lose efficiency and unbiasedness; therefore, estimation with splines is presented as a useful tool for this type of design. The asymptotic properties of the estimators allow an agile and fast verification of the model, because its similarity with weighted least squares is demonstrated. Therefore, the adaptation of widely used diagnostic tests in normal linear models can be used.

As future work, the evaluation in circular designs can be considered, to observe the carry-over effects in this type of design. Also the evaluation of responses that do not belong to the exponential family as distribution of the response variable.

In case of missing at random in the response variable, the analysis with random effects of the subject could be explored for the analysis through GEE. If the missing data are not negligible abandonment (death, recovery, etc.), it is necessary to adapt the methodology presented in this work and it remains as future research.

Bibliography

Agresti, A. (2002). *Categorical Data Analysis Second Edition.* Wiley Series in Probability and Statistics, New Jersey.

Amatya, A. and Demirtas, H. (2017). PoisNor: An R package for generation of multivariate data with Poisson and normal marginals. *Communications in Statistics-Simulation and Computation*, 46(3):2241–2253.

Bailey, R. A., Cameron, P. J., Filipiak, K., Kunert, J., and Markiewicz, A. (2017). On optimality and construction of circular repeated-measurements designs. *Statistica Sinica*, 27:1–22.

Basu, S. and Santra, S. (2010). A joint model for incomplete data in crossover trials. *Journal of Statistical Planning and Inference*, 140(10):2839–2845.

Biabani, M., Farrell, M., Zoghi, M., Egan, G., and Jaberzadeh, S. (2018). Crossover design in transcranial direct current stimulation studies on motor learning: potential pitfalls and difficulties in interpretation of findings. *Reviews in the Neurosciences*, 29(4):463–473.

Boik, R. J. (1991). Scheffés mixed model for multivariate repeated measures: a relative efficiency evaluation. *Communications in Statistics-Theory and Methods*, 20(4):1233–1255.

Bunch, J. R. and Hopcroft, J. E. (1974). Triangular factorization and inversion by fast matrix multiplication. *Mathematics of Computation*, 28(125):231–236.

Carey., V. J. (2019). *gee: Generalized Estimation Equation Solver.* R package version 4.13-20.

Chalikias, M. and Kounias, S. (2012). Extension and necessity of Cheng and Wu conditions. *Journal of Statistical Planning and Inference*, 142(7):1794–1800.

Chalikias, M. S. (2017). Two treatment repeated measurement designs with uncorrelated observations: A compact review. *Journal of Advanced Statistics*, 2(1):27.

Chard, A. N., Trinies, V., Edmonds, C. J., Sogore, A., and Freeman, M. C. (2019). The impact of water consumption on hydration and cognition among schoolchildren: Methods and results from a crossover trial in rural Mali. *PloS one*, 14(1):e0210568.

Chasiotis, V. (2021). On optimality and construction of two-treatment circular cross-over designs. *Communications in Statistics-Theory and Methods*, 10(1):1–10.

Chasiotis, V. and Kounias, S. (2021). Optimal two treatment circular repeated measurements designs up to four periods. *Communications in Statistics-Theory and Methods*, 50(20):4867–4878.

Cordeiro, G. M. (2004). On Pearson's residuals in generalized linear models. *Statistics & Probability letters*, 66(3):213–219.

Cordeiro, G. M. and McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):629–643.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):248–265.

Cruz, N. A., López Pérez, L. A., and Melo, O. O. (2023a). Analysis of cross-over experiments with count data in the presence of carry-over effects. *Statistica Neerlandica*.

Cruz, N. A., Melo, O. O., and Martinez, C. A. (2023b). A correlation structure for the analysis of Gaussian and non-Gaussian responses in crossover experimental designs with repeated measures. *Statistical Papers*.

Cruz, N. A., Melo, O. O., and Martinez, C. A. (2023c). Semiparametric generalized estimating equations for repeated measurements in cross-over designs. *Statistical Methods in Medical Research*, 32(5):1033–1050. PMID: 36919447.

Curtin, F. (2017). Meta-analysis combining parallel and crossover trials using generalised estimating equation method. *Research Synthesis Methods*, 8(3):312–320.

Da Silva, A. A., do Carmo, J. M., Li, X., Wang, Z., Mouton, A. J., and Hall, J. E. (2020). Role of hyperinsulinemia and insulin resistance in hypertension: metabolic syndrome revisited. *Canadian Journal of Cardiology*, 36(5):671–682.

Davies, R. (1980). Algorithm as 155: The distribution of a linear combination of chi-2 random variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(3):323–333.

Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Springer, San Diego.

Diaz, F. J., Berg, M. J., Krebill, R., Welty, T., Gidal, B. E., Alloway, R., and Privitera, M. (2013). Random-effects linear modeling and sample size tables for two special crossover designs of average bioequivalence studies: the four-period, two-sequence, two-formulation and six-period, three-sequence, three-formulation designs. *Clinical pharmacokinetics*, 52(12):1033–1043.

Dubois, A., Lavielle, M., Gsteiger, S., Pigeolet, E., and Mentré, F. (2011). Model-based analyses of bioequivalence crossover trials using the stochastic approximation expectation maximisation algorithm. *Statistics in medicine*, 30(21):2582–2600.

Duchesne, P. and Micheaux, L. (2010). Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. *Computational Statistics and Data Analysis*, 54:858–862.

Fanelli, E., Di Monaco, S., Pappaccogli, M., Eula, E., Fasano, C., Bertello, C., Veglio, F., and Rabbia, F. (2021). Comparison of nurse attended and unattended automated office blood pressure with conventional measurement techniques in clinical practice. *Journal of Human Hypertension*, 36(9):01–06.

Farebrother, R. (1984). Algorithm AS 204: The distribution of a positive linear combination of chi- squared random variables. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 33(3):332–339.

Forbes, A. B., Akram, M., Pilcher, D., Cooper, J., and Bellomo, R. (2015). Cluster randomised crossover trials with binary data and unbalanced cluster sizes: Application to studies of near-universal interventions in intensive care. *Clinical Trials*, 12(1):34–44.

Grayling, M. J., Mander, A. P., and Wason, J. M. (2018). Blinded and unblinded sample size reestimation in crossover trials balanced for period. *Biometrical Journal*, 60(5):917–933.

Hao, C., von Rosen, D., and von Rosen, T. (2015). Explicit influence analysis in two-treatment balanced crossover models. *Mathematical Methods of Statistics*, 24(1):16–36.

Hardin, J. W. and Hilbe, J. (2003). *Generalized Estimating Equations*. Chapman & Hall, Boca Raton.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*, volume 1. Springer, New York.

He, X., Zhu, Z.-Y., and Fung, W.-K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, 89(3):579–590.

Hinkelmann, K. and Kempthorne, O. (2005). *Design and Analysis of Experiments*, volume Volume 2 of *Wiley series in probability and mathematical statistics. Applied probability and statistics*. Wiley, New York.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48:419–426.

Jaime, G. (2019). *Uso de un residuo de papel como suplemento para vacas lecheras*. Tesis de maestría, Universidad Nacional de Colombia, Sede Bogota.

Jankar, J. and Mandal, A. (2021). Optimal crossover designs for generalized linear models: An application to work environment experiment. *Statistics and Applications*, 19(1):319–336.

Jankar, J., Mandal, A., and Yang, J. (2020). Optimal crossover designs for generalized linear models. *Journal of Statistical Theory and Practice*, 14(2):1–27.

Jones, B. and Kenward, M. G. (2015). *Design and Analysis of Cross-Over Trials Third Edition*. Chapman & Hall/CRC, Boca Raton.

Jorgensen, B. (1997). *The Theory of Dispersion Models*. Chapman & Hall, London.

Josephy, H., Vansteelandt, S., Vanderhasselt, M.-A., and Loeys, T. (2015). Within-subject mediation analysis in AB/BA crossover designs. *The international journal of biostatistics*, 11(1):1–22.

Kenward, M. G. and Jones, B. (1987). A log-linear model for binary cross-over data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36:192–204.

Kenward, M. G. and Roger, J. H. (2009). The use of baseline covariates in crossover studies. *Biostatistics*, 11(1):1–17.

Kitchenham, B., Madeyski, L., Curtin, F., et al. (2018). Corrections to effect size variances for continuous outcomes of cross-over clinical trials. *Statistics in medicine*, 37(2):320–323.

Kokonendji, C., Dossou-Gbete, S., and Demetrio, C. (2004). Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio. *Statistics and Operations Research Transactions*, 28:201–214.

Krzyśko, M. and Skorzybut, M. (2009). Discriminant analysis of multivariate repeated measures data with a kronecker product structured covariance matrices. *Statistical papers*, 50(4):817–835.

Lancaster, H. (1965). The helmert matrices. *The American Mathematical Monthly*, 72(1):4–12.

Layard, M. and Arvesen, J. (1978). Analysis of Poisson data in crossover experimental designs. *Biometrics*, pages 421–428.

Lehmann, E. L. and Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media.

Leorato, S. and Mezzetti, M. (2016). Spatial panel data model with error dependence: A bayesian separable covariance approach. *Bayesian Analysis*, 11(4):1035–1069.

Li, F., Forbes, A. B., Turner, E. L., and Preisser, J. S. (2018). Power and sample size requirements for gee analyses of cluster randomized crossover trials. *Statistics in Medicine*.

Li, F., Forbes, A. B., Turner, E. L., and Preisser, J. S. (2019). Power and sample size requirements for gee analyses of cluster randomized crossover trials. *Statistics in medicine*, 38(4):636–649.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, 96(455):1045–1056.

Lindsey, J. and Jones, B. (1997). Treatment–patient interactions for diagnostics of cross-over trials. *Statistics in medicine*, 16(17):1955–1964.

Liu, F. and Li, Q. (2016). A bayesian model for joint analysis of multivariate repeated measures and time to event data in crossover trials. *Statistical Methods in Medical Research*, 25(5):2180–2192.

Liu, H., Tang, Y., and Zhang, H. (2009). A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Computational Statistics and Data Analysis*, 53:853–856.

Longford, N. (1998). Count data and treatment heterogeneity in 2× 2 crossover trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(2):217–229.

Lui, K.-J. (2015). Test equality between three treatments under an incomplete block crossover design. *Journal of biopharmaceutical statistics*, 25(4):795–811.

Lui, K.-J. (2016). Notes on estimation in Poisson frequency data under an incomplete block crossover design. *Statistical Methodology*, 32:53–62.

Madeyski, L. and Kitchenham, B. (2018). Effect sizes and their variance for AB/BA crossover design studies. *Empirical Software Engineering*, 23(4):1982–2017.

Mathai, A. M. (1982). Storage capacity of a dam with Gamma type inputs. *Annals of the Institute of Statistical Mathematics*, 34(3)A:591–597.

McDaniel, L. S., Henderson, N. C., and Rathouz, P. J. (2013). Fast pure R implementation of GEE: application of the Matrix package. *The R Journal*, 5:181–187.

Melo, O., López, L., and Melo, S. (2007). *Diseño de experimentos, Métodos y aplicaciones*. Universidad Nacional de Colombia, Sede Bogotá, Bogotá.

Moschopoulos, P. G. and Canada, W. B. (1984). The distribution function of a linear combination of chi-squares. *Computers and Mathematics with Applications*, 10:383–386.

Ning, B., Wang, X., Yu, Y., Waqar, A. B., Yu, Q., Koike, T., Shiomi, M., Liu, E., Wang, Y., and Fan, J. (2015). High-fructose and high-fat diet-induced insulin resistance enhances atherosclerosis in watanabe heritable hyperlipidemic rabbits. *Nutrition & metabolism*, 12(1):1–11.

Oh, H. S., Ko, S.-g., and Oh, M.-S. (2003). A bayesian approach to assessing population bioequivalence in a 2 2 2 crossover design. *Journal of Applied Statistics*, 30(8):881–891.

Pan, W. (2001a). Akaike's information criterion in generalized estimating equations. *Biometrics*, 57:120–125.

Pan, W. (2001b). On the robust variance estimator in generalised estimating equations. *Biometrika*, 88(3):901–906.

Patterson, H. D. (1951). Change-over trials. *Journal of the Royal Statistical Society. Series B (Methodological)*, 13:256–271.

Pitchforth, J., Nelson-White, E., van den Helder, M., and Oosting, W. (2020). The work environment pilot: An experiment to determine the optimal office design for a technology company. *PloS one*, 15(5):e0232943.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ratkowsky, D., Alldredge, R., and Evans, M. (1992). *Cross-over Experiments*. Statistics A Series of Textbooks and Monographs, Washington.

Rosenkranz, G. K. (2015). Analysis of cross-over studies with missing data. *Statistical methods in medical research*, 24(4):420–433.

Rotnitzky, A. and Jewell, N. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for clustered data. *Biometrika*, 77:485–497.

Roy, A. and Khattree, R. (2005). On implementation of a test for kronecker product covariance structure for multivariate repeated measures data. *Statistical Methodology*, 2(4):297–306.

Shkedy, Z., Molenberghs, G., Craenendonck, H. V., Steckler, T., and Bijnens, L. (2005). A hierarchical Binomial-Poisson model for the analysis of a crossover design for correlated binary data when the number of trials is dose-dependent. *Journal of biopharmaceutical statistics*, 15(2):225–239.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(3):413–436.

Srivastava, M. S., von Rosen, T., and Von Rosen, D. (2008). Models with a kronecker product covariance structure: estimation and testing. *Mathematical Methods of Statistics*, 17(4):357–370.

Stergiou, G. S., Zourbaki, A. S., Skeva, I. I., and Mountokalakis, T. D. (1998). White coat effect detected using self-monitoring of blood pressure at home: comparison with ambulatory blood pressure. *American Journal of Hypertension*, 11(7):820–827.

Stoklosa, J. and Warton, D. I. (2018). A generalized estimating equation approach to multivariate adaptive regression splines. *Journal of Computational and Graphical Statistics*, 27(1):245–253.

Tsuyuguchi, A. B., Paula, G. A., and Barros, M. (2020). Analysis of correlated Birnbaum–Saunders data based on estimating equations. *TEST*, 29(3):661–681.

Vegas, S., Apa, C., and Juristo, N. (2016). Crossover designs in software engineering experiments: Benefits and perils. *IEEE Transactions on Software Engineering*, 42(2):120–135.

Wang, X. and Chinchilli, V. M. (2021). Analysis of crossover designs with nonignorable dropout. *Statistics in Medicine*, 40(1):64–84.

Wild, C. and Yee, T. (1996). Additive extensions to generalized estimating equation methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(4):711–725.

Yang, L. and Niu, X.-F. (2021). Semi-parametric models for longitudinal data analysis. *Journal of Finance and Economics*, 9(3):93–105.

Yu, L. and Peace, K. E. (2012). Spline nonparametric quasi-likelihood regression within the frame of the accelerated failure time model. *Computational Statistics & Data Analysis*, 56(9):2675–2687.

Zhang, H., Yu, Q., Feng, C., Gunzler, D., Wu, P., and Tu, X. (2012). A new look at the difference between the gee and the glmm when modeling longitudinal count responses. *Journal of Applied Statistics*, 39(9):2067–2079.

A. Codes in R chapter 2

A.1. Simulation 1

```r
library(ggplot2)
library(normtest)
library(tidyverse)
n = 200
nsim=500
beta=2

sim_norm_pois <- function(n=100, nsim=200, beta){
  medias=matrix(0, ncol=nsim, nrow=n)

  for(i in 1:n){
    for(j in 1:nsim){
      x=rpois(i, beta)
      medias[i,j] <-mean(x)
    }
    print(i)
  }


Esta_shapi=numeric(n)
p_valor=numeric(n)

for(i in 1:n){
    Prueba <- shapiro.test(medias[i,])
    Esta_shapi[i]<- Prueba$statistic
    p_valor[i] <- Prueba$p.value
    print(i)
  }

```

```
30   Resultados <- data.frame(Esta_shapi, p_valor, n=1:n, mu=beta)
31   return(Resultados)
32 }
33
34 beta <- c(0.1,0.2,0.4, 0.8, 1.6, 3.2, 6.4, 12.8)
35
36
37 Resultados <- beta %>%  map_df(~sim_norm_pois(n=200, nsim=100,
     beta=.x))
38
39 Resultados <- Resultados %>% mutate(mu=factor(mu, ordered=TRUE)
     )
40
41 fig <- ggplot(Resultados, aes(x=n, y=p_valor, colour=mu))
42 fig + geom_smooth(se = FALSE)+coord_cartesian(ylim=c(0.0,0.6),
     xlim=c(2,100), expand=FALSE)+
43   geom_hline(yintercept = 0.05)+ xlab("Number of replications
       by sequence")+
44   ylab("P-value")
```

A.2. Simulation 2

```
1 library(PoisNor)
2 library(geepack)
3 library(gee)
4 library(tidyverse)
5 library(gridExtra)
6 library(splines)
7 library(mvtnorm)
8 library(geeM)
9 library(readxl)
10 library(tidyverse)
11 library(ggcorrplot)
12 library(plotly)
13 library(cowplot)
14 ##### Cross over 2X2####
15
```

```
16 nsim <- 10
17 nReplicas <- c(5,10,20,50,100)
18
19 ResulSimN <- NULL
20 for(nk in 1:length(nReplicas)){
21 tB <- seq(-2, 2, 0.1)
22 ResulSim <- data.frame(tB=gl(length(tB), nsim, labels = tB), Tw
     =NA, pW=NA)
23 for(j in tB){
24 Tw_total <- numeric(nsim)
25 pW_total <- numeric(nsim)
26 nperiod=2
27 mu0=0
28 Pe2=runif(1, 0, 1)
29 CarryA=runif(1, 0,1)
30 n1=nReplicas[nk]
31 n2=nReplicas[nk]
32 alpha=c(0.5)
33 N=diag(nperiod)
34 N[lower.tri(N)]=alpha
35 TV=N+t(N)  ### TV es la matriz de correlacion
36 diag(TV)<-1
37 TV=kronecker(diag(nperiod), TV)
38 datos=list()
39 #### Secuencia 1
40
41 lamvec1=c(exp(mu0-Pe2-j), exp(mu0+Pe2+j+CarryA))
42 cstar1 = cmat.star(no.pois=nperiod, no.norm=nperiod, TV,
     lamvec1)
43 lamvec2=c(exp(mu0-Pe2+j), exp(mu0+Pe2-j-CarryA))
44 cstar2 = cmat.star(no.pois=nperiod, no.norm=nperiod, TV,
     lamvec2)
45 for(i in 1:nsim){
46
47 mydata=genPoisNor(n=n1, no.norm=nperiod,  no.pois=nperiod, cmat
     .star=cstar1,
48                   lamvec1, sd.vec=rep(1,nperiod), mean.vec=rep
                        (1, nperiod))
```

```
49
50 datos[[1]]<-mydata
51 datos[[1]][2,1] <- ifelse(datos[[1]][2,1]==0, 1, datos
     [[1]][2,1])
52 datos[[1]][2,2] <- ifelse(datos[[1]][2,2]==0, 1, datos
     [[1]][2,2])
53 #### Secuencia 2
54
55
56 mydata=genPoisNor(n=n2, no.norm=nperiod,  no.pois=nperiod, cmat
     .star=cstar2,
57                   lamvec2, sd.vec=rep(1,nperiod), mean.vec=rep
                        (1, nperiod))
58 datos[[2]]<-mydata
59 datos[[2]][2,1] <- ifelse(datos[[2]][2,1]==0, 1, datos
     [[2]][2,1])
60 datos[[2]][2,2] <- ifelse(datos[[2]][2,2]==0, 1, datos
     [[2]][2,2])
61
62 #### Construcci?n de la base
63
64 treatment=c(rep("A",n1), rep("B",n2), rep("B",n2), rep("A",n2))
65 period=c(rep(1,n1), rep(2,n1),rep(1,n2), rep(2,n2))
66 Carry_A=c(rep(0,n1), rep(1,n1),rep(0,n2), rep(-1,n2))
67 count=c(datos[[1]][,1],datos[[1]][,2],datos[[2]][,1],datos
     [[2]][,1])
68 ind=c(rep(1:n1,2),rep(n1+1:n1+n2,2))
69 period=factor(period)
70 data_count=data.frame(count, ind, period, treatment, Carry_A)
71 data_count=data_count[order(data_count$ind, data_count$period)
     ,]
72 data_count
73 mod1=geeglm(count~treatment+period+Carry_A, id=data_count$ind,
      family="poisson",
74              corstr = "exchangeable", data=data_count)
75 a <- summary(mod1)
76 Tw_total[i] <- a$coefficients$Wald[2]
77 pW_total[i] <- a$coefficients$`Pr(>|W|)`[2]
```

```
print(i)
}
ResulSim[ResulSim$tB==paste(j),2:3]=cbind(Tw_total, pW_total)
print(j)
}
ResulSimN <- rbind(ResulSimN, ResulSim %>% mutate(nRep=
   nReplicas[nk]))
print(nk)
}


save(ResulSim, file = "Sim_2x2_Tw.RData")

setwd("~/Primer articulo")
load("Sim_2x2_Tw.RData")
ggplot(ResulSim,aes(y=pW, x=tB))+geom_boxplot()

resumen <- ResulSimN %>% group_by(nRep,tB) %>% mutate(DH=ifelse
   (pW<0.05,1,0)) %>%
  summarise(nDH=sum(DH)/100)%>% mutate(tB=as.numeric(as.
     character(tB)),
                                       Replications=factor(
                                          nRep,
                                        levels = c("5","10", "
                                           20", "50","100"),
                                        ordered = T))



SplineSmooth <- function(formula, data, weights, span = 0.5,
   ...) {
  pred <- smooth.spline(data$x, data$y, df = length(data$y)*
     span,...)$y
  # print(pred[1:10])
  model <- list(x = data$x, pred = pred)
  class(model) <- "my_smooth"
  model
}
```

```
109
110 predictdf.my_smooth <- function(model, xseq, se, level) {
111   data.frame(x = model$x, y = model$pred)
112 }
113
114 p <- resumen %>%
115   ggplot(aes(x=tB, y=nDH, colour=Replications))+
116   geom_smooth(method = "SplineSmooth", method.args = list(span
        = 0.001))+
117   coord_cartesian(ylim=c(0,1))+ xlab("True value of treatment
        effect")+
118   ylab("Power of hipotesis")
119
120 p
121
122 p1 <-  ResulSimN %>% filter(tB=="-2") %>%
123   mutate(Replications=factor(nRep, levels = c("5","10", "20", "
        50","100"),ordered = T)) %>%
124   ggplot(aes(x=Tw,group=Replications, fill=Replications))+geom_
        density(alpha=0.5)
125
126 p1
127
128 p2 <-  ResulSimN %>% filter(tB=="-1") %>%
129   mutate(Replications=factor(nRep, levels = c("5","10", "20", "
        50","100"),ordered = T)) %>%
130   ggplot(aes(x=Tw,group=Replications, fill=Replications))+geom_
        density(alpha=0.5)
131
132 p2
133
134
135 p3 <-  ResulSimN %>% filter(tB=="0") %>%
136   mutate(Replications=factor(nRep, levels = c("5","10", "20", "
        50","100"),ordered = T)) %>%
137   ggplot(aes(x=Tw,group=Replications, fill=Replications))+geom_
        density(alpha=0.5)
138
```

```
139 p3
140
141 p4 <-  ResulSimN %>% filter(tB=="1") %>%
142   mutate(Replications=factor(nRep, levels = c("5","10", "20", "
        50","100"),ordered = T)) %>%
143   ggplot(aes(x=Tw,group=Replications, fill=Replications))+geom_
        density(alpha=0.5)
144
145 p4
146
147
148 p5 <-  ResulSimN %>% filter(tB=="2") %>%
149   mutate(Replications=factor(nRep, levels = c("5","10", "20", "
        50","100"),ordered = T)) %>%
150   ggplot(aes(x=Tw,group=Replications, fill=Replications))+geom_
        density(alpha=0.5)
151
152 p5
153
154 grid.arrange(p1, p2, p3, p4)
155
156
157 ggdraw()+draw_plot(p1, 0, 0.5,0.5,0.5)+
158   draw_plot(p2, 0.5, 0.5,0.5,0.5)+
159   draw_plot(p3, 0, 0,0.5,0.5)+
160   draw_plot(p4, 0.5,0,0.5,0.5)+
161   draw_plot_label(c("a)","b)", "c)", "d)"), c(0,0.5, 0, 0.5), c
        (1,1, 0.5, 0.5))
```

B. Codes in R chapter 3

B.1. Simulation of Kroncker Crossover

```
1  ######################################
2
3  library(gee)
4  library(MASS)
5  library(SimDesign)
6
7  library(SimCorMultRes)
8
9  ######################################
10 ##### Parameters ####################
11 ######################################
12
13 #### Simulated a unstructured #####
14 ### corr matrix  psi  #############
15
16
17
18 SinEstructura <- function(nperiod=3){
19   psi_m = matrix(0, nrow=nperiod, ncol=nperiod)
20 condicion = TRUE
21 while(condicion == TRUE){
22 for(i in 1:nperiod){
23   for(j in i:nperiod){
24     a = runif(1, -0.4,0.8)
25     psi_m[i,j]= a
26     psi_m[j,i] = a
27   }
28 }
29 diag(psi_m)=1
```

```
condicion = min(eigen(psi_m)$values)<0
}
return(psi_m)
}

SinEstructura(5)

################################
####### Una intercambiable ###

Intercam <- function(ntimes=10, alpha=0.5){
  psi_m = toeplitz(c(1, rep(alpha, ntimes-1)))
  return(psi_m)
}

Intercam(10,0.5)

####################
## AR1 ##############
Ar1 <- function(ntimes=10, alpha=0.5){
  psi_m = matrix(0, nrow=ntimes, ncol=ntimes)
  for(i in 1:ntimes){
    for(j in 1:ntimes){
      psi_m[i,j]=alpha^(abs(i-j))
    }
  }
  return(psi_m)
}

Ar1(10,-0.9)

##################################
#### Simular una normal ##########
##################################
Sim_Norm_Kron <- function(nperiod, ntimes, nseq, nind, corst =c
    ("Exch", "Ar1", "Unst"), alpha=0.5,
                          betas_p = NULL, betas_T = NULL,
                             tratamiento=1,
```

```
66                              disenno= c("2x2", "PE", "SB"),
                                   intercepto = 2){
67    P = nperiod
68    L = ntimes
69    if(corst == "Exch"){
70      RR = kronecker(SinEstructura(nperiod), Intercam(ntimes,
          alpha))
71    }else if (corst == "Ar1") {
72      RR = kronecker(SinEstructura(nperiod), Ar1(ntimes, alpha))
73    }else{RR = kronecker(SinEstructura(nperiod), SinEstructura(
        ntimes)) }
74    ## Periodo
75    if(is.null(betas_p)){
76      p_eff=rnorm(P-1)
77    }else{p_eff = betas_p}
78    ### Tratamiento
79    if(is.null(betas_T)){
80      time_effect=rnorm(L-1)
81    }else{time_effect = betas_T}
82    ## Secuencia
83    periodo=kronecker(matrix(1, ncol=1, nrow = nseq*nind),
84                      kronecker(matrix(1:P, ncol=1, nrow=P),
85                                matrix(1, ncol=1, nrow=L)))
86
87    tiempo=kronecker(matrix(1, ncol=1, nrow = nseq*nind),
88                     kronecker(matrix(1, ncol=1, nrow=P),
89                               matrix(1:L, ncol=1, nrow=L)))
90    secuencia=kronecker(matrix(1:nseq, ncol=1, nrow = nseq),
91                        kronecker(matrix(1, ncol=1, nrow=P*nind),
92                                  matrix(1, ncol=1, nrow=L)))
93    id=as.numeric(gl(nind*nseq,nperiod*ntimes))
94    datos=numeric(nind*nseq*nperiod*ntimes)
95    df=data.frame(resp=datos, Periodo=factor(periodo),Secuencia=
        factor(secuencia),
96                  Tiempo=factor(tiempo), Sujeto=id )
97    df=df[order(df$Sujeto,df$Periodo, df$Tiempo), ]
98    ### Estructura del tratamiento
99
```

```
100   df["Tratamiento"]=0
101   if(disenno=="2x2"){
102     ## AB, BA
103     df["Tratamiento"][df["Secuencia"]==1 & df["Periodo"]==2] =1
104
105     df["Tratamiento"][df["Secuencia"]==2 & df["Periodo"]==1] =1
106   }else if(disenno =="PE") {
107     ## ABB, BAA
108     df["Tratamiento"][df["Secuencia"]==1 & df["Periodo"]==2] =1
109
110     df["Tratamiento"][df["Secuencia"]==1 & df["Periodo"]==3] =1
111
112
113     df["Tratamiento"][df["Secuencia"]==2 & df["Periodo"]==1] =1
114   }else{
115     ## ABA, BAB
116     df["Tratamiento"][df["Secuencia"]==1 & df["Periodo"]==2] =1
117
118     df["Tratamiento"][df["Secuencia"]==2 & df["Periodo"]==1] =1
119     df["Tratamiento"][df["Secuencia"]==2 & df["Periodo"]==3] =1
120   }
121
122   XX = model.matrix(lm(resp~factor(Periodo)+ factor(Tratamiento
        )+factor(Tiempo), data=df))
123   beta_total = c(intercepto, p_eff, tratamiento, time_effect)
124   df["media"]= XX%*%beta_total
125
126   for(i in 1:(nind*nseq)){
127     df["resp"][df["Sujeto"]==i] = rmvnorm(1, df["media"][df["
          Sujeto"]==i], RR)
128   }
129   return(list(datos=df, corr=RR))
130 }
131
132
133 ### Simular binomiales ###
134 ### Basado en
135 # https://cran.r-project.org/web/packages/SimCorMultRes/
```

```
    vignettes/SimCorMultRes.html
136 ####################
137
138
139
140 Sim_Bin_Kron <- function(nperiod, ntimes, nseq, nind, corst =c(
      "Exch", "Ar1", "Unst"), alpha=0.5,
141                           betas_p = NULL, betas_T = NULL,
                                 tratamiento=1,
142                           disenno= c("2x2", "PE", "SB"),
                                 intercepto = 2){
143   P = nperiod
144   L = ntimes
145   if(corst == "Exch"){
146     RR = kronecker(SinEstructura(nperiod), Intercam(ntimes,
          alpha))
147   }else if (corst == "Ar1") {
148     RR = kronecker(SinEstructura(nperiod), Ar1(ntimes, alpha))
149   }else{RR = kronecker(SinEstructura(nperiod), SinEstructura(
        ntimes)) }
150   ## Periodo
151   if(is.null(betas_p)){
152     p_eff=abs(rnorm(P-1,sd = 0.1))
153   }else{p_eff = betas_p}
154   ### Tratamiento
155   if(is.null(betas_T)){
156     time_effect=rnorm(L-1)
157   }else{time_effect = betas_T}
158   ## Secuencia
159   periodo=kronecker(matrix(1, ncol=1, nrow = nseq*nind),
160                     kronecker(matrix(1:P, ncol=1, nrow=P),
161                               matrix(1, ncol=1, nrow=L)))
162
163   tiempo=kronecker(matrix(1, ncol=1, nrow = nseq*nind),
164                    kronecker(matrix(1, ncol=1, nrow=P),
165                              matrix(1:L, ncol=1, nrow=L)))
166   secuencia=kronecker(matrix(1:nseq, ncol=1, nrow = nseq),
167                       kronecker(matrix(1, ncol=1, nrow=P*nind),
```

```
168                                   matrix(1, ncol=1, nrow=L)))
169   id=as.numeric(gl(nind*nseq,nperiod*ntimes))
170   datos=numeric(nind*nseq*nperiod*ntimes)
171   df=data.frame(resp=datos, Periodo=factor(periodo),Secuencia=
         factor(secuencia),
172                 Tiempo=factor(tiempo), Sujeto=id )
173   df=df[order(df$Sujeto,df$Periodo, df$Tiempo), ]
174   ### Estructura del tratamiento
175
176   df["Tratamiento"]=0
177   if(disenno=="2x2"){
178     ## AB, BA
179     df["Tratamiento"][df["Secuencia"]==1 & df["Periodo"]==2] =1
180
181     df["Tratamiento"][df["Secuencia"]==2 & df["Periodo"]==1] =1
182   }else if(disenno =="PE") {
183     ## ABB, BAA
184     df["Tratamiento"][df["Secuencia"]==1 & df["Periodo"]==2] =1
185
186     df["Tratamiento"][df["Secuencia"]==1 & df["Periodo"]==3] =1
187
188
189     df["Tratamiento"][df["Secuencia"]==2 & df["Periodo"]==1] =1
190   }else{
191     ## ABA, BAB
192     df["Tratamiento"][df["Secuencia"]==1 & df["Periodo"]==2] =1
193
194     df["Tratamiento"][df["Secuencia"]==2 & df["Periodo"]==1] =1
195     df["Tratamiento"][df["Secuencia"]==2 & df["Periodo"]==3] =1
196   }
197   XX = model.matrix(lm(resp~Periodo+Tratamiento+Tiempo, data=df
         ))
198   beta_total = c(intercepto, p_eff, tratamiento, time_effect)
199   df["media"]= XX%*%beta_total
200   form <- formula(lm(`(Intercept)`~., data=as.data.frame(XX)))
201   df1 <- rbin(clsize = nperiod*ntimes,intercepts = intercepto,
         betas=beta_total[-1],
202                   xformula =form[c(1,3)],
```

```
                    xdata= XX[,-1],
                    cor.matrix = RR, link="logit")
      df["resp"] <- df1$simdata$y
      return(list(datos=df, corr=RR))
}


## Binomiales
datos <- Sim_Bin_Kron(nperiod = 3, ntimes=10, nseq=2, nind =15,
    corst = "Ar1", alpha = 0.2, disenno = "PE")

## normales
datos <- Sim_Norm_Kron(nperiod = 3, ntimes=5, nseq=2, nind =15,
    corst = "Ar1", alpha = 0.2, disenno = "PE")
```

B.2. Functions to estimate parameters

```
##############################################
#### Para estimar la kronecker ############
##############################################

library(gee)
library(corrplot)

## residuuales
## https://www.sfu.ca/sasdoc/sashtml/stat/chap29/sect38.htm

simulacion <- Sim_Bin_Kron(nperiod = 3, ntimes=10, nseq=2, nind
    =15, corst = "Ar1",
                       alpha = 0.8, disenno = "PE", tratamiento
                          = -2, intercepto = 1,
                       betas_p = runif(2,0.2, 0.3), betas_T =
                          rep(0,9))


##### son tres pasos
## Estimar beta con un R_alpha dado
```

```
## Estimar alpha dentro con GEE
### estimar Psi con lo anterior
### Actualizar y repetir
### binomial

obtalpha <- function(residu, datos, tipo=c("Ar1", "Exch")){
  if(tipo=="Ar1"){
    K1 = sum(table(datos$Per_id)-1)
    n = length(table(datos$Per_id))
    suma=0
    for(i in 1:n){
      residu_temp = residu[datos$Per_id==i]
      n_temp =length(residu_temp)
      suma = suma + sum(residu_temp[-n_temp]*residu_temp[-1])
    }
    alpha = suma/((K1-pp)*phi)
  }else if(tipo=="Exch"){
    K1 =sum((table(datos$Per_id)-1)*(table(datos$Per_id)))
    n = length(table(datos$Per_id))
    suma=0
    for(i in 1:n){
      residu_temp = residu[datos$Per_id==i]
      a_temp = outer(residu_temp, residu_temp)
      suma = suma + sum(a_temp[lower.tri(a_temp)])
    }
    alpha = suma/((K1-pp)*phi)
  }
  return(alpha)
}


############################
## obtener Psi #############

obtPsi <- function(residu, datos, Ralpha){
  datos["residu"] <- residu
  PP <- length(table(datos$Periodo))
  LL <- max(table(datos$Per_id))
```

```
56   n_suj <- length(table(datos$Sujeto))
57   Psi <- matrix(0, PP, PP)
58   ### las medias por periodo
59   r_media <- matrix(0, PP, LL)
60   for(ii in 1:PP){
61     media_temp = rep(0, LL)
62     datos_temp = datos[datos$Periodo==ii,]
63     for(jj in 1:n_suj){
64       media_temp = media_temp + datos_temp[datos_temp$Sujeto==
           jj, "residu"]
65     }
66     r_media[ii,] <- media_temp/(n_suj)
67   }
68   ###### cada Psi
69   for(ii in 1:PP){
70     for(jj in 1:PP){
71       suma_temp = 0
72       for( kk in 1:n_suj){
73         datos_temp1 = datos[datos$Periodo==ii,]
74         datos_temp2 = datos[datos$Periodo==jj,]
75         residu1 = datos_temp1[datos_temp1$Sujeto==kk, "residu"]
76         residu2 = datos_temp2[datos_temp2$Sujeto==kk, "residu"]
77         AAA = solve(Ralpha) %*%((residu1- r_media[ii,])%*% t((
             residu2- r_media[jj,])))
78         suma_temp  = suma_temp  + sum(diag(AAA))
79       }
80       Psi[ii,jj] = suma_temp /n_suj
81     }
82   }
83   a=Psi
84   return(list(Psi=(diag(1/sqrt(diag(a))))%*% a %*% (diag(1/sqrt
       (diag(a)))), Ra = Ralpha, Psi1=Psi))
85 }
86
87
88 ###
89 simulacion <- Sim_Norm_Kron(nperiod = 3, ntimes=5, nseq=2, nind
     =100, corst ="Exch",
```

```
90                              alpha = 0.8, disenno = "PE",
                                   tratamiento = -2, intercepto = 1,
91                              betas_p = runif(2,0.2, 0.3), betas_T
                                    = rep(0,4))
92
93 datos = simulacion$datos
94 datos["Per_id"]=as.numeric(
95   as.factor(paste(datos$Sujeto, datos$Periodo)))
96
97 formula = resp ~ factor(Periodo)+ Tratamiento
98 corrplot(simulacion$corr)
99
100 init=gee(formula,
101           data=datos, id=Sujeto)
102 pp <- length(init$coefficients)
103 #residu <- init$residuuals/sqrt(init$fitted.values-init$fitted.
      values^2)
104 residu <- init$residuals
105 phi <- sum(residu^2)/(length(residu)-pp)
106
107
108 Ralpha = Intercam(max(table(datos$Per_id)), obtalpha(residu,
      datos, tipo="Ar1"))
109 matrices <- obtPsi(residu/sqrt(phi), datos, Ralpha )
110 RR <- kronecker(matrices$Psi, matrices$Ra)
111 beta_init <- init$coefficients
112 n_iter = 4
113 diferencia <- numeric(n_iter)
114 for(iter in 1:n_iter){
115   modelo = gee(formula,
116                data=datos, id=Sujeto, corstr = "fixed",
117                R = RR, b = beta_init, maxiter = 1, tol=100)
118   diferencia[iter] = sum((beta_init-modelo$coefficients)^2)
119   beta_init = modelo$coefficients
120   #residu <- modelo$residuuals/sqrt(modelo$fitted.values-modelo
        $fitted.values^2)
121   residu <- modelo$residuals
122   phi <- sum(residu^2)/(length(residu)-pp)
```

```
123   Ralpha = Intercam(max(table(datos$Per_id)), obtalpha(residu,
         datos, tipo="Ar1"))
124   matrices <- obtPsi(residu/sqrt(phi), datos, Ralpha )
125   RR <- kronecker(matrices$Psi, matrices$Ra)
126   if(is.complex(eigen(RR)$values)){
127     RR = diag(1,nrow(RR))
128     beep(7)
129   }
130 }
131 plot(diferencia)
132
133 corrplot(RR)
134 corrplot(simulacion$corr)
135
136
137 round(RR,3)
138 round(simulacion$corr,3)
```

B.3. Code to analyze cow data

```
1  rm(list=ls())
2
3  library(data.table)
4  library(MASS)
5  library(stats)
6  library(tidyverse)
7  library(gee)
8  setwd("~/Doctorado/SegundoArticulo/pone.0232943.s001/data")
9  # GLMs of perception based variables
10
11 #Load data
12 anasurv <- fread("./cleansurvey.csv")
13
14 # Environmental variables
15
16 env <- fread( "./sensor_readings.csv")
17
```

```
18 # Commits
19 commits <- fread("./git_commits.csv")
20
21
22 # Occupancy data
23 occ <- fread("./desk_occupancy.csv")
24
25 occup_hora <- occ %>% group_by(area, wave, hours, sensor, prev_
       wave) %>%
26   summarise(n=n(), n_occup=sum(reading)) %>% arrange(sensor,
         wave) %>%
27   mutate(prop = n_occup/n, sensorid = as.numeric(as.factor(
         sensor)))
28
29
30
31
32 #### Modelo con GLMM
33
34 load(file="/content/Hora.RData")
35 occup_hora["Per_id"]= as.numeric(
36   as.factor(paste(occup_hora$sensorid, occup_hora$wave)))
37 datos = occup_hora
38
39 formula = cbind(n_occup, n-n_occup) ~area+wave+factor(hours)+
       prev_wave
40
41
42 mod1 = gee(formula, family= binomial,
43            data=datos, id=Per_id)
44
45 mod2 = gee(formula, family= binomial,
46            data=datos, id=sensorid)
47
48 mod3 = gee(formula, family= binomial,
49            data=datos, id=sensorid, corstr = "exchangeable")
50
51 mod4 = gee(formula, family= binomial,
```

```
52            data=datos, id=Per_id, corstr = "AR-M")
53
54
55
56 init=gee(formula, family= binomial,
57          data=datos, id=sensorid)
58 pp <- length(init$coefficients)
59 residu <- init$residuals/sqrt(init$fitted.values-init$fitted.
     values^2)
60 #residu <- init$residuals
61 phi <- sum(residu^2)/(length(residu)-pp)
62
63
64 Ralpha = Intercam(max(table(datos$Per_id)), obtalpha(residu,
     datos, tipo="Exch"))
65 matrices <- obtPsi(residu/sqrt(phi), datos, Ralpha )
66 RR <- kronecker(matrices$Psi, matrices$Ra)
67 if(is.complex(eigen(RR)$values) | min(eigen(RR)$values)<0){
68   RR = diag(1,nrow(RR))
69   beep(7)
70 }
71 beta_init <- init$coefficients
72 n_iter = 4
73 diferencia <- numeric(n_iter)
74 for(iter in 1:n_iter){
75   modelo = gee(formula, family= binomial,
76                data=datos, id=sensorid, corstr = "fixed",
77                R = RR, b = beta_init, maxiter = 1, tol=100)
78   diferencia[iter] = sum((beta_init-modelo$coefficients)^2)
79   beta_init = modelo$coefficients
80   residu <- modelo$residuals/sqrt(modelo$fitted.values-modelo$
       fitted.values^2)
81   #residu <- modelo$residuals
82   phi <- sum(residu^2)/(length(residu)-pp)
83   Ralpha = Intercam(max(table(datos$Per_id)), obtalpha(residu,
       datos, tipo="Exch"))
84   matrices <- obtPsi(residu/sqrt(phi), datos, Ralpha )
85   RR <- kronecker(matrices$Psi, matrices$Ra)
```

```
86   if(is.complex(eigen(RR)$values) | min(eigen(RR)$values)<0){
87     RR = diag(1,nrow(RR))
88     beep(7)
89   }
90 }
91
92 mod5 = modelo
93
94
95 ## El AR 1
96 init=gee(formula, family= binomial,
97          data=datos, id=sensorid)
98 pp <- length(init$coefficients)
99 residu <- init$residuals/sqrt(init$fitted.values-init$fitted.
     values^2)
100 #residu <- init$residuals
101 phi <- sum(residu^2)/(length(residu)-pp)
102
103 Ralpha = Ar1(max(table(datos$Per_id)), obtalpha(residu, datos,
     tipo="Ar1"))
104 matrices <- obtPsi(residu/sqrt(phi), datos, Ralpha )
105 RR <- kronecker(matrices$Psi, matrices$Ra)
106 if(is.complex(eigen(RR)$values) | min(eigen(RR)$values)<0){
107   RR = diag(1,nrow(RR))
108   beep(7)
109 }
110 beta_init <- init$coefficients
111 n_iter = 4
112 diferencia <- numeric(n_iter)
113 for(iter in 1:n_iter){
114   modelo = gee(formula, family= binomial,
115                data=datos, id=sensorid, corstr = "fixed",
116                R = RR, b = beta_init, maxiter = 1, tol=100)
117   diferencia[iter] = sum((beta_init-modelo$coefficients)^2)
118   beta_init = modelo$coefficients
119   residu <- modelo$residuals/sqrt(modelo$fitted.values-modelo$
        fitted.values^2)
120   #residu <- modelo$residuals
```

```
121   phi <- sum(residu^2)/(length(residu)-pp)
122   Ralpha = Ar1(max(table(datos$Per_id)), obtalpha(residu, datos
        , tipo="Ar1"))
123   matrices <- obtPsi(residu/sqrt(phi), datos, Ralpha )
124   RR <- kronecker(matrices$Psi, matrices$Ra)
125   if(is.complex(eigen(RR)$values) | min(eigen(RR)$values)<0){
126     RR = diag(1,nrow(RR))
127     beep(7)
128   }
129 }
130 mod6 = modelo
131
132
133 QIC=numeric(6)
134
135 QIC[1]<--2*sum(mod1$y * log(mod1$fitted.values/(1-mod1$fitted.
      values))+log(1-mod1$fitted.values))+
136   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod1$naive
        .variance))
137 QIC[2]<--2*sum(mod2$y * log(mod2$fitted.values/(1-mod2$fitted.
      values))+log(1-mod2$fitted.values))+
138   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod2$
        robust.variance))
139 QIC[3]<--2*sum(mod3$y * log(mod3$fitted.values/(1-mod3$fitted.
      values))+log(1-mod3$fitted.values))+
140   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod3$
        robust.variance))
141 QIC[4]<--2*sum(mod4$y * log(mod4$fitted.values/(1-mod4$fitted.
      values))+log(1-mod4$fitted.values))+
142   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod4$
        robust.variance))
143 QIC[5]<--2*sum(mod5$y * log(mod5$fitted.values/(1-mod5$fitted.
      values))+log(1-mod5$fitted.values))+
144   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod5$
        robust.variance))
145 QIC[6]<--2*sum(mod6$y * log(mod6$fitted.values/(1-mod6$fitted.
      values))+log(1-mod6$fitted.values))+
146   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod6$
```

```
    robust.variance))
```

B.4. Code to analyze arterial pressure data

```
library(gee)
library(xtable)
library(readxl)
library(tidyverse)
library(ggcorrplot)
library(corrplot)
library(geepack)
library(MASS)
source("~/Doctorado/SegundoArticulo/Sim_general_CrosKron.R")

source("~/Doctorado/SegundoArticulo/Estimacion_funciones.R")

Datos_con_base <- read_excel("Doctorado/Datos_con_base.xlsx")
library(nlme)
Datos_con_base=Datos_con_base[order(Datos_con_base$Sujeto,
                                    Datos_con_base$Periodo,
                                       Datos_con_base$Tiempo),
                                       ]

Datos_con_base["Per_id"]=as.numeric(
  as.factor(paste(Datos_con_base$Sujeto, Datos_con_base$Periodo
    )))


formula=Presion~Base+Tiempo+factor(Periodo)+Tratamiento+CarryA+
    CarryB+I(Tiempo^2)

datos = Datos_con_base %>% data.frame()
datos["resp"] = datos$dprod

init=gee(formula,
         data=datos, id=Sujeto, corstr = "AR-M")
pp <- length(init$coefficients)
```

```
#resid <- init$residuals/sqrt(init$fitted.values-init$fitted.values^2)
resid <- init$residuals
phi <- sum(resid^2)/(length(resid)-pp)

init

Ralpha = Intercam(max(table(datos$Per_id)), obtalpha(resid, datos, tipo="Exch"))

matrices <- obtPsi(resid, datos, Ralpha )

corrplot(kronecker(matrices$Psi, matrices$Ra))


Ralpha = Intercam(max(table(datos$Per_id)), obtalpha(resid, datos, tipo="Exch"))
RR <- kronecker(matrices$Psi, matrices$Ra)
beta_init <- init$coefficients

n_iter = 20
diferencia <- numeric(n_iter)
for(iter in 1:n_iter){
  modelo = gee(formula,
               data=datos, id=Sujeto, corstr = "fixed",
               R = RR, b = beta_init, maxiter = 1, tol=100)
  diferencia[iter] = sum((beta_init-modelo$coefficients)^2)
  beta_init = modelo$coefficients
  pp=2
  #resid <- modelo$residuals/sqrt(modelo$fitted.values-modelo$fitted.values^2)
  resid <- modelo$residuals
  phi <- sum(resid^2)/(length(resid)-pp)
  matrices <- obtPsi(resid, datos, Ralpha )
  Ralpha = Intercam(max(table(datos$Per_id)), obtalpha(resid, datos, tipo="Exch"))
  RR <- kronecker(matrices$Psi, matrices$Ra)
}
```

```
63
64
65 plot(diferencia)
66 a=summary(modelo)
67 a=a$coefficients
68 pvalor = 1-pnorm(abs(a[,5]))
69 a=cbind(a, pvalor)
70 xtable(a[,c(1,4,5,6)], digits=4)
71
72
73 modelo1 = gee(formula,
74               data=datos, id=Sujeto, corstr = "exchangeable")
75
76 summary(modelo1)
77
78
79 corrplot(modelo1$working.correlation)
80 corrplot(modelo$working.correlation)
81
82
83
84
85 mod1=gee(formula,data=datos, id=Per_id,corstr="independence")
86 mod2=gee(formula,data=datos, id=Per_id,corstr="AR-M")
87 mod3=gee(formula,data=datos, id=Sujeto,corstr="AR-M")
88 mod4=gee(formula,data=datos, id=Sujeto,corstr="exchangeable")
89 mod5=gee(formula,data=datos, id=Sujeto,corstr="fixed",R =RR)
90
91 QIC=numeric(5)
92
93 QIC[1]<-sum((mod1$residuals)^2)+
94   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod1$naive
      .variance))
95 QIC[2]<-sum((mod2$residuals)^2)+
96   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod2$
      robust.variance))
97 QIC[3]<-sum((mod3$residuals)^2)+
98   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod3$
```

```
      robust.variance))
99 QIC[4]<-sum((mod4$residuals)^2)+
100   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod4$
      robust.variance))
101 QIC[5]<-sum((mod5$residuals)^2)+
102   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod5$
      robust.variance))
103
104
105 QIC
106
107
108 summary(modelo)
```

B.5. Code to analyze occupancy data

```
1  rm(list=ls())
2
3  library(data.table)
4  library(MASS)
5  library(stats)
6  library(tidyverse)
7  library(gee)
8  setwd("~/Doctorado/SegundoArticulo/pone.0232943.s001/data")
9  # GLMs of perception based variables
10
11 #Load data
12 anasurv <- fread("./cleansurvey.csv")
13
14 # Environmental variables
15
16 env <- fread( "./sensor_readings.csv")
17
18 # Commits
19 commits <- fread("./git_commits.csv")
20
21
```

```r
# Occupancy data
occ <- fread("./desk_occupancy.csv")

occup_hora <- occ %>% group_by(area, wave, hours, sensor, prev_wave) %>%
  summarise(n=n(), n_occup=sum(reading)) %>% arrange(sensor, wave) %>%
  mutate(prop = n_occup/n, sensorid = as.numeric(as.factor(sensor)))



#### Modelo con GLMM

load(file="/content/Hora.RData")
occup_hora["Per_id"]= as.numeric(
  as.factor(paste(occup_hora$sensorid, occup_hora$wave)))
datos = occup_hora

formula = cbind(n_occup, n-n_occup) ~area+wave+factor(hours)+prev_wave


mod1 = gee(formula, family= binomial,
           data=datos, id=Per_id)

mod2 = gee(formula, family= binomial,
           data=datos, id=sensorid)

mod3 = gee(formula, family= binomial,
           data=datos, id=sensorid, corstr = "exchangeable")

mod4 = gee(formula, family= binomial,
           data=datos, id=Per_id, corstr = "AR-M")
```

```
56 init=gee(formula, family= binomial,
57           data=datos, id=sensorid)
58 pp <- length(init$coefficients)
59 residu <- init$residuals/sqrt(init$fitted.values-init$fitted.
     values^2)
60 #residu <- init$residuals
61 phi <- sum(residu^2)/(length(residu)-pp)
62
63
64 Ralpha = Intercam(max(table(datos$Per_id)), obtalpha(residu,
     datos, tipo="Exch"))
65 matrices <- obtPsi(residu/sqrt(phi), datos, Ralpha )
66 RR <- kronecker(matrices$Psi, matrices$Ra)
67 if(is.complex(eigen(RR)$values) | min(eigen(RR)$values)<0){
68   RR = diag(1,nrow(RR))
69   beep(7)
70 }
71 beta_init <- init$coefficients
72 n_iter = 4
73 diferencia <- numeric(n_iter)
74 for(iter in 1:n_iter){
75   modelo = gee(formula, family= binomial,
76                data=datos, id=sensorid, corstr = "fixed",
77                R = RR, b = beta_init, maxiter = 1, tol=100)
78   diferencia[iter] = sum((beta_init-modelo$coefficients)^2)
79   beta_init = modelo$coefficients
80   residu <- modelo$residuals/sqrt(modelo$fitted.values-modelo$
       fitted.values^2)
81   #residu <- modelo$residuals
82   phi <- sum(residu^2)/(length(residu)-pp)
83   Ralpha = Intercam(max(table(datos$Per_id)), obtalpha(residu,
       datos, tipo="Exch"))
84   matrices <- obtPsi(residu/sqrt(phi), datos, Ralpha )
85   RR <- kronecker(matrices$Psi, matrices$Ra)
86   if(is.complex(eigen(RR)$values) | min(eigen(RR)$values)<0){
87     RR = diag(1,nrow(RR))
88     beep(7)
89   }
```

```
}

mod5 = modelo


## El AR 1
init=gee(formula, family= binomial,
         data=datos, id=sensorid)
pp <- length(init$coefficients)
residu <- init$residuals/sqrt(init$fitted.values-init$fitted.
   values^2)
#residu <- init$residuals
phi <- sum(residu^2)/(length(residu)-pp)

Ralpha = Ar1(max(table(datos$Per_id)), obtalpha(residu, datos,
   tipo="Ar1"))
matrices <- obtPsi(residu/sqrt(phi), datos, Ralpha )
RR <- kronecker(matrices$Psi, matrices$Ra)
if(is.complex(eigen(RR)$values) | min(eigen(RR)$values)<0){
  RR = diag(1,nrow(RR))
  beep(7)
}
beta_init <- init$coefficients
n_iter = 4
diferencia <- numeric(n_iter)
for(iter in 1:n_iter){
  modelo = gee(formula, family= binomial,
               data=datos, id=sensorid, corstr = "fixed",
               R = RR, b = beta_init, maxiter = 1, tol=100)
  diferencia[iter] = sum((beta_init-modelo$coefficients)^2)
  beta_init = modelo$coefficients
  residu <- modelo$residuals/sqrt(modelo$fitted.values-modelo$
     fitted.values^2)
  #residu <- modelo$residuals
  phi <- sum(residu^2)/(length(residu)-pp)
  Ralpha = Ar1(max(table(datos$Per_id)), obtalpha(residu, datos
     , tipo="Ar1"))
  matrices <- obtPsi(residu/sqrt(phi), datos, Ralpha )
```

```
124   RR <- kronecker(matrices$Psi, matrices$Ra)
125   if(is.complex(eigen(RR)$values) | min(eigen(RR)$values)<0){
126     RR = diag(1,nrow(RR))
127     beep(7)
128   }
129 }
130 mod6 = modelo
131
132
133 QIC=numeric(6)
134
135 QIC[1]<--2*sum(mod1$y * log(mod1$fitted.values/(1-mod1$fitted.
      values))+log(1-mod1$fitted.values))+
136   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod1$naive
        .variance))
137 QIC[2]<--2*sum(mod2$y * log(mod2$fitted.values/(1-mod2$fitted.
      values))+log(1-mod2$fitted.values))+
138   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod2$
        robust.variance))
139 QIC[3]<--2*sum(mod3$y * log(mod3$fitted.values/(1-mod3$fitted.
      values))+log(1-mod3$fitted.values))+
140   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod3$
        robust.variance))
141 QIC[4]<--2*sum(mod4$y * log(mod4$fitted.values/(1-mod4$fitted.
      values))+log(1-mod4$fitted.values))+
142   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod4$
        robust.variance))
143 QIC[5]<--2*sum(mod5$y * log(mod5$fitted.values/(1-mod5$fitted.
      values))+log(1-mod5$fitted.values))+
144   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod5$
        robust.variance))
145 QIC[6]<--2*sum(mod6$y * log(mod6$fitted.values/(1-mod6$fitted.
      values))+log(1-mod6$fitted.values))+
146   2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod6$
        robust.variance))
```

C. Codes in R chapter 4

C.1. Simulation

```
library(beepr)
library(mvtnorm)
library(geeM)
library(gee)
library(readxl)
library(tidyverse)
library(ggcorrplot)
library(geepack)
library(MASS)
library(corrplot)
library(plotly)
library(MASS)
library(splines)

Sim_kron_cross=function(betakk=0.5, P=3,L=5,nseq=2 ,nind=3,
    corr=c("ar1","exch"), distr=c("normal", "poisson")){
  if(corr=="ar1"){
    condicion=TRUE
    while(condicion==TRUE){
      alpha1=runif(1, -1,1); alpha2=runif(1,-1,1)
      Phi=diag(rep(1,P))
      R_alpha=matrix(0, ncol=L, nrow=L)
      for(i in 1:L){
        for(j in 1:L){
          R_alpha[i,j]=alpha2^(abs(i-j))
        }
      }

      RR=kronecker(Phi, R_alpha)
```

```
29        condicion<-min(eigen(RR)$values)<0
30      }
31    }else{
32      condicion=TRUE
33      while(condicion==TRUE){
34        alpha1=runif(1, -1,0.5); alpha2=runif(1,-1,1)
35        Phi=diag(rep(1,P))
36        R_alpha=matrix(0, ncol=L, nrow=L)
37        for(i in 1:L){
38          for(j in 1:L){
39            R_alpha[i,j]=alpha2^(1-as.numeric(i==j))
40          }
41        }
42
43        RR=kronecker(Phi, R_alpha)
44        condicion<-min(eigen(RR)$values)<0
45      }
46    }
47    aaaa=rnorm(1,0, 1)
48    p_eff=c(0,3,3)
49    seq_effect=c(aaaa, aaaa+betakk)
50    ttt=(1:L)
51    time_effect=(cos(ttt)-cos(ttt)[1])
52    carry_effect=(sin(ttt)-sin(ttt)[1])
53    periodo=kronecker(matrix(1, ncol=1, nrow = nseq*nind),
54                      kronecker(matrix(1:P, ncol=1, nrow=P),
55                                matrix(1, ncol=1, nrow=L)))
56
57    tiempo=kronecker(matrix(1, ncol=1, nrow = nseq*nind),
58                     kronecker(matrix(1, ncol=1, nrow=P),
59                               matrix(1:L, ncol=1, nrow=L)))
60    secuencia=kronecker(matrix(1:nseq, ncol=1, nrow = nseq),
61                        kronecker(matrix(1, ncol=1, nrow=P*nind),
62                                  matrix(1, ncol=1, nrow=L)))
63    id=gl(nseq*nind, P*L)
64    datos=numeric(length(periodo))
65    df=data.frame(resp=datos,mu=datos, Periodo=periodo,Secuencia=
        secuencia,Tiempo=tiempo, Sujeto=id )
```

```
66   df = df %>% mutate(Carry=ifelse((Periodo==2 & Secuencia==2)|(
       Periodo==3 & Secuencia==1),1,0))
67   for(i in 1:nrow(df)){
68     df$mu[i]<-seq_effect[df[, "Secuencia"][i]]+p_eff[df[, "
         Periodo"][i]]+
69       time_effect[df[, "Tiempo"][i]]+ df[,"Carry"][i]*carry_
           effect[df[, "Tiempo"][i]]
70   }
71   df=df[order(df$Sujeto,df$Periodo, df$Tiempo), ]
72   if(distr=="normal"){
73     for(i in 1:nrow(df)){
74       df[,"resp"][i] <- rnorm(1, df[,"mu"][i], sd=0.1)
75     }
76   }else{
77     aaaa=rnorm(1,0, 1)
78     p_eff=c(0,3,3)
79     seq_effect=c(aaaa, aaaa+betakk)
80     ttt=(1:L)
81     time_effect=(cos(ttt)-cos(ttt)[1])
82     carry_effect=(sin(ttt)-sin(ttt)[1])
83     for(i in 1:nrow(df)){
84       df$mu[i]<-seq_effect[df[, "Secuencia"][i]]+p_eff[df[, "
           Periodo"][i]]+
85         time_effect[df[, "Tiempo"][i]]+ df[,"Carry"][i]*carry_
             effect[df[, "Tiempo"][i]]
86     }
87     for(i in 1:nrow(df)){
88       df[,"resp"][i] <- rpois(1, exp(df[,"mu"][i]))
89     }
90   }
91   df["Per_id"]=as.numeric(as.factor(paste(df$Sujeto,df$Periodo)
       ))
92   beta=c(seq_effect[1]+p_eff[1],c(seq_effect-seq_effect[1])[-1]
       ,c(p_eff-p_eff[1])[-1])
93   return(list(df, R_alpha,beta))
94 }
95
96
```

```
97 sim_QIC_normal=function(betakk=0.5, nsim=100, nind=10, P=3, L
     =10, nseq=3, corr="exch", distr="normal"){
98   QICTotal=matrix(0, ncol=4, nrow=nsim)
99   cobertura=list(matrix(0, nrow=nsim, ncol=(P+nseq-2)),
100                  matrix(0, nrow=nsim, ncol=(P+nseq-2)),
101                  matrix(0, nrow=nsim, ncol=(P+nseq-2)))
102  accuracy=list(matrix(0, nrow=nsim, ncol=(P+nseq-2)),
103                  matrix(0, nrow=nsim, ncol=(P+nseq-2)),
104                  matrix(0, nrow=nsim, ncol=(P+nseq-2)))
105  sesgoes=list(matrix(0, nrow=nsim, ncol=(P+nseq-2)),
106                 matrix(0, nrow=nsim, ncol=(P+nseq-2)),
107                 matrix(0, nrow=nsim, ncol=(P+nseq-2)))
108  n_param=P+nseq-2
109  for(ii in 1:nsim){
110    bb=Sim_kron_cross(betakk=betakk, P=P, nseq=nseq,nind=nind,L
       =L,corr=corr, distr=distr)
111    datos=bb[[1]]
112    beta=c(betakk,3,3)
113    m=2 ## Numero de nodos
114    Tiempo_total=datos$Tiempo
115    splines1= bs(Tiempo_total, knots = quantile(Tiempo_total,
       1:m/(m+1)))
116    Tiempo_carryA=datos %>% filter(Carry!=0) %>% dplyr::select(
       Tiempo) %>% data.frame()
117    carrya=bs(Tiempo_carryA[,1], knots = quantile(Tiempo_carryA
       [,1], 1:m/(m+1)))
118    splinescarryA=matrix(0, ncol=ncol(splines1), nrow=nrow(
       splines1))
119    j=1
120    for(i in 1:nrow(datos)){
121      if(datos$Carry[i]!=0){
122        splinescarryA[i,]<- carrya[j,]
123        j=j+1
124      }
125    }
126    Data= data.frame(splines1, splinescarryA, datos)
127    mod1=gee(resp~factor(Secuencia)+factor(Periodo)+X1+X2+X3+X4
       +X5+
```

```
128                 X1.1+X2.1+X3.1+X4.1+X5.1, data=Data, maxiter =
                       10,
129               corstr = "AR-M", id=Per_id)
130     mod2=gee(resp~factor(Secuencia)+factor(Periodo)+Tiempo,
           data=Data, maxiter = 10,
131               corstr = "AR-M", id=Per_id)
132     mod3=gee(resp~factor(Secuencia)+factor(Periodo)+Tiempo + I(
           Tiempo^2), data=Data, maxiter = 10,
133               corstr = "AR-M", id=Per_id)
134     liminf=mod1$coefficients[2:(n_param+1)]-1.96*sqrt(diag(mod1
           $robust.variance)[2:(n_param+1)])
135     limsup=mod1$coefficients[2:(n_param+1)]+1.96*sqrt(diag(mod1
           $robust.variance)[2:(n_param+1)])
136     for(w in 1:length(liminf)){
137       if(beta[w]>=liminf[w] & beta[w]<=limsup[w]){
138         cobertura[[1]][ii,w]<-1
139       }
140     }
141     for(w in 1:length(liminf)){
142       if(0>=liminf[w] & 0<=limsup[w]){
143         accuracy[[1]][ii,w]<-1
144       }
145       sesgoes[[1]][ii,]<-(beta-mod1$coefficients[2:(n_param+1)
              ])^2
146     }
147
148
149     liminf=mod2$coefficients[2:(n_param+1)]-1.96*sqrt(diag(mod2
           $robust.variance)[2:(n_param+1)])
150     limsup=mod2$coefficients[2:(n_param+1)]+1.96*sqrt(diag(mod2
           $robust.variance)[2:(n_param+1)])
151
152     for(w in 1:length(liminf)){
153       if(beta[w]>=liminf[w] & beta[w]<=limsup[w]){
154         cobertura[[2]][ii,w]<-1
155       }
156     }
157     for(w in 1:length(liminf)){
```

```
158       if(0>=liminf[w] & 0<=limsup[w]){
159         accuracy[[2]][ii,w]<-1
160       }
161       sesgoes[[2]][ii,]<-(beta-mod2$coefficients[2:(n_param+1)
            ])^2
162     }
163
164     liminf=mod3$coefficients[2:(n_param+1)]-1.96*sqrt(diag(mod3
          $robust.variance)[2:(n_param+1)])
165     limsup=mod3$coefficients[2:(n_param+1)]+1.96*sqrt(diag(mod3
          $robust.variance)[2:(n_param+1)])
166     for(w in 1:length(liminf)){
167       if(beta[w]>=liminf[w] & beta[w]<=limsup[w]){
168         cobertura[[3]][ii,w]<-1
169       }
170     }
171     for(w in 1:length(liminf)){
172       if(0>=liminf[w] & 0<=limsup[w]){
173         accuracy[[3]][ii,w]<-1
174       }
175       sesgoes[[3]][ii,]<-(beta-mod3$coefficients[2:(n_param+1)
            ])^2
176     }
177
178     QIC=numeric(3)
179     QIC[1]<-sum((mod1$residuals)^2)+
180       2*sum(diag(ginv(as.matrix(mod1$robust.variance))%*%mod1$
            robust.variance))-32
181     QIC[2]<-sum((mod2$residuals)^2)+
182       2*sum(diag(ginv(as.matrix(mod2$robust.variance))%*%mod2$
            robust.variance))
183     QIC[3]<-sum((mod3$residuals)^2)+
184       2*sum(diag(ginv(as.matrix(mod3$robust.variance))%*%mod3$
            robust.variance))
185   QICTotal[ii,]=c(QIC, min(QIC[1:3]))
186   print(ii)
187   }
188 return(list(QICTotal, cobertura, accuracy, sesgoes))
```

```
189 }
190
191 P=3
192 L=10
193 nseq=2
194 betakk=1
195
196 n=50
197 QIC=list()
198 xx=seq(2,n,3)
199 cobertura=list(matrix(0, nrow=length(xx), ncol=(P+nseq-2)),
200                matrix(0, nrow=length(xx), ncol=(P+nseq-2)),
201                matrix(0, nrow=length(xx), ncol=(P+nseq-2)))
202 sesgoes=cobertura
203 accuracy=sesgoes
204 mejores=matrix(0, ncol=3, nrow=length(xx))
205 nsim=200
206 for(i in 1:length(xx)){
207   QIC_total=sim_QIC_normal(betakk=betakk, nsim=nsim, nind=xx[i
      ], corr="exch", P=P, L=10, nseq = nseq)
208   QIC[[i]]=QIC_total[[1]]
209   for(m in 1:3){
210     cobertura[[m]][i,]=colSums(QIC_total[[2]][[m]])/nsim
211     accuracy[[m]][i,]=colSums(QIC_total[[3]][[m]])/nsim
212     sesgoes[[m]][i,]=colSums(QIC_total[[4]][[m]])/nsim
213   }
214   QIC_total=QIC_total[[1]]
215   mejores[i,]=c(sum(QIC_total[,1]-QIC_total[,4]==0)/nrow(QIC_
      total),
216   sum(QIC_total[,2]-QIC_total[,4]==0)/nrow(QIC_total),
217   sum(QIC_total[,3]-QIC_total[,4]==0)/nrow(QIC_total))
218   print(paste(rep(i, 100)) )
219 }
220 beep(8)
221
222 save(cobertura, sesgoes, mejores, file="SIM_NORMAL_beta1.RData"
    )
223
```

```
224
225
226 plot(xx, cobertura[[1]][,1])
227
228 library(xtable)
229
230
231 mod1 <- cbind(xx, cobertura[[1]])
232 xtable(mod1)
233
234 mod2 <- cbind(cobertura[[2]], xx)
235
236
237 mod3 <- cbind(cobertura[[2]], xx)
238
239
240
241
242 mejores=cbind(cobertura[[1]],cobertura[[1]],cobertura[[1]])
243 for(i in 1:3){
244   mejores1[,i]=predict(loess(mejores[,i]~xx), se = TRUE)$fit
245   mejores1[,3+i]=mejores1[,i]+1.96*predict(loess(mejores[,i]~xx
        ), se = TRUE)$se
246   mejores1[,6+i]=mejores1[,i]-1.96*predict(loess(mejores[,i]~xx
        ), se = TRUE)$se
247 }
248
249
250 fig=plot_ly(x=xx, y=mejores[,1], type="scatter", mode="lines",
      name = "Treatment effect")%>%
251   add_trace(y=mejores[,2], type="scatter", mode="lines", name="
        Second period effect")%>%
252   add_trace(y=mejores[,3], type="scatter", mode="lines", name="
        Third period effect") %>%
253   layout(xaxis=list(title="Number of replications by sequence")
        ,
254          yaxis=list(title="Proportions of times that H0 is
               accepted"))
```

```
255 fig
256
257
258
259 ### Efecto de secuencia
260 fig=plot_ly(y=mejores[,1], type="box", name = "Independence",
      marker = list(color = 'rgb(9,56,125)'),
261              line = list(color = 'rgb(9,56,125)'))%>%
262   add_trace(y=mejores[,2], type="box", name="Exchangeable")%>%
263   add_trace(y=mejores[,3], type="box", name="Autoregressive")
        %>%
264   add_trace(y=mejores[,4], type="box", name="Kronecker")%>%
265   layout(xaxis=list(title="Correlation Matrix\n  "),
266          yaxis=list(title="Proportions of lowest QIC"))
267 ax <- list(
268   zeroline = TRUE,
269   showline = TRUE,
270   mirror = "ticks",
271   zerolinecolor = toRGB("red"),
272   zerolinewidth = 1,
273   linecolor = toRGB("black"),
274   linewidth = 0.11,
275   title="Correlation matrix using in GEE\n"
276 )
277 ay <- list(
278   zeroline = TRUE,
279   showline = TRUE,
280   mirror = "ticks",
281   zerolinecolor = toRGB("red"),
282   zerolinewidth = 1,
283   linecolor = toRGB("black"),
284   linewidth = 0.11,
285   range=c(0,1),
286   title="Proportions"
287 )
288
289
290 fig %>% layout(xaxis = ax, yaxis = ay,font=list(size=14),
```

```
    showlegend = FALSE)
```

C.2. Code for analyze arterial pressure data

```
##################
##################
## Cargar librerias

library(splines)
library(mvtnorm)
library(geeM)
library(gee)
library(readxl)
library(tidyverse)
library(ggcorrplot)

library(geepack)
library(MASS)
library(corrplot)
library(plotly)
library(gridExtra)
library(cowplot)

Datos <- read_excel("Doctorado/Datos_con_base.xlsx")
Datos=Datos[order(Datos$Sujeto,Datos$Periodo, Datos$Tiempo), ]

Datos["Per_id"]=as.numeric(as.factor(paste(Datos$Sujeto, Datos$
    Periodo)))
m=5 ## Numero de nodos

formula_beta=Presion~Base+factor(Periodo)+Tratamiento
mod1=gee(Presion~Base+factor(Periodo)+Tratamiento,data=Datos,
    corstr = "AR-M", id=Per_id)
beta0=coef(mod1)
Xbeta=model.matrix(lm(formula_beta, data=Datos))

Tiempo_total=Datos$Tiempo
```

```
32 splines1= bs(Tiempo_total, knots = quantile(Tiempo_total, 1:m/(
     m+1)))
33
34 Tiempo_carryA=Datos %>% filter(CarryA!=0) %>% dplyr::select(
     Tiempo) %>% data.frame()
35
36 carrya=bs(Tiempo_carryA[,1], knots = quantile(Tiempo_carryA
     [,1], 1:m/(m+1)))
37 splinescarryA=matrix(0, ncol=ncol(splines1), nrow=nrow(splines1
     ))
38 j=1
39 for(i in 1:nrow(Datos)){
40   if(Datos$CarryA[i]!=0){
41     splinescarryA[i,]<- carrya[j,]
42     j=j+1
43   }
44 }
45
46 Tiempo_carryB=Datos %>% filter(CarryB!=0) %>% dplyr::select(
     Tiempo) %>% data.frame()
47 carryb=bs(Tiempo_carryB[,1], knots = quantile(Tiempo_carryB
     [,1], 1:m/(m+1)))
48 splinescarryB=matrix(0, ncol=ncol(splines1), nrow=nrow(splines1
     ))
49 j=1
50 for(i in 1:nrow(Datos)){
51   if(Datos$CarryB[i]!=0){
52     splinescarryB[i,]<- carryb[j,]
53     j=j+1
54   }
55 }
56
57
58 formula_beta=Presion~Base+factor(Periodo)+Tratamiento
59 dif=numeric(10)
60 for(kk in 1:10){
61
62 beta0=coef(mod1)
```

```
63 Xbeta=model.matrix(lm(formula_beta, data=Datos))
64 R_alpha=mod1$working.correlation
65 Data= data.frame(xbeta=Xbeta %*%beta0, splines1, splinescarryA,
       splinescarryB, Datos)
66
67 mod1=gee(Presion~X1+X2+X3+X4+X5+X6+X7+X8+
68             X1.1+X2.1+X3.1+X4.1+X5.1+X6.1+X7.1+X8.1+
69             X1.2+X2.2+X3.2+X4.2+X5.2+X6.2+X7.2+X8.2+ offset(
                  xbeta), data=Data, R=R_alpha, maxiter = 10,
70           corstr = "fixed", id=Per_id)
71 alpha_tiempo=coef(mod1)[2:9]
72 Z1=Data[,2:9]
73
74 alpha_carryA=coef(mod1)[10:17]
75 Z2=Data[,10:17]
76
77 alpha_carryB=coef(mod1)[18:25]
78 Z3=Data[,18:25]
79
80 T1=as.matrix(Z1)%*%alpha_tiempo
81 T2=as.matrix(Z2)%*%alpha_carryA
82 T3=as.matrix(Z3)%*%alpha_carryB
83 Data1=data.frame(Datos, T1, T2, T3)
84 mod1=gee(Presion~Base+factor(Periodo)+Tratamiento+offset(T1)+
       offset(T2)+offset(T3),
85           data=Data1,corstr = "AR-M", id=Per_id)
86
87 dif[kk]=sum((beta0-coef(mod1))^2)
88 }
89
90
91
92
93 plot(dif)
94
95
96 n_tiempos=10
97 x=Tiempo_total[1:n_tiempos]
```

```
y=as.matrix(Z1[1:n_tiempos,])%*%t(t(alpha_tiempo))
a=lm(y~bs(x, knots = 40))

x=seq(-30,240,1)
aaa=predict(a, newdata = data.frame(x=x), interval = "prediction")
liminf=(aaa[,1]-aaa[,2])/sqrt(36)
limsup=(-aaa[,1]+aaa[,3])/sqrt(36)
aaa[,2]=aaa[,1]-liminf
aaa[,3]=aaa[,1]+limsup

par(mfrow=c(2,2))

data_con=data.frame(aaa)
colnames(data_con)=c("Estimate_time_effect", "lower_band","upper_band")
data_con$Time =seq(-30,240,1)
plot1 <-data_con %>% ggplot( aes(x = Time, y =Estimate_time_effect)) +
  # plot CI-s as a ribbon arround each line
  geom_ribbon(aes(ymin = lower_band, ymax = upper_band), alpha=0.4) +
  # set manually the fill color for ribbons

  # add the average lines on top of CI ribbons
  geom_line(color="black",lwd=.8, linetype='solid') +
  # NOTE: to set the order in legend as desired one needs to mention that in breaks=()

  # Final adjustments (optional)
  # set axis labels
  xlab("Time in minutes")+ylab("Change in blood systolic pressure")+
  geom_hline(yintercept = 0)+geom_vline(xintercept = 0)



x=Tiempo_total[1:n_tiempos]
```

```
130 y=as.matrix(Z2[51:60,])%*%t(t(alpha_carryA))
131 a=lm(y~bs(x, knots = 50))
132 x=seq(-30,240,1)
133 aaa=predict(a, newdata = data.frame(x=x), interval = "
      prediction")
134 liminf=(aaa[,1]-aaa[,2])/sqrt(11)
135 limsup=(-aaa[,1]+aaa[,3])/sqrt(11)
136 aaa[,2]=aaa[,1]-liminf
137 aaa[,3]=aaa[,1]+limsup
138
139 fig=plot_ly(x=x, y=aaa[,1], type="scatter", mode="lines", name
      = "Estimate")%>%
140   add_trace(y=aaa[,2], type="scatter", mode="lines", name="
        lower band",
141             line=list(dash="dash"))%>%
142   add_trace(y=aaa[,3], type="scatter", mode="lines", name="
        upper band",
143             line=list(dash="dot")) %>%
144   layout(xaxis=list(title="time of observation\n", zeroline=T,
        showline=T),
145          yaxis=list(title="Carry over of A treatment"), title="
              Estimate of carry over effect")
146 fig
147
148 data_con=data.frame(aaa)
149 colnames(data_con)=c("Estimate_time_effect", "lower_band","
      upper_band")
150 data_con$Time =seq(-30,240,1)
151 plot2 <- data_con %>% ggplot( aes(x = Time, y =Estimate_time_
      effect)) +
152   # plot CI-s as a ribbon arround each line
153   geom_ribbon(aes(ymin = lower_band, ymax = upper_band), alpha
        =0.4) +
154   # set manually the fill color for ribbons
155
156   # add the average lines on top of CI ribbons
157   geom_line(color="black",lwd=.8, linetype='solid') +
158   # NOTE: to set the order in legend as desired one needs to
```

```
      mention that in breaks=()
159
160   # Final adjustments (optional)
161   # set axis labels
162   xlab("Time in minutes")+ylab("First order carry-over effect
        of A")+
163   geom_hline(yintercept = 0)+geom_vline(xintercept = 0)
164
165
166
167
168 x=Tiempo_total[1:n_tiempos]
169 y=as.matrix(Z3[21:30,])%*%t(t(alpha_carryB))
170 a=lm(y~bs(x, knots = 50))
171 x=seq(-30,240,1)
172 aaa=predict(a, newdata = data.frame(x=x), interval = "
      prediction")
173 liminf=(aaa[,1]-aaa[,2])/sqrt(14)
174 limsup=(-aaa[,1]+aaa[,3])/sqrt(14)
175 aaa[,2]=aaa[,1]-liminf
176 aaa[,3]=aaa[,1]+limsup
177 fig=plot_ly(x=x, y=aaa[,1], type="scatter", mode="lines", name
      = "Estimate")%>%
178   add_trace(y=aaa[,2], type="scatter", mode="lines", name="
        lower band",
179             line=list(dash="dash"))%>%
180   add_trace(y=aaa[,3], type="scatter", mode="lines", name="
        upper band",
181             line=list(dash="dot")) %>%
182   layout(xaxis=list(title="time of observation\n", zeroline=T,
        showline=T),
183          yaxis=list(title="Carry over of B treatment"), title="
               Estimate of carry over effect")
184 fig
185
186 data_con=data.frame(aaa)
187 colnames(data_con)=c("Estimate_time_effect", "lower_band","
      upper_band")
```

```
188 data_con$Time =seq(-30,240,1)
189 plot3 <- data_con %>% ggplot( aes(x = Time, y =Estimate_time_
      effect)) +
190   # plot CI-s as a ribbon arround each line
191   geom_ribbon(aes(ymin = lower_band, ymax = upper_band), alpha
        =0.4) +
192   # set manually the fill color for ribbons
193
194   # add the average lines on top of CI ribbons
195   geom_line(color="black",lwd=.8, linetype='solid') +
196   # NOTE: to set the order in legend as desired one needs to
        mention that in breaks=()
197
198   # Final adjustments (optional)
199   # set axis labels
200   xlab("Time in minutes")+ylab("First order carry-over effect
        of B")+
201   geom_hline(yintercept = 0)+geom_vline(xintercept = 0)
202
203
204
205
206
207 mod1=geeglm(Presion~Base+factor(Periodo)+Tratamiento+offset(T1)
      +offset(T2)+offset(T3),
208           data=Data1,corstr = "ar1", id=Per_id)
209 a=summary(mod1)
210 xtable(a$coefficients)
211
212
213 library(qqplotr)
214
215 qqnorm(mod1$residuals/sd(mod1$residuals))
216 aaa=data.frame(res=mod1$residuals/sd(mod1$residuals))
217 plot4 <-  ggplot(aaa,mapping = aes(sample = res)) +
218   stat_qq_band(distribution = "norm", conf=0.99) +
219   stat_qq_line() +
220   stat_qq_point() +
```

```
221   labs(x = "Theoretical Quantiles", y = "Sample Quantiles")
222 plot4
223 abline(a=0, b=1)
224
225 grid.arrange(plot1, plot2, plot3, plot4)
226
227
228 ggdraw()+draw_plot(plot1, 0, 0.5,0.5,0.5)+
229   draw_plot(plot2, 0.5, 0.5,0.5,0.5)+
230   draw_plot(plot3, 0, 0,0.5,0.5)+
231   draw_plot(plot4, 0.5,0,0.5,0.5)+
232   draw_plot_label(c("a)","b)", "c)", "d)"), c(0,0.5, 0, 0.5), c
        (1,1, 0.5, 0.5))
```

C.3. Code for analyze rabbits data

```
1 ###################
2 ###################
3 ## Cargar librerias
4 rm(list=ls())
5 library(splines)
6 library(mvtnorm)
7 library(geeM)
8 library(xtable)
9 library(gee)
10 library(readxl)
11 library(tidyverse)
12 library(ggcorrplot)
13 library(geepack)
14 library(MASS)
15 library(corrplot)
16 library(plotly)
17 library(gridExtra)
18 library(cowplot)
19
20
21 Datos <- read_delim("Doctorado/Rabbits.csv", ";", escape_double
```

```
    = FALSE, trim_ws = TRUE)
Datos$Sujeto <-Datos$rabbit
Datos$Tiempo <- Datos$hours
Datos$Periodo <- Datos$Period
Datos$Tratamiento <- Datos$Treatment
Datos=Datos[order(Datos$Sujeto,Datos$Periodo, Datos$Tiempo), ]
Datos$Secuencia <- factor(Datos$sequence)
levels(Datos$Secuencia) <- c("ABBA", "BABA")
Datos$Per_seq <-paste("Period ",Datos$Period, " of ", Datos$
    Secuencia, sep="")

Datos %>%  ggplot(aes(x=hours, y=Blood_Sugar,group=interaction(
    rabbit,Period, Treatment),
                      color=Treatment))+
  geom_line()+facet_wrap(.~Per_seq, nrow = 4)+
  xlab("Time of observation")+ ylab("Blood Suggar levels")


Datos["Per_id"]=as.numeric(as.factor(paste(Datos$Sujeto, Datos$
    Periodo)))
m=1 ## Numero de nodos
Datos$Blood_Sugar[Datos$Blood_Sugar<1]<-1

formula_beta=Blood_Sugar~factor(Periodo)*Tratamiento
mod1=gee(formula_beta,data=Datos,corstr = "AR-M", id=Per_id,
    family = Gamma(link="log"))
beta0=coef(mod1)
Xbeta=model.matrix(lm(formula_beta, data=Datos))

Tiempo_total=Datos$Tiempo
splines1= bs(Tiempo_total, knots = quantile(Tiempo_total, 1:m/(
    m+1)))

Tiempo_carryA=Datos %>% filter(CarryA==1) %>% dplyr::select(
    Tiempo) %>% data.frame()

carrya=bs(Tiempo_carryA[,1], knots = quantile(Tiempo_carryA
    [,1], 1:m/(m+1)))
```

```
52 splinescarryA=matrix(0, ncol=ncol(splines1), nrow=nrow(splines1
     ))
53 j=1
54 for(i in 1:nrow(Datos)){
55   if(Datos$CarryA[i]==1){
56     splinescarryA[i,]<- carrya[j,]
57     j=j+1
58   }
59 }
60
61
62
63 dif=numeric(10)
64 for(kk in 1:10){
65
66 beta0=coef(mod1)
67 Xbeta=model.matrix(lm(formula_beta, data=Datos))
68 R_alpha=mod1$working.correlation
69 Data= data.frame(xbeta=Xbeta %*%beta0, splines1, splinescarryA,
       Datos)
70
71 mod1=gee(Blood_Sugar~X1+X2+X3+X4+
72            X1.1+X2.1+X3.1+X4.1+ offset(xbeta), data=Data, R=R_
                alpha, maxiter = 10,
73          corstr = "fixed", id=Per_id, family = Gamma(link="log"
              ))
74 alpha_tiempo=coef(mod1)[2:5]
75 Z1=Data[,2:5]
76
77 alpha_carryA=coef(mod1)[6:9]
78 Z2=Data[,6:9]
79
80 T1=as.matrix(Z1)%*%alpha_tiempo
81 T2=as.matrix(Z2)%*%alpha_carryA
82 Data1=data.frame(Datos, T1, T2)
83 mod1=gee(Blood_Sugar ~factor(Periodo)*Tratamiento+offset(T1)+
     offset(T2),
84          data=Data1,corstr = "AR-M", id=Per_id, family = Gamma(
```

```
            link="log"))
85
86 dif[kk]=sum((beta0-coef(mod1))^2)
87 }
88
89
90
91
92 plot(dif)
93
94
95 n_tiempos=5
96 x=Tiempo_total[1:n_tiempos]
97 y=as.matrix(Z1[1:n_tiempos,])%*%t(t(alpha_tiempo))
98 a=lm(y~bs(x, knots = 40))
99
100 x=seq(0,6,0.1)
101 aaa=predict(a, newdata = data.frame(x=x), interval = "
      prediction")
102 liminf=(aaa[,1]-aaa[,2])/sqrt(44)
103 limsup=(-aaa[,1]+aaa[,3])/sqrt(44)
104 aaa[,2]=aaa[,1]-liminf
105 aaa[,3]=aaa[,1]+limsup
106
107 plot(x, aaa[,1] ,type="l", ylim=c(min(aaa[,2]), max(aaa[,3])),
108      xlab="Tiempo desde la aplicacion", ylab="Cambio en la
           presion sistolica")
109
110 lines(x, aaa[,2])
111 lines(x, aaa[,3])
112 abline(h=0, col=2)
113
114
115 data_con=data.frame(aaa)
116 colnames(data_con)=c("Estimate_time_effect", "lower_band","
      upper_band")
117 data_con$Time =seq(0,6,0.1)
118 plot1 <- data_con %>% ggplot( aes(x = Time, y =Estimate_time_
```

```
    effect)) +
  # plot CI-s as a ribbon arround each line
  geom_ribbon(aes(ymin = lower_band, ymax = upper_band), alpha
     =0.4) +
  # set manually the fill color for ribbons

  # add the average lines on top of CI ribbons
  geom_line(color="black",lwd=.8, linetype='solid') +
  # NOTE: to set the order in legend as desired one needs to
     mention that in breaks=()

  # Final adjustments (optional)
  # set axis labels
  xlab("Time in hours")+ylab("Change of blood sugar levels")+
  geom_hline(yintercept = 0)+geom_vline(xintercept = 0)



x=Tiempo_total[1:n_tiempos]
y=as.matrix(Z2[6:10,])%*%t(t(alpha_carryA))
a=lm(y~bs(x, knots = 40))
x=seq(0,6,0.1)
aaa=predict(a, newdata = data.frame(x=x), interval = "
    prediction")
liminf=(aaa[,1]-aaa[,2])/sqrt(11)
limsup=(-aaa[,1]+aaa[,3])/sqrt(11)
aaa[,2]=aaa[,1]-liminf
aaa[,3]=aaa[,1]+limsup

fig=plot_ly(x=x, y=aaa[,1], type="scatter", mode="lines", name
    = "Estimate")%>%
  add_trace(y=aaa[,2], type="scatter", mode="lines", name="
     lower band",
            line=list(dash="dash"))%>%
  add_trace(y=aaa[,3], type="scatter", mode="lines", name="
     upper band",
            line=list(dash="dot")) %>%
  layout(xaxis=list(title="time of observation\n", zeroline=T,
```

```
      showline=T),
150          yaxis=list(title="Carry over of A treatment"), title="
                 Estimate of carry over effect")
151 fig
152
153 data_con=data.frame(aaa)
154 colnames(data_con)=c("Estimate_time_effect", "lower_band","
    upper_band")
155 data_con$Time =seq(0,6,0.1)
156 plot2 <- data_con %>% ggplot( aes(x = Time, y =Estimate_time_
    effect)) +
157   # plot CI-s as a ribbon arround each line
158   geom_ribbon(aes(ymin = lower_band, ymax = upper_band), alpha
      =0.4) +
159   # set manually the fill color for ribbons
160
161   # add the average lines on top of CI ribbons
162   geom_line(color="black",lwd=.8, linetype='solid') +
163   # NOTE: to set the order in legend as desired one needs to
      mention that in breaks=()
164
165   # Final adjustments (optional)
166   # set axis labels
167   xlab("Time in hours")+ylab(" First order carry-over effect of
       A")+
168   geom_hline(yintercept = 0)+geom_vline(xintercept = 0)
169
170
171
172
173 mod1=geeglm(Blood_Sugar~factor(Periodo)+Tratamiento+offset(T1)+
    offset(T2),
174             data=Data1,corstr = "ar1", id=Per_id, family=Gamma(
                link="log"))
175
176 QIC(mod1)
177
178 mod1=geeglm(Blood_Sugar~factor(Periodo)+Tratamiento+offset(T1)+
```

```
    offset(T2),
179             data=Data1,corstr = "exchangeable", id=Per_id,
                   family=Gamma(link="log"))
180
181 QIC(mod1)
182 a=summary(mod1)
183 xtable(a$coefficients)
184 a
185
186 QIC(mod1)
187 print(paste("QIC", QIC(mod1)))
188
189
190
191 library(qqplotr)
192
193 qqnorm(mod1$residuals/sd(mod1$residuals))
194 aaa=data.frame(res=mod1$residuals/sd(mod1$residuals))
195 plot3 <- ggplot(aaa,mapping = aes(sample = res)) +
196   stat_qq_band(distribution = "norm", conf=0.9999
197                ) +
198   stat_qq_line() +
199   stat_qq_point() +
200   labs(x = "Theoretical Quantiles", y = "Sample Quantiles")
201 plot3
202
203 abline(a=0, b=1)
204
205
206 ggdraw()+draw_plot(plot1, 0, 0.5,1,0.5)+
207   draw_plot(plot2, 0, 0,0.5,0.5)+
208   draw_plot(plot3, 0.5, 0,0.5,0.5)+
209   draw_plot_label(c("a)","b)", "c)"), c(0, 0, 0.5), c(1,0.5,
        0.5))
210 options(digits=10)
211 mean((mod1$y-predict(mod1, type="response"))^2)
212
213 sqrt(sum(diag(vcov(mod1))))
```

```
214
215
216 mod4=glm(Blood_Sugar~factor(Periodo)+Tratamiento+offset(T1)+
      offset(T2),
217             data=Data1, family=Gamma(link="log"))
218
219 extractAIC(mod4)
```