UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXIV

New permutation methodologies to deal with the multiplicity issue: multiple comparisons and multiple tests with applications to single-case experiments and to regression analysis

Direttore della Scuola: Ch.ma Prof. Alessandra Salvan

Supervisore: Ch.mo Prof. Fortunato Pesarin

Co-supervisori: Ch.mo Prof. Luigi Salmaso e Ch.mo Prof. Patrick Onghena

Dottoranda: Francesca Solmi

31 Gennaio 2012

Riassunto

Questa tesi presenta nuove metodologie di permutazione per risolvere problemi reali di natura complessa. Spesso i dati reali sono risultati di complesse fasi di pianificazione dell'esperimento, o sono di loro natura complessi. Risposte multiple sono spesso di interesse e, fatto che aumenta ulteriormente la complessità, le strutture di dipendenza presenti all'interno dei dati sono, oltre che complicate, sconosciute (*problema della molteplicità*). Sono qui stati affrontati due problemi reali: i così detti single-case experiments e l'analisi di dati ordinali.

Nella tesi vengono proposte soluzioni sia univariate che multivariate, che mostrano di risolvere il problema in modo soddisfacente tramite l'utilizzo di test di permutazione e della loro combinazione non parametrica. Riguardo i single-case experiments viene presentata una soluzione complessa basata sulla combinazione di tecniche di lisciamento e della teoria di permutazione. Per l'analisi di dati ordinali, invece, si propongono alcuni test di permutazione che utilizzano stime non parametriche come statistiche test, creando in questo modo un collegamento tra soluzione del problema via parametrica e non parametrica.

Diversi studi di simulazione e applicazioni a dati reali mostrano il buon comportamento e l'utilità dei metodi proposti.

Abstract

The thesis presents new results within the *permutation* testing approach in order to deal with *real complex* problems. Very often real datasets are the result of complicated planning phases of the study or they are *complex* by themselves. Multiple outcomes are often of interest and, a fact which increases further on their complexity, complicated and unknown dependence structures can underlie such multivariate responses (i.e. the *multiplicity issue*). Two particular applied problems are faced: single-case experiments and regression analysis of ordinal data.

Both univariate and multivariate solutions to such issues are proposed in this thesis, which show to successfully handle the data complexity by means of permutation tests and their nonparametric combination. Regarding the single-case experiments problem a complex solution is developed which exploits the joint use of smoothing techniques and permutation theory. For ordinal data analysis instead, we propose some permutation solutions that use parametric estimates as test statistics, creating a link between parametric and nonparametric problem solving.

Several simulation studies and real case applications show the good behavior and the usefulness of the presented procedures.

Acknowlegments

I would like to thank Prof. Fortunato Pesarin for having introduced me to the fashinating topic of permutation testing and for being my guide over the last 6 years. I would like to thank Prof. Luigi Salmaso for having always encouraged me in what I was doing, before and during my Ph.D.. I would also like to thank Prof. Onghena for the time and the energy he spent with me during my visiting period in Belgium and also after that, for his always helpful comments and suggestions, and last but not least, for his kindness, which made me feel at home since the first day I was in Leuven.

A special thank goes to my family, who has always believed in me and supported me in all my choices. In particular I thank my mum, for having spent her entire life for and with Marco and me. All the things I know and I believe in come from her. And I thank Marco for always understanding how I feel and what I need, without me needing to say anything. He is the best brother a sister could have. I thank my dad for his support, together with Tania, in taking my important decisions; even if most of the times I still do what I want, it's very important for me to have those moments of discussion with him. I thank Pupo, for being a super grandpa, and for having said to be happy now of becoming even an "international grandpa", he is the best! I thank my aunt Betty for being always very close to me and for the irreplaceable talks together. I thank my aunt Carla for her always enthusiastic encouragements in trying to make the best of everything. I thank my uncle Fabrizio for the many great skiing and trekking days together, since I was a child. I thank my cousins Cri and Stefy for being the best part of the family meetings. I thank my cousin Chiara too, for the very special link I have with her.

I thank my "old" (or better "longer lasting") friends for all the things they have done for and with me, not only in the last three years: for having believed in me, having fought, laughed, cried, dreamed, planned, had fun with me. So, thank you Chiara, Fedy, Ba, Silvia, Gian, Peter, Lety and all the others. I want to thank also my "new" (better "more recent") friends, the beautiful people I met in Belgium, for having made my year there an amazing experience. I was used to think that, when you grow up, it becomes more difficult to find new real friends... I was obviously mistaking. So, thank you Ila, Cla, Laura, Palo, Velina, Mariela, Tamzin, Christine, Ilse, Manu, Magnus and all the others.

I also want to thank my colleagues, the XXIV cycle Ph.D. students, I guess that to say that "my cycle is different" is enough.
I thank a lot Dario too, for the precious help he gave me with this thesis, and as important as that, for not having killed me after one year of house sharing!

I thank MY little Poldo for being MY little Poldo.

And finally I want to thank Bas, for being so great that he managed to make me change my life to be with him. I hope that what we have just started will keep on being as incredibly nice as it is now.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter provides an introduction to the research work presented in this thesis. It gives an overview of the the topics we dealt with during the second and third year of my Ph.D., discussing the research background and describing the motivations that brought us to explore this particular topics. A description of the structure of the thesis is also given. Moreover the main contributions are presented, together with a more detailed description of the faced problems and of the proposed solutions.

1.1 Overview

The thesis presents new results within the *permutation* testing approach in order to deal with *real complex* problems. Very often real datasets are the result of complicated planning phases of the study or they are *complex* by themselves: multiple outcomes are often of interest and, a fact which increases further on their complexity, complicated and unknown dependence structures can underlie such multivariate responses (i.e. the *multiplicity issue*). As it can be a difficult job to try to solve *real complex* problems, and as we are aware of the fact that statistics can play an important role in many business or health-care decisions, we think that the data analysis should be performed in a suitable way for the problem at hand. Nonparametric statistics can be classified, in this sense, among the most powerful statistical approaches, since it avoids too many assumptions on the data and performs a distribution-free inference. Among nonparametric testing methods, *permutation* techniques are known to be well performing while not requiring too strong assumptions on the data, and this is the reason why we decided to adopt such methodology in order to solve the problems we present in this work. That is also why we chose this title for this thesis: we will introduce new *permutation* methodologies to deal with the *multiplicity issue* in *real* applications.

The thesis is divided in two parts, related to two main projects which have been developed during this research period: a first one is related to the study of single-case experiments (see Onghena (2005) and Diggle et al. (2002)), where longitudinal data are recorded on one subject regarding some response of interest and multiple treatment conditions are alternated on time and compared among each other. The second problem we faced is the analysis of ordinal data,

which is nowadays a very "hot" topic: very often statistical surveys, having ordinal outcomes, are conducted in many fields (e.g. in marketing research) and ordinal responses are often recorded also in clinical research while, for instance, comparing different treatments or drugs. The awareness of the complexity of real data and the presence, very often, of multivariate responses is the reason why for the two real problems we present both a basic univariate solution and an extension to the multivariate case. Within the single-case experiments, the situation of replication of such kind of study on more than one subject, also allowing the presence of a multivariate response, will be considered. On the other hand the analysis of survey data with more than one outcome is also solved through the use of a multivariate technique.

Chapters 2 and 3 are devoted to the description of the solution proposed for single-case experiments while the Chapters 4 and 5 refer to the CUB models problem (such acronym is originated by the introduction of Covariates in the mixture of Uniform and shifted Binomial distributions). As already said, univariate and multivariate solutions are provided for both the issues, hence separated chapters are dedicated to the two solutions. Chapter 2 describes the univariate solution to test for treatments' effect difference in single-case experiments, giving the motivation and the description of the method and exploring its performances by means of a simulation study and a real application. The extension to the general case of replication of single-case experiments with multivariate response is considered in Chapter 3, where the permutation strategy is presented and its behavior again checked through a simulation study. In this chapter two applications are also presented. Chapter 4 introduces the CUB models approach and proposes the permutation test as an alternative to likelihood based inference; a simulation study shows its reliability and power and a meaningful application to real data is also presented. Chapter 5 describes the permutation solution to test for covariates influence in the case of multivariate ordinal responses, it discusses the results of a simulation study which shows the good performances of the method and again two real applications are presented. The conclusions of the thesis are presented in Chapter 6. Finally an Appendix is given with the presentation and description of the `R` codes which can be used in order to perform all the permutation solutions presented in this thesis: the first section of the Appendix refers to the univariate and replicated-multivariate solution for single-case experiments, while the second section to the `R` codes for CUB models. The codes are described and two toy examples are also introduced, which can be used in oder to practice with the `R` scripts and the use and interpretation of the outputs.

1.2 Main Contributions of the Thesis

Single-case experiments are clinical experiments which are conducted on single subjects in order to test if two or more treatments (or active treatments and placebo) have different effects on some outcome variable of interest for the particular patient under study. In such experiments different treatments are administered to the subject during a certain period of time; the order of the treatments in time is randomized and usually the experiment is double-blind, in the sense that neither the physician nor the patient know the actual sequence of treatments. Patient-specific results can be extended to a population level

through repetition of single-case experiments: hence the same randomization schedule is applied to several subjects while the treatments are compared in terms of the same outcome variables. The randomization of treatments during the time period of observation allows to find a natural solution to such problem by means of permutation tests: exploiting the randomization scheme we can build the randomization (or permutation) distribution of the test statistic of interest and then use it to make inference without relying on too strong assumptions on the data. The use of permutation tests for inference on single-case experiments can be found already in Bulté and Onghena (2008) and Bulté and Onghena (2009): there the single subject's observations are differently labelled according to treatments and, after the sample means are computed on the several sub-sequences, a permutation test is performed to compare such means. In Bulté and Onghena (2009) the authors suggest the possibility of computing other test statistics on the sub-series (instead of the sample mean), so that differences in trend (and not only in average level) can be detected. Our proposal aims to improve the one just described, still keeping the idea of considering separately the differently treatments' labelled observations; we work within the framework of alternation design single-case experiments, where the randomization schedule allows all the treatments to be present at any time occasion. In our solution these sub-time-series are smoothed and the resulting curves, instead of other summary statistics, are then compared. Moreover in this work we propose an extension of such method to the general case of replicated single-cases with multivariate response, being this kind of studies often used in order to extend the single patient related results to a population of interest (through replication), and at the same time measure the effects of the several treatments on more aspects which can be relevant for the health of the patients (multivariate response). The method is, according to current literature, the first ever proposed to handle such a complex problem with the aim of controlling the multiplicity issue. The global permutation technique is obtained by nonparametric combination of dependent permutation tests and it is the result of a two steps procedure.

The second part of this thesis is devoted to the analysis of ordinal data through permutation methodologies within the CUB models framework (see Piccolo (2003), D'Elia and Piccolo (2005) and Piccolo (2006)). CUB models are regression models which can be used to analyze the influence of some covariates on an ordinal response. Rating problems arise in survey data when the respondents are asked to rate a product or more generally an item; it can also be the case of clinical studies when the patients are asked to rate the level of efficacy of a certain medical treatment. Here the choice of the respondent is modeled as a mixture of two components, the feeling of the subject for the item and the uncertainty due to the fact that a choice among several values has to be done. Inference using this kind of model has been developed in a parametric framework for which asymptotic results based on maximum likelihood theory are available and the method is known to be quite flexible and to give good and reliable performances when the sample size is not small. After an *iid* sample is drawn from the population of interest, the parameters associating the covariates to the response can be estimated via maximum likelihood using an EM algorithm (see Piccolo (2006) and Iannario and Piccolo (2009)). The linear predictor given by covariates values and associated parameters is linked directly to the feeling and uncertainty parameters, which shape the distribution of the response variable.

In order to test the influence of a covariate on the response Wald type or likelihood ratio based tests can be performed. However when the sample size is not high such asymptotic results lack of reliability and alternative solutions must be adopted. We propose some permutation tests based on the maximum likelihood estimates of the parameters; considering the Wald or the likelihood ratio type statistic as simple statistics of the data, then we can then use their permutation distribution in order to make inference. The proposed solution is based on the constrained permutation of raw data and it is valid when the covariates which are not tested, but are included in the null model, are categorical. Such alternative to parametric inference is reliable and powerful also when the sample size is low; in such case the type I error is not controlled by the standard parametric solution. The performance of the permutation tests tends to coincide with the parametric one as we increase the sample size. Such permutation solution is proposed and some of its properties are shown in Bonnini et al. (2011). However the presence of multivariate responses is very common for this kind of data: think about survey data, where there is never a single item which is involved in the study; also in clinical studies often the outcomes of interest can be related to several health level indicators. The necessity of analyzing more than one outcome at the same time is very common, but no parametric method seems to be available in this case. We propose to nonparametrically solve the problem by performing separated permutation tests on the several responses of interest and then building a global test nonparametrically combining the partial tests: in this way the underlying dependence structure can be handled without the need of estimating it.

Hence, this thesis presents new permutation methods to deal with the real scenarios described above. All the proposals have been developed to improve the already existing ones, in order to obtain better performances in a general context. We can summarize the main contributions of the thesis as follows:

- univariate framework: we improve an already existing permutation solution for single-case experiments, which is able to detect general differences in the effects of more treatments;
- univariate framework: we present a valid alternative to classical parametric inference for CUB models, in order to deal with cases in which the asymptotic approximation is not reliable;
- multivariate framework: we assert external validity in replicated single-case experiments, together with the possibility of getting also single patients' level results. Such solution also takes care of the multiplicity issue.
- multivariate framework: we test for covariates effects on multivariate ordinal responses within the CUB models framework. Also in this case we take care of the multiplicity issue.

Chapter 2

Single-case experiments

2.1 Introduction

This chapter is devoted to testing some treatment effects on a response variable, which can be carried out by performing a clinical experiment in which the effect of several treatments on the outcome variable is registered. Such research can be broadly divided into relatively large-scale clinical trials and single-case studies. Here we concentrate on single-case experiments since often large-n trials cannot be carried out, due to financial and human resourcing constraints. For this reason single-case experiments are becoming more and more important.

As described in Onghena (2005), single-case experiments are characterized by the fact that one entity is observed repeatedly during a certain period of time, under different levels ("treatments") of at least one independent variable. Hence we observe only one entity (single-case, $n = 1$), manipulating independent variable(s) (experiment); this way the entity is exposed to all levels of the independent variable, and we handle repeated measures or observations of the dependent variable of interest (longitudinal data, see Diggle et al. (2002)). As argued in Onghena (2005), single-case experiments have a long history in behavioral science; often they are the only viable options if rare or unique conditions are involved, and they can be motivated thinking that actually they mimic the care for the individual patient that is needed in clinical work.

The goal of single-case experiments is usually to investigate the presence of a difference in the effects of the treatments considered in the study. Instead of using standard parametric techniques, which in this setting cannot always provide valid inference, nonparametric tools can be a valid alternative to analyze this kind of data (see Todman and Dugard (2001) and Onghena (2005)). In particular, randomization of assignment of the measurement occasions to treatments allows us to find an interesting solution to the problem in permutation tests. We can thus ensure the study's internal and statistical-conclusion validity. The basic idea is to perform a randomization of the repeated observations in time, instead of in individuals as is usually done, and then choose a statistic to evaluate the difference between treatments.

In this work we aim to motivate the use of permutation techniques in order to analyze single-case data; for this reason new proposals are presented to successfully solve the problem in a permutation framework. Such proposals are

shown to improve the performances of other already existing solutions. In the second Section we give a main argument for the use of randomization methods, describing how randomization is dealt with in these kinds of experiments, and hence how these methods are exact by construction, due to the random allocation in the real experiment. In Section 3 we describe the general approach to the solution of the testing for treatment effects in single-case designs, underlining the need to adopt time-series analysis tools in order to model data which actually display some autocorrelation. Section 4 is devoted to the more in depth description of a new permutation time-dependent solution, describing the general idea behind the permutation methodology for single-case experiments, and discussing proposals for the choice of "best" time-series analysis tools. Section 5 shows the results of a simulation study in which the proposed permutation solutions are tested under a wide variety of data scenarios, and these methods are applied to real clinical psychologist data in Section 6. Conclusions about the obtained results are discussed in Section 7.

2.2 Randomization in single-case designs

The starting point for correctly analyzing single-case designs is to underline the difference between random sampling and random allocation. A crucial point is understanding that random allocation does not apply exclusively to the allocation of participants to treatment conditions. In single-case designs, it also applies to the allocation of opportunities for exposure to treatment conditions. Both these points are essential to understand that the permutation approach really does make sense in this kind of problem.

Random sampling from a large, well-defined population is the usual formal requirement for applying parametric techniques, and it justifies the generalization of results, allowing their external validity. However, it is often impractical to have a genuine random sample from a given population of interest, both because it is difficult to define the population of interest itself, and because costs and time requirements can be prohibitive. Instead, randomization in human experimentation is much more likely than random sampling; each exposure opportunity is considered to be randomly placed into separate urns, each of which represents a particular combination of experimental conditions (independent variables). Hence no general wider population is supposed to exist here, and we are able to control for confounding variables.

The simple and natural randomization mechanism considered in the study can then be used to construct a test - the permutation test - simply by mimicking this underlying random assignment. Hence the permutation test is by construction the most natural candidate for use in data analysis following a given randomization design. Indeed we can easily construct the distribution of the test statistic under the null hypothesis of interest by randomly reassigning the condition urns to exposure opportunities.

2.2.1 Different randomization designs

The randomization designs in single-case experiments can broadly be divided into two categories - alternation and phase designs (see Todman and Dugard (2001) and Onghena (2005)). The former class is characterized by the fact

that any level of the independent variable can be present at each measurement occasion; hence we randomly determine the treatment sequence by taking into account the number of levels of the independent variables and the number of measurement occasions for each level. For example, for one independent variable (treatment) with two levels (say A and B) and three measurement occasions for each level, we simply have to randomly select a sequence from a number of possible ones: $AAABBB$, $AABABB$, $AABBAB$, $AABBBA$, $ABAABB$, $ABABAB$, $ABABBA$, $BAABBA$, $BAABAB$, $BAAABB$, etc. Of course, depending on the specific characteristics of the study, we can impose a minimum, or a maximum, number of consecutive measurement occasions for each level of the independent variable.

Phase designs are useful when fast and frequent alternation of treatments is difficult to perform. In this case several consecutive measurements are performed for each phase, the order of the phases is fixed and we randomize the moments of phase change. Hence we randomly determine the treatments' phases sequence by taking into account the order of the phases, the number of treatments, the minimum length of the phases and the total length of the experiment. For example, for one independent variable (treatment) with two levels (say A and B), six measurement occasions and at least one measurement for each phase, we simply have to randomly select a sequence from the following possible ones: $ABBBBB$, $AABBBB$, $AAABBB$, $AAAABB$, $AAAAAB$.

2.3 Testing for treatment effects in single-case experiments

To make inference in single-case designs is not as straightforward as in large-n problems. Indeed, classic parametric techniques are often not suitable for making statistical decisions while correctly controlling the inferential errors. In this thesis we propose a solution to test for treatment effects in single-case designs; let us use $\{X_{A^j}(t)\}_{t\in T}$, where T is a continuous support, to indicate the time-process of the response under treatment A^j, with $j = 1, \ldots, C$. The null and alternative hypotheses that we are going to consider in this work can then be formalized as follows

$$\begin{cases} H_0: & X_{A^j}(t) \overset{\mathcal{M}}{=} X_{A^l}(t) \qquad \forall j < l, \quad j,l = 1,\ldots,C, \quad \forall t \in T \\ H_1: & \exists\, j,l \in \{1,\ldots,C\}, \quad j \neq l \quad | \quad X_{A^j}(t) \overset{\mathcal{M}}{\neq} X_{A^l}(t), \end{cases},$$

where the notations (t) and $\mathcal{M}$ emphasize that we are considering the model underlying the response time process. Hence we aim to test the null hypothesis that the several treatments have the same effect on the time process underlying the response against the alternative hypothesis that there is at least one treatment effect which is different from the others.

Various approaches have been proposed (see Gorman and Allison (1996) for an in-depth description). Among them, time series analysis provides a valid set of procedures, such as autoregressive integrated moving average (ARIMA) models. This solution is often useful and works well, provided that we handle a sufficient number of observations.

An alternative solution is given by classic ANOVA and least squares regression approaches. Indeed, even if in principle parametric statistics require

normality and homoscedasticity assumptions, in practice they are robust to violations of these assumptions. However their robustness is less strong in the case of small and unequal sample sizes, as actually often arises in single-case experiments. Additionally, the use of such parametric techniques can be criticized for the presence of another strong assumption - that the errors are uncorrelated. This latter problem has to be taken into account, since, as argued in Gorman and Allison (1996) and Levin et al. (1978), autocorrelation very often arises in single-case designs, and classic ANOVA and least squares regression approaches are not robust to violations of this assumption.

A third class of solutions is given by nonparametric tests. In general they represent a valid alternative to parametric solutions, especially in small-n and in large-n cases when the required assumptions may be violated. They can be successfully applied in single-case designs; the positive results of the use of nonparametric rank tests is shown in Edgington (1996).

To summarize, as argued in Todman and Dugard (2001), for single-case designs with a large number of observations (say at least 50 per phase) the time-series analysis approach can be successfully applied. When the number of observations per phase is not that large, we should instead use a permutation test to solve the problem.

2.4 The permutation solution

The permutation-based solution for single-case experiments has been discussed by many authors. A good review is given in Todman and Dugard (2001), where the authors also provide an in-depth discussion of the motivations for using randomization tests in this framework. Recently, this methodology has been discussed in Bulté and Onghena (2008) and Bulté and Onghena (2009), where the description of a specific R package is also given. It is also applied to real data problems in, among other works, de Jong et al. (2005) and ter Kuile et al. (2009). The basic idea is to randomize the repeated observations in time, instead of in individuals as is usually done, and then choose a statistic to evaluate the difference between the treatments. More specifically, the kind of randomization needed to perform the test should be based on the design aspect that is random in the experiment; hence we can randomize either the treatments at each measurement occasion, in alternation designs, or the moments of phase change, in phase designs (see Onghena (2005)).

In this chapter we propose a permutation solution to test for treatment effects in alternation single-case experiments. The aim of our research is to improve the existing proposals on the choice of test statistic to be used to compare the performance of the different treatments. For example, in Bulté and Onghena (2008) and Bulté and Onghena (2009) the authors simply use the sample mean difference, ignoring the time ordering of the observations.

In this section of the chapter we firstly point out the importance of taking care of the autocorrelation between observations, and we underline how permutation tests are able to handle this problem. Then the randomization scheme and the permutation procedure itself are well defined. Furthermore we discuss the problem of choice of test statistic, proposing the use of time-series analysis tools.

2.4.1 Autocorrelation and randomization scheme

As pointed out previously, autocorrelation, i.e. correlation between the residuals of scores in time-series data, is of course present in single-case data. Parametric techniques such as t-test and ANOVA require the assumption of independence of observations, and it is known that violation of this assumption leads to higher type I errors. Therefore it is very important to take care of the serial dependence between the observations.

One main feature of randomization tests for single-case designs is that they are not sensitive to the effect of serial dependency because under the null hypothesis there would be identical responses across occasions. As such the randomization scheme allows us to mimic the true distribution of the test statistic under the null hypothesis. However, this does not mean that any test statistic is good for making inference about the phenomenon of interest, and, if serial dependence exists among the observed data, we should choose a suitable test statistic that can capture the main feature of interest. This means that, excluding the simple case of no trend and effect level of interest, a time-series analysis tool is needed. This aspect will be treated in more detail in the next paragraph.

2.4.2 Time-series solution for the test statistic

Here we discuss the problem of choosing a suitable test statistic for testing the treatment effect in single-case designs. The reader is referred to Bulté and Onghena (2008) where the authors propose the sample mean difference as a test statistic to compare the effect of the two treatments; this choice is motivated by the argument that the focus is on an expected difference in level, which can be reflected by a difference between means. The same proposal can also be found in Bulté and Onghena (2009), where it is suggested that if needed, also other test statistics, such as differences in slopes or intercepts, can be adopted. As far as we are aware, the problem of choosing the type of statistic has not yet been considered, and the *R* package presented in Bulté and Onghena (2008) and Bulté and Onghena (2009) does not offer much choice. We think this is a particularly relevant topic, since the sample mean can be sometimes not suitable to distinguish between time-series: in this kind of problems the shape of the observed points in relation to time is of primary interest, as, for instance, different trends can be found in different time-series while they can register the same average behavior.

The basic idea behind the development of our proposals is that a longitudinal data analysis tool is needed in this case. In order to provide a completely nonparametric solution, nonparametric tools may seem more appropriate. It is true, on the other hand, that even using a parametric tool on each randomized sample, then using its permutation distribution, might be a solution given that we do in fact treat the resulting parametric statistic as a simple choice for the test statistic, completely disregarding its inferential properties.

More precisely, our idea is to estimate the time functional shape of observations (with a nonparametric smoother, as splines, kernel smoother, local regression, as well as a parametric time-series analysis tool) and then use a summary statistic of the resulting curve (say $\psi(\cdot)$). By randomizing the observations according to the randomization design, we are able to reproduce the null distribution of the test statistic, and hence evaluate the p-value for the test

of interest. Thus, we estimate the time functional shape of observations for the sub-datasets of various treatments, and then use a summary statistic which is able to measure the difference between these smoothed time-processes. Indeed, under the null hypothesis the outcome variable should follow similarly shaped time processes for the different treatments; hence by randomizing the labels of the treatment in order to construct the permutation distribution of the test statistic under the null hypothesis, we expect to end up with a set of low values for the difference between the various smoothed time-series. Any extreme value of the observed test statistic will then denote a departure of the data from the null hypothesis.

Formally, we can describe the general randomization procedure as follows:

i. consider the original time-series $\boldsymbol{x} = (x_1, \ldots, x_T)$, and compute the observed value for the test statistic $t^{obs} = \psi(x)$;

ii. according to the randomization scheme, perform a randomization of the treatments' labels for $\boldsymbol{x}$, obtaining the permuted time-series $\boldsymbol{x}^{*,1}$, and compute the value for the test statistic $t^{*,1} = \psi(\boldsymbol{x}^{*,1})$;

iii. repeat Step ii. B times, getting the values of the test statistic $t^{*,i} = \psi(\boldsymbol{x}^{*,i})$, for $i = 1, \ldots, B$;

iv. construct the permutation distribution of the test statistic under the null hypothesis from the vector of values $t^* = (t^{*,1}, \ldots, t^{*,B})$;

v. compute the p-value of the test for treatment effects, for example rejecting the null hypothesis for high values of the test statistic, as $p = \sum_{i=1}^{B} \mathbf{1}_{(t^{obs},\infty)}(t^{*,i}(\boldsymbol{x}^{*,i}))/B$.

We remark that B can be either the number of all the possible permutations of the data (the whole permutation space) or a lower number if we are interested in using only a Monte Carlo sample of them (for further details see Pesarin and Salmaso (2010)). In the following section we provide a more in-depth description of the proposed randomization solution.

2.4.3 Formalization of the procedure

In alternation designs each treatment can arise at any time occasion in the single-case experiment. Once the treatment sequence is randomized, we handle a time process in which we alternate the considered treatments. As already pointed out, we propose to separately smooth the different treatment-labeled time-processes, and then use a measure of the distance between the curves as a test statistic. Firstly we define the concept of exchangeability under the null hypothesis when the underlying model for the response time-process is of interest. Then we discuss both the choice of smoother and of test statistic.

Let us use f to denote the density of response X, $f^{(n)}(\boldsymbol{x})$ to denote the density of the sampling variable $X^{(n)}$, and $\boldsymbol{x} = (x_1, \ldots, x_n)$ to denote the data set. In general the exchangeability of the observed data with respect to groups under the null hypothesis is said to hold if $f^{(n)}(x_1, \ldots, x_n) = f^{(n)}(x_{u_1^*}, \ldots, x_{u_n^*})$, where $(u_1^*, \ldots, u_n^*)$ is any permutation of $(1, \ldots, n)$. We can adapt this characterization to the case of time-dependent data as follows: firstly we introduce time into the formulation, and substitute the concept of 'density' with that of

'model'. Thus we speak about the model underlying the sampling variable time-process $\mathcal{M}^{(n)}(\boldsymbol{x};t)$. We then recall that in the single-case framework, and in particular in the case of alternation designs, the experimental random quantity is the treatment assignment. Hence we can say that the exchangeability of the observed data with respect to treatments under the null hypothesis is said to hold if $\mathcal{M}^{(n)}(x_1,\ldots,x_n;t) = \mathcal{M}^{(n)}(x_{u_1^*},\ldots,x_{u_n^*};t)$, where $(u_1^*,\ldots,u_n^*)$ is any permutation of the original treatment sequence.

As regards the choice of smoother, the best candidate, or best time-series modeling tool in general, of course depends on the real data being handled. The same holds for the choice of most suitable test statistic. An interesting feature of permutation tests is that in principle we can choose any statistic depending on the data we wish to analyze and the study's objective. On the other hand, the opinion of the expert who presents the problem can also be taken into consideration in making this choice. If, for example, the expert expects that all treatments will lead to a trend in the outcome, and that what is of interest is the difference between the slopes and the intercepts of these trends, then we should choose a time-series analysis tool capable of taking into account the underlying slopes and intercepts of the different treatment trends. Therefore the nature of the data is also important. The number of observations per treatment for example can help us to decide which statistical instrument to use. We strongly believe that several implicit and explicit aspects of the data should be taken into account when choosing the test statistic. In this work we propose some possible choices, which of course do not exclude the possibility of applying different ones depending on specific real data features. However, we aim to provide a unique main proposal both for the smoother and the test statistic, which should be better able to detect treatment effects for any feature of the data than the others, at least in the field of educational and behavioural sciences. Two categories of techniques for modeling time-series data are nonparametric smoothers and classical ARIMA models. Nonparametric smoothers can be a good choice when we have no idea what the time-process underlying the data is (for instance when we handle an experiment concerning a new treatment, and not even the field expert knows what to expect from the data). On the other hand, ARIMA modeling can be a useful tool when pilot studies are available, or at least previous analyses on similar problems. In what follows we will consider only the former class of models, for one main reason: ARIMA models are only applicable with equispaced time-series data, which poses a problem for the solution we are proposing. Indeed, since each treatment can arise in every time occasion, then each single treatment time-series does not in general have equispaced observations across time. Nonparametric smoothing on the other hand is particularly suitable for our purposes. We need to somehow measure the distance between the several treatment curves, hence nice objects we can handle are continuous curves to compare, which actually is the output of smoothers. Among nonparametric smoothers many choices exist, such as splines, kernel smoothers and local regression. As a starting point, we consider local regression to be a good choice. It is a widely accepted method, it gives good fitting results even with relatively short time-series, and it is quite simple to understand even for users out of the statistical world.

As regards the choice of test statistics, it should, as already said, be a measure of the distance between the curves. Several choices can be made and, according to Fisher (1935), in principle there is no preferable statistic for all

data features. Depending on different features of the same null hypothesis, different test statistics may be more appropriate. This leads to the Multi-Aspect (MA) testing issue (see Pesarin (2001)). Therefore in our particular framework, we propose to nonparametrically combine several test statistics $t'_m(x)$, for m=1,...,M, about which more details will be given in Section 5.

The permutation procedure for alternation designs can be summarized as follows: let us consider the general case of C treatments, denoted by $A^1, A^2, \ldots, A^C$; also let $s_{A^j}(t; \boldsymbol{x}_{A^j})$, for $j = 1, \ldots, C$, be the resulting local regression smoothed processes from, respectively, data labelled with treatments A^j, for $j = 1, \ldots, C$; $s_{A^j}(t; \boldsymbol{x}_{A^j})$ emphasizes the fact that the smoother is a function of time t, hence defined in every time occasion, given the sub-series $\boldsymbol{x}_{A^j}$ only. Then:

i. consider the original time-series $\boldsymbol{x} = (x_1, \ldots, x_T)$, and the C sub-series $\boldsymbol{x}_{A^j} = (x_{A^j,1_{A^j}}, \ldots, x_{A^j,T_{A^j}})$. Compute the C smoothed processes $s_{A^j}(t; \boldsymbol{x}_{A^j})$, for $j = 1, \ldots, C$;

ii. compute the observed value for the partial test statistics $t'^{,obs}_m(\boldsymbol{x}) = \psi_m(\boldsymbol{x})$ $= \psi_m(s_{A^1}(t; \boldsymbol{x}_{A^1}), \ldots, s_{A^C}(t; \boldsymbol{x}_{A^C}))$, for $m = 1, \ldots, M$;

iii. according to the randomization scheme, perform a randomization of the treatments' labels for $\boldsymbol{x}$, getting the randomized time-series $\boldsymbol{x}^{*,1}$ and the C permuted sub-series $\boldsymbol{x}^{*,1}_{A^j}$. Compute the C permuted smoothed processes $s^{*,1}_{A^j}(t; \boldsymbol{x}^{*,1}_{A^j})$, for $j = 1, \ldots, C$;

iv. compute the value of the partial test statistics $t'^{,*,1}_m = \psi_m(\boldsymbol{x}^{*,1})$ $= \psi_m(s^{*,1}_{A^1}(t; \boldsymbol{x}^{*,1}_{A^1}), \ldots, s^{*,1}_{A^C}(t; \boldsymbol{x}^{*,1}_{A^C}))$, for $m = 1, \ldots, M$;

v. repeat Steps iii. and iv. B times, getting the values of the partial test statistics $t'^{,*,i}_m = \psi_m(\boldsymbol{x}^{*,i}) = \psi_m(s^{*,i}_{A^1}(t; \boldsymbol{x}^{*,i}_{A^1}), \ldots, s^{*,i}_{A^C}(t; \boldsymbol{x}^{*,i}_{A^C}))$, for $i = 1, \ldots, B$, $m = 1, \ldots, M$;

vi. construct the permutation distributions of the partial test statistics under the null hypothesis from the vectors of values $t'^{,*}_m = (t'^{,*,1}_m, \ldots, t'^{,*,B}_m)$, for $m = 1, \ldots, M$;

vii. nonparametrically combine the partial permutation tests, obtaining the permutation distribution of the Multi-Aspect global test $t''_{MA} = \Psi(t'_1, \ldots, t'_M)$;

viii. compute the p-value of the global Multi-Aspect test for treatment effects, for example rejecting the null hypothesis for high values of the test statistic, as $p = \sum_{i=1}^{B} \mathbf{1}_{(t''^{,obs},\infty)}(t''^{,*,i}(\boldsymbol{x}^{*,i}))/B$.

2.5 Power behavior

A Monte Carlo simulation study was performed to demonstrate the reliability and power of the proposed procedures in single-case experiments with univariate responses, in the case of a bidirectional alternative hypothesis. The study was

divided into two main parts. At first, we considered the simplest case of $C = 2$ treatments, where the tested hypotheses are

$$\begin{cases} H_0: & X_{A^1}(t) \overset{\mathcal{M}}{=} X_{A^2}(t) \qquad \forall t \in T \\ H_1: & X_{A^1}(t) \overset{\mathcal{M}}{\neq} X_{A^2}(t), \end{cases},$$

Then, we considered the more general case of $C > 2$ treatments. The reliability of the methods was measured through the estimation of the type I error, while the capability of the procedures to detect treatment effects was deduced from the estimation of power. In the latter case different data scenarios were simulated under the alternative hypothesis, considering several underlying models for the treatment time-processes. We compared the new permutation solutions with the proposal in Bulté and Onghena (2008). As regards the smoother, we performed a local polynomial regression with degree 2, span parameter $\lambda = 0.75$ and tricubic weighting (proportional to $(1 - (\text{dist}/\text{maxdist})^3)^3$).

In the case of $C = 2$, we propose two MA solutions which use the following partial test statistics. The first global test (hereafter labeled $t^{''}_{MA_1}$) is constructed combining the partial tests

$$t^{'}_1(\boldsymbol{x}) = \frac{\sum_{j=1}^{l_{grid}} (s^{*,i}_{A^1}(t_j; \boldsymbol{x}^{*,i}_{A^1}) - s^{*,i}_{A^2}(t_j; \boldsymbol{x}^{*,i}_{A^2}))^2}{l_{grid}}$$
$$t^{'}_2(\boldsymbol{x}) = \max_{j=1,\ldots,l_{grid}} (|s^{*,i}_{A^1}(t_j; \boldsymbol{x}^{*,i}_{A^1}) - s^{*,i}_{A^2}(t_j; \boldsymbol{x}^{*,i}_{A^2})|).$$

Notice that $t^{'}_1$ is a proxy of the area between the two smoothed time subprocesses. In practice this quantity has been approximated by the mean of the squared distances on a regular time grid $t_1, \ldots, t_{l_{grid}}$ of length $l_{grid} < 100$. Instead $t^{'}_2$ is the well-known two-sample Kolmogorov-Smirnov type statistic. The second global test (hereafter labeled $t^{''}_{MA_2}$) is constructed combining the above presented partial tests with the one proposed in Bulté and Onghena (2008) (hereafter denoted by t_{BO}), which uses the absolute value of the difference between the sample means in the two treatment groups. In order to increase the procedure's power, we propose to use Tippett combining function (see Pesarin (2001) and Pesarin and Salmaso (2010)), given that under the alternative hypothesis it chooses the smallest partial p-value. Hence

$$t^{''}_{MA_k}(\boldsymbol{x}) = \Psi(t^{'}_1, \ldots, t^{'}_{M_k}) = \max\left(1 - \lambda^{'}_1, \ldots, 1 - \lambda^{'}_{M_k}\right),$$

where $\lambda^{'}_m$, $m = 1, \ldots, M_k$, indicates the permutation p-value of the mth partial test, so in our case $k \in \{1, 2\}$ and $M_k \in \{2, 3\}$.

The simulation study was performed generating $CMC = 1000$ samples from the assumed underlying models; for the implementation of the permutation tests we used $B = 1000$ permutations (drawing a Monte Carlo random sample from the permutation space). We considered errors coming from both the standard normal and the Student t distribution with 2 degrees of freedom. We considered four lengths for the entire time series, $n = 30, 50, 70, 100$, and five underlying ARIMA models, $AR(1)$, $MA(1)$, $ARMA(1,1)$, and $ARIMA(1,1,1)$ with and without constant term $\mu = 1$, each with several values for the autoregressive and moving-average parameters, $\phi, \theta = -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5$.

Table 2.1 shows, for the normal errors case, the reliability of the proposed methods under the null hypothesis: the obtained results suggest that all the three methods behave in a reliable way under the null hypothesis, controlling the I type error at the nominal level.

Table 2.1: Estimated I type error for the alternation design solutions, $C = 2$ treatments; errors simulated from the standard normal distribution.

		t''_{MA_1}, $l_{grid} = 100$ Length of the series				t''_{MA_2}, $l_{grid} = 100$ Length of the series				t_{BO} Length of the series			
Model	*Parameters*	$n = 30$	$n = 50$	$n = 70$	$n = 100$	$n = 30$	$n = 50$	$n = 70$	$n = 100$	$n = 30$	$n = 50$	$n = 70$	$n = 100$
$AR(1)$	$\phi = -0.3$	0.054	0.039	0.041	0.047	0.047	0.042	0.060	0.049	0.040	0.046	0.048	0.040
	$\phi = -0.2$	0.053	0.061	0.054	0.055	0.054	0.055	0.053	0.065	0.047	0.053	0.046	0.068
	$\phi = -0.1$	0.037	0.057	0.051	0.046	0.040	0.048	0.052	0.043	0.063	0.037	0.048	0.034
	$\phi = 0.1$	0.055	0.045	0.049	0.050	0.056	0.044	0.031	0.060	0.050	0.046	0.042	0.056
	$\phi = 0.2$	0.050	0.046	0.050	0.042	0.052	0.042	0.064	0.044	0.066	0.047	0.051	0.052
	$\phi = 0.3$	0.056	0.044	0.053	0.054	0.056	0.044	0.053	0.054	0.046	0.041	0.045	0.044
	$\phi = 0.4$	0.031	0.046	0.052	0.047	0.042	0.050	0.052	0.041	0.044	0.055	0.049	0.059
	$\phi = 0.5$	0.045	0.041	0.040	0.058	0.040	0.053	0.055	0.054	0.052	0.052	0.054	0.055
$MA(1)$	$\theta = -0.3$	0.056	0.045	0.059	0.059	0.056	0.045	0.059	0.059	0.055	0.046	0.046	0.057
	$\theta = -0.2$	0.051	0.039	0.054	0.049	0.051	0.039	0.054	0.049	0.052	0.037	0.056	0.058
	$\theta = -0.1$	0.035	0.042	0.049	0.060	0.035	0.042	0.049	0.060	0.050	0.048	0.046	0.065
	$\theta = 0.1$	0.055	0.052	0.052	0.045	0.055	0.052	0.052	0.045	0.052	0.058	0.055	0.048
	$\theta = 0.2$	0.045	0.057	0.050	0.056	0.045	0.057	0.050	0.056	0.046	0.049	0.048	0.046
	$\theta = 0.3$	0.057	0.054	0.051	0.052	0.057	0.054	0.051	0.052	0.045	0.044	0.061	0.046
	$\theta = 0.4$	0.055	0.048	0.047	0.055	0.052	0.048	0.047	0.052	0.050	0.047	0.041	0.057
	$\theta = 0.5$	0.058	0.039	0.042	0.054	0.058	0.039	0.042	0.054	0.053	0.043	0.042	0.066
$ARMA$ $(1,1)$	$\phi, \theta = -0.3$	0.050	0.045	0.054	0.052	0.053	0.050	0.058	0.060	0.053	0.047	0.054	0.056
	$\phi, \theta = -0.2$	0.037	0.041	0.043	0.053	0.051	0.034	0.042	0.057	0.061	0.052	0.039	0.059
	$\phi, \theta = -0.1$	0.049	0.052	0.056	0.046	0.051	0.049	0.052	0.049	0.049	0.045	0.043	0.049
	$\phi, \theta = 0$	0.051	0.049	0.042	0.057	0.041	0.049	0.039	0.053	0.041	0.047	0.038	0.046
	$\phi, \theta = 0.1$	0.053	0.061	0.052	0.046	0.043	0.054	0.053	0.056	0.054	0.054	0.053	0.052
	$\phi, \theta = 0.2$	0.050	0.051	0.048	0.055	0.055	0.051	0.054	0.060	0.053	0.052	0.063	0.057
	$\phi, \theta = 0.3$	0.052	0.048	0.040	0.059	0.043	0.044	0.040	0.054	0.041	0.057	0.052	0.048
	$\phi, \theta = 0.4$	0.055	0.054	0.057	0.052	0.062	0.057	0.054	0.047	0.051	0.071	0.059	0.040
	$\phi, \theta = 0.5$	0.046	0.052	0.050	0.051	0.053	0.042	0.047	0.052	0.046	0.044	0.046	0.058
$ARIMA$ $(1,1,1)$	$\phi, \theta = -0.3$, $\mu = 0$	0.054	0.044	0.054	0.050	0.051	0.056	0.055	0.057	0.046	0.053	0.050	0.058
	$\phi, \theta = -0.2$, $\mu = 0$	0.048	0.053	0.055	0.058	0.043	0.049	0.062	0.047	0.047	0.051	0.059	0.034
	$\phi, \theta = -0.1$, $\mu = 0$	0.045	0.052	0.056	0.045	0.052	0.055	0.049	0.046	0.051	0.051	0.051	0.049
	$\phi, \theta = 0$, $\mu = 0$	0.044	0.039	0.062	0.049	0.053	0.049	0.059	0.037	0.051	0.057	0.049	0.044
	$\phi, \theta = 0.1$, $\mu = 0$	0.053	0.056	0.040	0.052	0.053	0.052	0.041	0.047	0.055	0.036	0.048	0.034
	$\phi, \theta = 0.2$, $\mu = 0$	0.050	0.042	0.048	0.048	0.052	0.051	0.049	0.045	0.056	0.050	0.055	0.052
	$\phi, \theta = 0.3$, $\mu = 0$	0.048	0.031	0.041	0.051	0.056	0.027	0.053	0.050	0.067	0.043	0.050	0.058
	$\phi, \theta = 0.4$, $\mu = 0$	0.047	0.056	0.048	0.046	0.049	0.066	0.056	0.043	0.057	0.062	0.051	0.037
	$\phi, \theta = 0.5$, $\mu = 0$	0.048	0.041	0.053	0.039	0.043	0.056	0.050	0.050	0.046	0.068	0.058	0.051
$ARIMA$ $(1,1,1)$	$\phi, \theta = -0.3$, $\mu = 1$	0.054	0.036	0.049	0.046	0.051	0.038	0.047	0.043	0.040	0.057	0.041	0.049
	$\phi, \theta = -0.2$, $\mu = 1$	0.043	0.052	0.040	0.057	0.051	0.056	0.050	0.051	0.053	0.052	0.052	0.041
	$\phi, \theta = -0.1$, $\mu = 1$	0.055	0.053	0.044	0.047	0.050	0.053	0.040	0.046	0.054	0.053	0.040	0.046
	$\phi, \theta = 0$, $\mu = 1$	0.049	0.038	0.052	0.049	0.048	0.047	0.053	0.056	0.044	0.048	0.045	0.049
	$\phi, \theta = 0.1$, $\mu = 1$	0.055	0.053	0.042	0.042	0.048	0.054	0.037	0.054	0.048	0.056	0.046	0.043
	$\phi, \theta = 0.2$, $\mu = 1$	0.059	0.046	0.041	0.049	0.050	0.041	0.058	0.053	0.047	0.048	0.059	0.046
	$\phi, \theta = 0.3$, $\mu = 1$	0.065	0.045	0.043	0.050	0.055	0.045	0.053	0.051	0.054	0.054	0.041	0.051
	$\phi, \theta = 0.4$, $\mu = 1$	0.068	0.046	0.048	0.054	0.065	0.043	0.055	0.051	0.068	0.047	0.061	0.052
	$\phi, \theta = 0.5$, $\mu = 1$	0.043	0.049	0.068	0.058	0.048	0.042	0.056	0.048	0.053	0.046	0.045	0.053

The power of the procedures was also investigated under different data scenarios. Again we considered four lengths for the entire time series, $n = 30, 50, 70, 100$, and generated the difference in treatment effects by adding two different deterministic parts to the same underlying stochastic ARMA models. The rationale behind this choice is that we expect a patient's response to different treatments to remain the same in the underlying autocorrelation structure,

and change only in the deterministic component of the model. More precisely, data has been simulated as follows:

$$\begin{aligned} x_t =& \text{deterministic component} + \text{stochastic component} \\ =& \begin{cases} f_A(t) + \phi x_{t-1} + \theta\epsilon_{t-1} + \epsilon_t & \text{if } Treatment_t = A \\ f_B(t) + \phi x_{t-1} + \theta\epsilon_{t-1} + \epsilon_t & \text{if } Treatment_t = B \end{cases}, \end{aligned}$$

with $\epsilon_t \sim N(0,1)$ or $\epsilon_t \sim t_2$ *iid*. As regards the deterministic part of the model, we simulated treatment A as both a placebo and an active treatment, and treatment B as the active treatment, mimicking both a difference in level and slope of the trend between the two treatment effects. We simulated a difference in levels between the treatments as follows:

$$x_t = \begin{cases} \phi x_{t-1} + \theta\epsilon_{t-1} + \epsilon_t & \text{if } Treatment_t = A \\ \delta + \phi x_{t-1} + \theta\epsilon_{t-1} + \epsilon_t & \text{if } Treatment_t = B \end{cases}.$$

Three ARMA models, $AR(1)$, $MA(1)$ and $ARMA(1,1)$ were considered, each with several values for the autoregressive and moving-average parameters, $\phi, \theta = -0.3, -0.1, 0.1, 0.3, 0.5$. Additionally, several values were considered for the level differences $\delta = 1, 3, 5, 10$. Therefore we simulated a difference in the slope of the trend between treatments as follows:

$$x_t = \begin{cases} \beta_A + x_{t-1} + \epsilon_t & \text{if } Treatment_t = A \\ \beta_B + x_{t-1} + \epsilon_t & \text{if } Treatment_t = B \end{cases}.$$

Notice that treatment A mimics either a 'no treatment' setting when $\beta_A = 0$ or an active treatment when $\beta_A \neq 0$, while treatment B always has a linear trend in the response. Indeed, again several values have been considered for the slope differences combining different values for the slopes $\beta_A = -50, -30, -10, -5, -1, 0, 1$ and $\beta_B = 1, 5, 10, 30, 50$. The results are reported in Tables 2.2 and 2.3 for the level and slope difference respectively. Some of the results for the normal error case are also displayed in Figures 2.1 and 2.2 for the case of a treatment difference in level and a treatment difference in slope respectively. While t_{BO} shows greater power when the two treatments are simulated with different levels, it performs worse when we consider a difference in slopes; in the latter case the estimated rejection probability is greater than the nominal level α only when the slopes of the two treatments have opposite signs. Regarding the MA solutions, t''_{MA_2} performs better than t''_{MA_1} in the case of a difference in levels, but not a difference in slopes, where t''_{MA_1} often reports greater power. These results are coherent with the way in which the MA solutions are constructed. In t''_{MA_2}, use of the partial test on the difference between the sample means allows us to increase power when the treatments differ in level, but it decreases it in the case of different slopes. It should therefore be noted that in the case of different slopes, t''_{MA_2} is always more powerful than t_{BO}. At the end, the power of all procedures increases with the number of observations and with the difference between the deterministic parts of the two models. Moreover, the power values reach 1 as n and δ increase when we simulate differences in levels, while this is not the case for differences in slopes, where the power function seems to have a horizontal asymptote below 1. We must remark that the model used to simulate the data with difference in slopes produces a non stationary process, where both the average level and

the variance of the process are not fixed but they increase with time. This, in particular, means that the population mean is not finite under the null hypothesis, when the number of time observations goes to infinity: this means that the data simulated in this way do not respect the first sufficient condition for the weak consistency of permutation tests (see Pesarin and Salmaso (2011)), leading our test in this case to be not consistent. However it is worth underlying that this is a very particular situation, which is usually treated, for instance within the ARMA theory, differentiating the series and then estimating the model of the differentiated time-series. One could follow the same idea, then, in order to solve the problem of the horizontal asymptote below 1 for the power function of the test: a possible solution could be to differentiate the series, getting a stationary time-process similar to the ones simulated for the difference in levels. More specifically we would get a process with only homoscedastic innovations and no autocorrelation, that represents a particular case of the model considered for the difference in levels. For this simulation model the good performances of the permutations tests have already been shown. Regarding the effect of the autocorrelation on the power behavior of the proposed methods, in general the estimated rejection probabilities change when changing the autocorrelation values. They slightly decrease as the autocorrelation parameters of the ARMA models (it does not matter which model is considered) move, in absolute value, away from zero. Similar results have been obtained for the case of errors

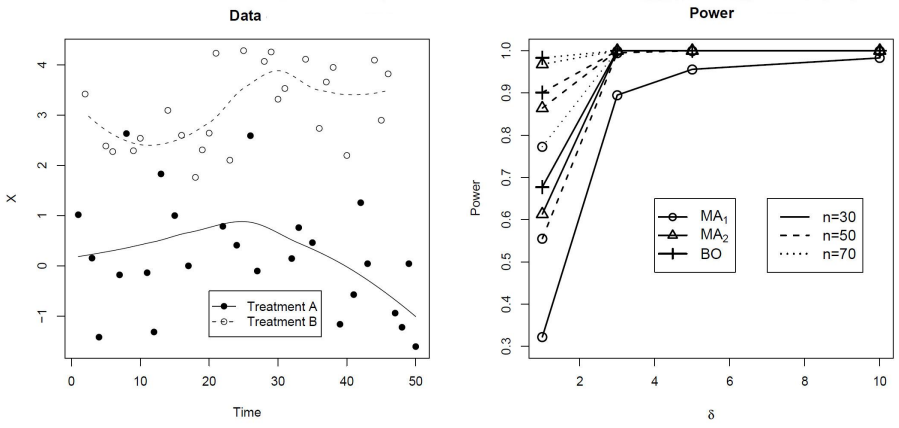

Figure 2.1: Simulation study. On the left an example of simulated data with a difference between levels with an $ARMA(1,1)$ model for the errors with parameters $\phi, \theta = 0.1$, $n = 50$ and $\delta = 3$; on the right power functions for the three considered permutation solutions in the case of an $ARMA(1,1)$ model for the errors with parameters $\phi, \theta = 0.1$, while n and δ are varying.

generated from the t_2 distribution. All the procedures control the type-I error at the nominal level; the power simply decreases with respect to the normal errors case, but the behaviour of the three procedures remains the same.

Considering the more general case of $C > 2$ treatments, the new solution is built performing the MA permutation tests on all pairwise comparisons between treatments, and then combining them in a global test. Formally, for $i = 1, \ldots, B$,

Table 2.2: Estimated power for the alternation design solutions, $C = 2$ treatments. Simulated difference in level between the two treatments' effects; errors simulated from the standard normal distribution.

			t''_{MA_1}, $l_{grid} = 100$ Length of the series				t''_{MA_2}, $l_{grid} = 100$ Length of the series				t_{BO} Length of the series			
δ	*Model*	*Parameters*	$n = 30$	$n = 50$	$n = 70$	$n = 100$	$n = 30$	$n = 50$	$n = 70$	$n = 100$	$n = 30$	$n = 50$	$n = 70$	$n = 100$
		$\phi = -0.3$	0.298	0.540	0.720	0.796	0.616	0.855	0.949	0.992	0.688	0.902	0.966	0.993
		$\phi = -0.1$	0.333	0.593	0.753	0.831	0.644	0.870	0.970	0.998	0.704	0.910	0.982	0.999
	$AR(1)$	$\phi = 0.1$	0.324	0.572	0.741	0.824	0.622	0.895	0.959	1.000	0.706	0.919	0.974	1.000
		$\phi = 0.3$	0.271	0.550	0.716	0.791	0.579	0.839	0.954	0.992	0.655	0.884	0.967	0.996
		$\phi = 0.5$	0.254	0.453	0.632	0.729	0.499	0.750	0.903	0.970	0.563	0.807	0.931	0.988
		$\theta = -0.3$	0.278	0.495	0.708	0.800	0.602	0.858	0.955	0.993	0.673	0.897	0.971	0.997
		$\theta = -0.1$	0.305	0.570	0.748	0.838	0.620	0.885	0.966	0.998	0.701	0.920	0.981	1.000
1	$MA(1)$	$\theta = 0.1$	0.329	0.589	0.778	0.827	0.639	0.877	0.975	0.999	0.700	0.912	0.987	0.999
		$\theta = 0.3$	0.325	0.603	0.725	0.821	0.615	0.865	0.951	0.993	0.676	0.891	0.968	0.994
		$\theta = 0.5$	0.295	0.555	0.702	0.778	0.564	0.820	0.935	0.979	0.625	0.865	0.953	0.985
		$\phi, \theta = -0.3$	0.236	0.386	0.571	0.695	0.503	0.734	0.892	0.966	0.582	0.798	0.937	0.986
		$\phi, \theta = -0.1$	0.306	0.575	0.725	0.791	0.629	0.888	0.967	0.994	0.700	0.929	0.981	0.994
	$ARMA$ $(1,1)$	$\phi, \theta = 0.1$	0.322	0.555	0.773	0.829	0.613	0.864	0.968	0.991	0.677	0.901	0.983	0.997
		$\phi, \theta = 0.3$	0.273	0.499	0.634	0.719	0.535	0.776	0.897	0.973	0.606	0.827	0.922	0.983
		$\phi, \theta = 0.5$	0.202	0.328	0.468	0.513	0.337	0.537	0.716	0.845	0.385	0.604	0.761	0.884
		$\phi = -0.3$	0.881	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi = -0.1$	0.899	0.994	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$AR(1)$	$\phi = 0.1$	0.883	0.998	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi = 0.3$	0.875	0.990	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi = 0.5$	0.766	0.982	0.994	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = -0.3$	0.881	0.996	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = -0.1$	0.898	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	$MA(1)$	$\theta = 0.1$	0.912	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = 0.3$	0.910	0.996	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = 0.5$	0.889	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = -0.3$	0.821	0.995	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = -0.1$	0.882	0.997	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$ARMA$ $(1,1)$	$\phi, \theta = 0.1$	0.895	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = 0.3$	0.847	0.988	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = 0.5$	0.742	0.951	0.990	0.997	0.994	1.000	1.000	1.000	0.997	1.000	1.000	1.000
		$\phi = -0.3$	0.943	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi = -0.1$	0.950	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$AR(1)$	$\phi = 0.1$	0.961	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi = 0.3$	0.922	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi = 0.5$	0.869	0.992	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = -0.3$	0.961	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = -0.1$	0.950	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	$MA(1)$	$\theta = 0.1$	0.968	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = 0.3$	0.958	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = 0.5$	0.963	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = -0.3$	0.927	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = -0.1$	0.957	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$ARMA$ $(1,1)$	$\phi, \theta = 0.1$	0.956	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = 0.3$	0.931	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = 0.5$	0.852	0.988	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi = -0.3$	0.966	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi = -0.1$	0.968	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$AR(1)$	$\phi = 0.1$	0.963	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi = 0.3$	0.958	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi = 0.5$	0.922	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = -0.3$	0.966	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = -0.1$	0.976	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	$MA(1)$	$\theta = 0.1$	0.980	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = 0.3$	0.982	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\theta = 0.5$	0.977	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = -0.3$	0.958	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = -0.1$	0.975	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$ARMA$ $(1,1)$	$\phi, \theta = 0.1$	0.983	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = 0.3$	0.947	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = 0.5$	0.899	0.998	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

the ith permutation value of the test is constructed as follows:

$$\begin{aligned} t^{*,i} =& \psi(s^{*,i}_{A^1}(t; \boldsymbol{x}^{*,i}_{A^1}), \ldots, s^{*,i}_{A^C}(t; \boldsymbol{x}^{*,i}_{A^C})) \\ =& \sum_{l<s,\, l,s \in \{1,\ldots,C\}} \frac{t^{''*,i}_{MA_k;l,s}(\boldsymbol{x})}{sd\left(t^{''*,}_{MA_k;l,s}(\boldsymbol{x})\right)}, \end{aligned}$$

where $t^{''*,i}_{MA_k;l,s}(x)$ denotes the MA permutation test related to the comparison of treatments A^l and A^s. Notice that the procedure coincides with operating a

Table 2.3: Estimated power for the alternation design solutions, $C = 2$ treatments. Simulated difference in slope between the two treatments' effects; errors simulated from the standard normal distribution.

		$t''_{MA_1}, l_{grid} = 100$ Length of the series				$t''_{MA_2}, l_{grid} = 100$ Length of the series				t_{BO} Length of the series			
Slope A	*Slope B*	$n = 30$	$n = 50$	$n = 70$	$n = 100$	$n = 30$	$n = 50$	$n = 70$	$n = 100$	$n = 30$	$n = 50$	$n = 70$	$n = 100$
$\beta_A = 0$	$\beta_B = 1$	0.097	0.089	0.124	0.120	0.081	0.082	0.111	0.093	0.060	0.054	0.066	0.047
	$\beta_B = 5$	0.289	0.367	0.388	0.363	0.229	0.266	0.298	0.284	0.068	0.064	0.063	0.051
	$\beta_B = 10$	0.336	0.430	0.467	0.415	0.261	0.317	0.352	0.306	0.066	0.060	0.057	0.059
	$\beta_B = 30$	0.351	0.451	0.468	0.446	0.265	0.330	0.371	0.328	0.053	0.058	0.061	0.039
$\beta_A = 1$	$\beta_B = 1$	0.043	0.056	0.042	0.062	0.045	0.061	0.039	0.053	0.047	0.046	0.049	0.042
	$\beta_B = 5$	0.282	0.346	0.385	0.335	0.195	0.248	0.296	0.255	0.048	0.040	0.064	0.050
	$\beta_B = 10$	0.353	0.415	0.474	0.402	0.274	0.297	0.354	0.296	0.049	0.052	0.049	0.047
	$\beta_B = 30$	0.358	0.434	0.488	0.463	0.276	0.321	0.374	0.346	0.058	0.058	0.052	0.051
	$\beta_B = 50$	0.345	0.439	0.487	0.439	0.247	0.331	0.374	0.314	0.057	0.051	0.062	0.057
$\beta_A = -1$	$\beta_B = 1$	0.163	0.230	0.209	0.206	0.183	0.225	0.221	0.209	0.174	0.189	0.183	0.180
$\beta_A = -5$	$\beta_B = 5$	0.339	0.442	0.460	0.402	0.383	0.460	0.477	0.450	0.362	0.398	0.382	0.423
$\beta_A = -10$	$\beta_B = 10$	0.351	0.452	0.496	0.417	0.405	0.451	0.498	0.492	0.379	0.425	0.421	0.423
$\beta_A = -30$	$\beta_B = 30$	0.377	0.443	0.463	0.436	0.438	0.478	0.476	0.492	0.408	0.428	0.427	0.440
$\beta_A = -50$	$\beta_B = 50$	0.361	0.449	0.514	0.434	0.414	0.470	0.519	0.493	0.385	0.414	0.423	0.43

nonparametric combination of the *MA* pairwise comparisons permutation tests using a Direct combining function (see Pesarin (2001) and Pesarin and Salmaso (2010)). In the summation we divide the terms by their standard deviation in order to standardize the combined test statistics.

Again we worked with $CMC = 1000$, $B = 1000$, and the cases of $C = 3$ and $C = 5$ were considered. For $C = 3$ and $C = 5$, as lengths we simulated for the entire time series respectively $n = 45, 75, 150$ and $n = 75, 125, 250$ and three underlying ARMA models, $AR(1)$, $MA(1)$ and $ARMA(1, 1)$, each with several values for the autoregressive and moving-average parameters, $\phi, \theta = -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5$. The reliability of the methods is reported in Tables 2.4 and 2.5 for the cases $C = 3$ and $C = 5$ respectively. The results confirm the reliability of the proposed solutions, showing the control of the type-I error at the nominal level $\alpha = 0.05$.

We performed a simulation study to show that the power of the *MA* solutions increases with sample size, distance from the null hypothesis and number of pairwise comparisons under the alternative hypothesis. As with the case of $C = 2$, we generated the treatment effect adding different treatments to the data generating processes and different deterministic parts to the same stochastic underlying ARMA models. We considered both a difference in level and slope of the trend among the several treatment effects, following a similar scheme to the one used for $C = 2$. We simulated a difference in levels among treatments considering three ARMA models, $AR(1)$, $MA(1)$ and $ARMA(1, 1)$, each with several values for the autoregressive and moving-average parameters, $\phi, \theta = -0.3, -0.1, 0.1, 0.3, 0.5$. Moreover, for both the cases of $C = 3$ and $C = 5$, four settings were considered for the level differences, which are described in Table 2.6. The settings differ both in terms of the total difference among the treatment levels (δ_{tot}) and the number of false null hypotheses in the family of all pairwise comparisons. The results, for normal errors, are reported in Tables 2.7 and 2.8 for the cases $C = 3$ and $C = 5$. Some of the results for the case of $C = 3$ are also displayed in Figure 2.3. Notice that the power of both MA solutions increases with sample size; furthermore, we obtain a higher power when we also increase the distance from the global null hypothesis (in Figure 2.3 summarized as a combination between the number of false partial null hypotheses and the

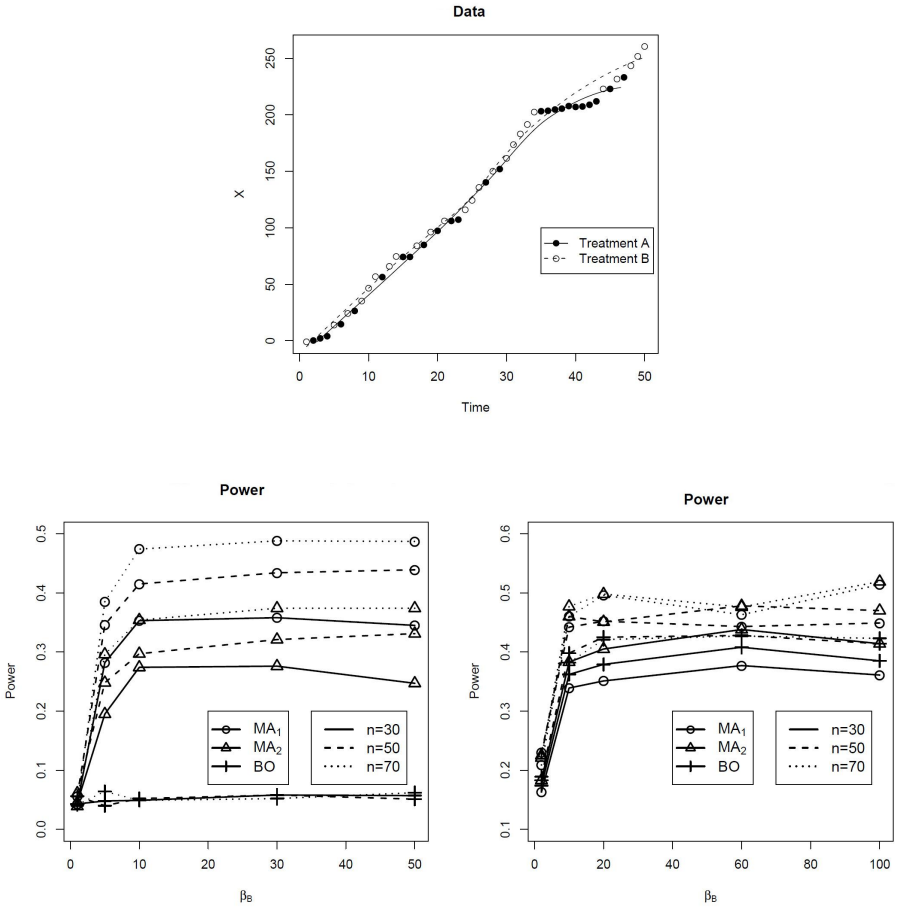

Figure 2.2: Simulation study. At the top an example of simulated data with a difference between slopes with $n = 50$, $\beta_A = 1$ and $\beta_B = 10$; below left power functions for the three considered permutation solutions in the case of $\beta_A = 1$ while n and β_B are varying; below right power functions for the three considered permutation solutions in the case of $\beta_A = -\beta_B$ while n and β_B are varying.

total difference among the treatment levels, δ_{tot}); this behavior is less clear in the case of $C = 5$, probably due to the coexistence of too many different deterministic parts in the full time series, which makes it more difficult for the tests to distinguish between different effects.

We simulated a difference in slopes among treatments considering four settings for both the cases of $C = 3$ and $C = 5$, described in Table 2.9: the settings differ both in terms of the total difference among the treatment slopes (β_{tot}) and the number of false null hypotheses in the family of all pairwise comparisons. The results are reported in Tables 2.10 and 2.11 for the cases $C = 3$ and $C = 5$ respectively. Again in this case the power of both MA solutions increases with sample size; furthermore we obtain a higher power when we also increase the number of false null hypotheses in the family, even if this behavior is less clear

Table 2.4: Estimated I type error for the alternation design solutions t''_{MA_1} and t''_{MA_2}, $C = 3$ treatments; errors simulated from the standard normal distribution.

		t''_{MA_1}, $l_{grid} = 100$, $C = 3$ Length of the series			t''_{MA_2}, $l_{grid} = 100$, $C = 3$ Length of the series		
Model	*Parameters*	$n = 45$	$n = 75$	$n = 150$	$n = 45$	$n = 75$	$n = 150$
$AR(1)$	$\phi = -0.3$	0.046	0.059	0.045	0.043	0.032	0.050
	$\phi = -0.2$	0.054	0.065	0.055	0.052	0.036	0.056
	$\phi = -0.1$	0.049	0.047	0.044	0.065	0.051	0.066
	$\phi = 0.1$	0.055	0.052	0.047	0.054	0.047	0.054
	$\phi = 0.2$	0.058	0.056	0.047	0.048	0.048	0.055
	$\phi = 0.3$	0.054	0.054	0.056	0.048	0.054	0.051
	$\phi = 0.4$	0.062	0.046	0.052	0.040	0.043	0.056
	$\phi = 0.5$	0.051	0.053	0.053	0.041	0.049	0.045
$MA(1)$	$\theta = -0.3$	0.061	0.047	0.060	0.061	0.047	0.060
	$\theta = -0.2$	0.041	0.055	0.065	0.041	0.055	0.065
	$\theta = -0.1$	0.047	0.052	0.055	0.047	0.052	0.055
	$\theta = 0.1$	0.038	0.046	0.055	0.038	0.046	0.055
	$\theta = 0.2$	0.058	0.056	0.048	0.058	0.056	0.048
	$\theta = 0.3$	0.063	0.049	0.055	0.063	0.049	0.055
	$\theta = 0.4$	0.055	0.052	0.056	0.055	0.049	0.048
	$\theta = 0.5$	0.052	0.053	0.046	0.052	0.053	0.046
$ARMA$ $(1,1)$	$\phi, \theta = -0.3$	0.043	0.060	0.043	0.054	0.047	0.048
	$\phi, \theta = -0.2$	0.047	0.056	0.045	0.044	0.046	0.046
	$\phi, \theta = -0.1$	0.044	0.052	0.043	0.058	0.043	0.048
	$\phi, \theta = 0$	0.060	0.053	0.041	0.044	0.056	0.047
	$\phi, \theta = 0.1$	0.047	0.050	0.058	0.071	0.064	0.051
	$\phi, \theta = 0.2$	0.041	0.053	0.063	0.055	0.044	0.057
	$\phi, \theta = 0.3$	0.060	0.053	0.054	0.036	0.050	0.054
	$\phi, \theta = 0.4$	0.045	0.052	0.055	0.049	0.054	0.058
	$\phi, \theta = 0.5$	0.049	0.046	0.055	0.042	0.050	0.053

than in the case of a difference in levels among the treatment effects. Again we find a less clear increase of power in the case of $C = 5$. An the end, in the case of both a difference in levels and a difference in slopes among the treatments, it has to be underlined that the order between the power of the two MA solutions is the same as the order obtained in the related simulation settings for the case of $C = 2$. The coherence between the simple case and the generalization to $C > 2$ is of course a good feature of the method. Regarding the effect of the autocorrelation on the power behavior, we can notice a similar behavior in case of $C = 2$ treatments: the results show a decrease of the estimated rejection probabilities as the autocorrelation parameters move, in absolute value, away from zero.

2.6 Real clinical psychology data application

We applied the proposed MA solutions to real data from a single-case experiment carried out at the Virga Jesse Hospital in Hasselt, Belgium (see Baplu (2005)). The data were collected from a 17-year old adolescent who had concentration

Table 2.5: Estimated I type error for the alternation design solutions t''_{MA_1} and t''_{MA_2}, $C = 5$ treatments; errors simulated from the standard normal distribution.

		t''_{MA_1}, $l_{grid} = 100$, $C = 5$ Length of the series			t''_{MA_2}, $l_{grid} = 100$, $C = 5$ Length of the series		
Model	*Parameters*	$n = 75$	$n = 125$	$n = 250$	$n = 75$	$n = 125$	$n = 250$
$AR(1)$	$\phi = -0.3$	0.053	0.072	0.059	0.049	0.053	0.048
	$\phi = -0.2$	0.067	0.046	0.055	0.049	0.049	0.049
	$\phi = -0.1$	0.049	0.053	0.063	0.052	0.042	0.043
	$\phi = 0.1$	0.056	0.047	0.052	0.043	0.051	0.056
	$\phi = 0.2$	0.057	0.062	0.031	0.056	0.062	0.044
	$\phi = 0.3$	0.048	0.039	0.048	0.045	0.051	0.055
	$\phi = 0.4$	0.047	0.063	0.049	0.038	0.054	0.046
	$\phi = 0.5$	0.052	0.049	0.066	0.056	0.033	0.051
$MA(1)$	$\theta = -0.3$	0.052	0.055	0.049	0.053	0.055	0.048
	$\theta = -0.2$	0.063	0.061	0.052	0.063	0.061	0.052
	$\theta = -0.1$	0.059	0.052	0.052	0.059	0.052	0.052
	$\theta = 0.1$	0.051	0.050	0.041	0.051	0.050	0.041
	$\theta = 0.2$	0.051	0.053	0.058	0.051	0.053	0.058
	$\theta = 0.3$	0.058	0.056	0.046	0.058	0.056	0.046
	$\theta = 0.4$	0.053	0.051	0.048	0.047	0.050	0.046
	$\theta = 0.5$	0.045	0.046	0.047	0.045	0.046	0.047
$ARMA$ $(1,1)$	$\phi, \theta = -0.3$	0.050	0.053	0.052	0.052	0.055	0.062
	$\phi, \theta = -0.2$	0.036	0.066	0.049	0.043	0.048	0.043
	$\phi, \theta = -0.1$	0.066	0.047	0.053	0.056	0.049	0.058
	$\phi, \theta = 0$	0.060	0.054	0.052	0.049	0.051	0.042
	$\phi, \theta = 0.1$	0.056	0.057	0.042	0.053	0.056	0.044
	$\phi, \theta = 0.2$	0.045	0.040	0.060	0.050	0.046	0.052
	$\phi, \theta = 0.3$	0.042	0.047	0.047	0.049	0.042	0.043
	$\phi, \theta = 0.4$	0.047	0.052	0.046	0.043	0.049	0.065
	$\phi, \theta = 0.5$	0.043	0.052	0.042	0.047	0.050	0.046

Table 2.6: Settings of treatments' levels δ_j, $j \in \{A, B, C, D, E\}$; cases of $C = 3$ and $C = 5$.

C	Setting	δ_{tot}	Proportion of false H_0	δ_A	δ_B	δ_C	δ_D	δ_E
$C = 3$	1	2	2/3	0	0	1		
	2	4	3/3	0	1	2		
	3	4	2/3	0	0	2		
	4	8	3/3	0	2	4		
$C = 5$	5	4	4/10	0	0	0	0	1
	6	10	7/10	0	0	0	1	2
	7	8	4/10	0	0	0	0	2
	8	20	7/10	0	0	0	2	4

problems, but who did not meet the formal DSM-IV-TR criteria for Attention Deficit Disorder (ADD). The patient came to consultation with his mother and complained about increasing attention problems at school and worsening of his school grades. They had heard about the effect of methylphenidate (Ritalin) on concentration and wondered whether they could obtain a prescription. Although

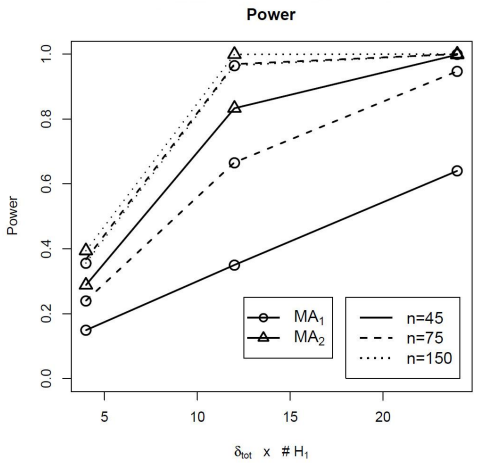

Figure 2.3: Simulation study. Power functions for the two considered MA solutions in the case of $C = 3$, an $ARMA(1, 1)$ model for the errors with parameters $\phi, \theta = 0.1$, while n, δ_{tot} and the number of false partial null hypotheses are varying.

there is some evidence that Ritalin has a positive effect on concentration, even in the absence of ADD, the general use of this drug for mild conditions (or to improve intellectual performance) is still very controversial, and the effect could vary strongly from individual to individual (see Agay et al. (2010) and Crutchley and Temlett (1999)). Therefore, the physician, in close collaboration with a researcher from the Katholieke Universiteit of Leuven (Belgium), proposed to design a single-case experiment to determine the effectiveness (and also the possible side-effects) of this drug, and only to continue using of the drug after convincing experimental results.

The experiment was approved by the Ethics Committee of the hospital and the patient and his parents gave informed consent including a detailed account of the purpose, risks, and design of the study. The study was set up as a randomized, double-blind, placebo-controlled single-case experiment, comparing Ritalin to placebo. The study was randomized because six treatment periods and six placebo periods were randomly assigned to twelve available time blocks (i.e., one treatment order was randomly selected from $12!/6!6! = 924$ possibilities). A time block consisted of two days. Because Ritalin washes out within five hours, no carryover effects from one day to the next were expected (see Tannock et al. (1989)). The study was double-blind because neither the patient and his family, nor the physician and the researcher knew the actual treatment order. A sealed envelope containing the actual treatment order was prepared by the researcher's supervisor and handed directly to the hospital pharmacist who prepared identical capsules (same shape, color, and taste) for the active medication and the placebo. The capsules had to be taken twice a day, and also the primary outcome measure was taken twice a day. Because a time block consisted of two days, this means that four measurements were available for each treatment period as well as for each placebo period. The outcome measure

was the sum of three scores on items from the ADHD Rating Scale IV, which were selected as most relevant by the patient, his parents, and the physician. The patient gave a score on an anchored scale from 0 (never or rarely) to 3 (very often) in a diary for each of the following items:

- Does not follow through on instructions and fails to finish work.
- Fails to give close attention to details or makes careless mistakes in schoolwork.
- Has difficulty sustaining attention in tasks or play activities.

In the diary also more general comments, other complaints and potential adverse or side effects could be registered.

Figure 2.4 shows the summed scores of these three items. We wish to highlight the fact that the same application was considered in Bulté and Onghena (2008), where a directional alternative hypothesis was considered. We performed

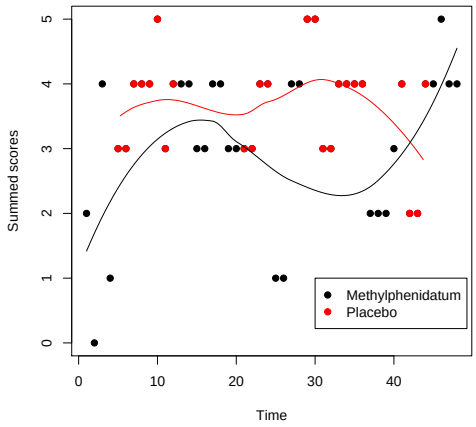

Figure 2.4: Real data from Baplu (2005): summed scores and smoothed time sub-processes.

the two *MA* proposals (t''_{MA_1} and t''_{MA_2}) and the t_{BO} solution; the results are shown in Table 4.11. A visual inspection of Figure 2.4 might suggest that the intake of Rilatine results in better concentration scores as compared to placebo, but the permutation test shows that there is only weak evidence against the null hypothesis. The three test statistics provide p-values around 10%, with t''_{MA_1} providing the smallest p-value and t''_{MA_2} providing the largest p-value. On the basis of his results the adolescent decided not to continue the medication. Furthermore, in a debriefing interview after the study, it was clear that he could not discriminate between the treatment and the placebo periods.

2.7 Main results

In this chapter we presented a permutation solution to test for treatment effects in single-case experiments with an alternation design. The aim of our research

is to improve the already existing solutions and provide an instrument which behaves well in a variety of data scenarios.

The joint use of nonparametric smoothing and permutation theory provides a good solution for the problem of interest; additionally, considering an *MA* solution, we are able to satisfactorily handle several data scenarios. A simulation study showed that the proposed tool is reliable under the null hypothesis and powerful under the alternative. As regards comparison with an already existing solution, the *MA* solutions generally behave better than their competitor, always showing an estimated rejection probability greater than the nominal level of the test under the alternative hypothesis; this is not the case, instead, for the considered competitor. In general a good feature of the proposed solutions is the increase in their power as the data move further away from the null hypothesis.

The new techniques were also applied to real data, where we confirmed the results of the simulation study and demonstrated the flexibility and usefulness of the methods.

Several further developments can be made using the tools presented here. For instance, it may be of interest to study performances when a smoother other than local regression is used to draw the smoothed time sub-processes, as well as when we simply change the choice of smoothing parameter. The role played by the use of different partial test statistics in the MA solution could also be studied. Furthermore, the performance of the methods could be explored under many other data scenarios, for instance letting the several treatments have also different underlying correlation structures besides the differences in levels or slopes. Another possible kind of presence of carry-over is when there is the effect of the number of treatment administrations that precede a particular administration (see Edgington and Onghena (2007)). Hence two scenarios can happen: the carry-over is identical for the several treatments (identical carry-over), or it is different, meaning that the subject's response to a treatment at certain time depends also on the particular treatments that were given at the previous treatment times (differential carry-over). In the former case we do not have any difference among treatments' effects, hence we are under the null hypothesis, and inference made through randomization tests in alternation designs is not affected by such a carry-over effect because the effect is constant over all data permutations. In the latter case instead, we are under the alternative hypothesis, as there is a difference in the effects of the several treatments; in this case, inference made through randomization tests is affected by this kind of effect, as this is not constant over all data permutations. It would surely be interesting to check the behavior of the proposed procedure in such case.

Table 2.7: Estimated power for the alternation design solutions (using direct combining function on standardized test statistics), $C = 3$ treatments. Simulated difference in level among the three treatments' effects; errors simulated from the standard normal distribution.

			t''_{MA_1}, $l_{grid} = 100$ Length of the series			t''_{MA_2}, $l_{grid} = 100$ Length of the series		
Setting	*Model*	*Parameters*	$n = 45$	$n = 75$	$n = 150$	$n = 45$	$n = 75$	$n = 150$
Setting 1	$AR(1)$	$\phi = -0.3$	0.078	0.097	0.172	0.115	0.173	0.294
		$\phi = -0.1$	0.078	0.127	0.199	0.107	0.168	0.321
		$\phi = 0.1$	0.085	0.128	0.204	0.128	0.181	0.328
		$\phi = 0.3$	0.084	0.117	0.187	0.112	0.190	0.290
		$\phi = 0.5$	0.089	0.103	0.188	0.115	0.171	0.281
	$MA(1)$	$\theta = -0.3$	0.066	0.104	0.175	0.112	0.177	0.301
		$\theta = -0.1$	0.094	0.103	0.206	0.143	0.172	0.301
		$\theta = 0.1$	0.080	0.131	0.201	0.123	0.209	0.324
		$\theta = 0.3$	0.078	0.114	0.184	0.112	0.166	0.313
		$\theta = 0.5$	0.080	0.108	0.184	0.117	0.158	0.282
	$ARMA$ $(1,1)$	$\phi, \theta = -0.3$	0.073	0.102	0.146	0.116	0.159	0.255
		$\phi, \theta = -0.1$	0.084	0.130	0.206	0.114	0.187	0.309
		$\phi, \theta = 0.1$	0.091	0.106	0.207	0.128	0.174	0.314
		$\phi, \theta = 0.3$	0.069	0.089	0.160	0.099	0.134	0.258
		$\phi, \theta = 0.5$	0.080	0.068	0.114	0.095	0.105	0.173
Setting 2	$AR(1)$	$\phi = -0.3$	0.140	0.221	0.495	0.271	0.458	0.816
		$\phi = -0.1$	0.153	0.261	0.531	0.323	0.534	0.848
		$\phi = 0.1$	0.140	0.253	0.552	0.299	0.534	0.849
		$\phi = 0.3$	0.147	0.233	0.492	0.272	0.484	0.817
		$\phi = 0.5$	0.129	0.220	0.423	0.220	0.385	0.711
	$MA(1)$	$\theta = -0.3$	0.117	0.230	0.501	0.273	0.483	0.823
		$\theta = -0.1$	0.131	0.251	0.514	0.289	0.524	0.823
		$\theta = 0.1$	0.148	0.261	0.526	0.304	0.529	0.848
		$\theta = 0.3$	0.161	0.232	0.528	0.309	0.503	0.823
		$\theta = 0.5$	0.141	0.252	0.510	0.289	0.464	0.788
	$ARMA$ $(1,1)$	$\phi, \theta = -0.3$	0.116	0.202	0.417	0.215	0.389	0.734
		$\phi, \theta = -0.1$	0.140	0.256	0.511	0.316	0.525	0.860
		$\phi, \theta = 0.1$	0.141	0.254	0.511	0.310	0.501	0.836
		$\phi, \theta = 0.3$	0.146	0.215	0.413	0.238	0.386	0.726
		$\phi, \theta = 0.5$	0.097	0.155	0.257	0.188	0.236	0.460
Setting 3	$AR(1)$	$\phi = -0.3$	0.148	0.208	0.353	0.272	0.364	0.407
		$\phi = -0.1$	0.159	0.250	0.346	0.296	0.389	0.411
		$\phi = 0.1$	0.156	0.233	0.333	0.277	0.355	0.389
		$\phi = 0.3$	0.147	0.208	0.371	0.269	0.321	0.408
		$\phi = 0.5$	0.144	0.204	0.346	0.244	0.339	0.418
	$MA(1)$	$\theta = -0.3$	0.150	0.239	0.351	0.287	0.378	0.396
		$\theta = -0.1$	0.148	0.233	0.363	0.274	0.371	0.411
		$\theta = 0.1$	0.155	0.260	0.353	0.289	0.409	0.414
		$\theta = 0.3$	0.146	0.249	0.335	0.253	0.371	0.393
		$\theta = 0.5$	0.140	0.235	0.398	0.282	0.337	0.453
	$ARMA$ $(1,1)$	$\phi, \theta = -0.3$	0.129	0.202	0.356	0.233	0.349	0.442
		$\phi, \theta = -0.1$	0.162	0.212	0.371	0.316	0.330	0.422
		$\phi, \theta = 0.1$	0.159	0.248	0.346	0.294	0.367	0.408
		$\phi, \theta = 0.3$	0.135	0.221	0.350	0.233	0.346	0.417
		$\phi, \theta = 0.5$	0.134	0.176	0.296	0.199	0.263	0.418
Setting 4	$AR(1)$	$\phi = -0.3$	0.623	0.929	1.000	0.997	1.000	1.000
		$\phi = -0.1$	0.641	0.933	1.000	0.999	1.000	1.000
		$\phi = 0.1$	0.645	0.941	1.000	0.998	1.000	1.000
		$\phi = 0.3$	0.606	0.918	1.000	0.994	1.000	1.000
		$\phi = 0.5$	0.496	0.836	0.994	0.954	0.993	1.000
	$MA(1)$	$\theta = -0.3$	0.325	0.655	0.949	0.789	0.969	0.999
		$\theta = -0.1$	0.347	0.641	0.964	0.852	0.977	1.000
		$\theta = 0.1$	0.377	0.672	0.972	0.860	0.970	1.000
		$\theta = 0.3$	0.377	0.669	0.965	0.810	0.967	0.999
		$\theta = 0.5$	0.284	0.615	0.948	0.754	0.942	1.000
	$ARMA$ $(1,1)$	$\phi, \theta = -0.3$	0.584	0.882	0.999	0.992	0.999	1.000
		$\phi, \theta = -0.1$	0.647	0.930	0.999	0.998	1.000	1.000
		$\phi, \theta = 0.1$	0.655	0.942	1.000	0.999	1.000	1.000
		$\phi, \theta = 0.3$	0.575	0.913	1.000	0.985	1.000	1.000
		$\phi, \theta = 0.5$	0.433	0.768	0.984	0.889	0.998	1.000

Table 2.8: Estimated power for the alternation design solutions (using direct combining function on standardized test statistics), $C = 5$ treatments. Simulated difference in level among the three treatments' effects; errors simulated from the standard normal distribution.

			t''_{MA_1}, $l_{grid} = 100$ Length of the series			t''_{MA_2}, $l_{grid} = 100$ Length of the series		
Setting	*Model*	*Parameters*	$n = 45$	$n = 75$	$n = 150$	$n = 45$	$n = 75$	$n = 150$
Setting 1	$AR(1)$	$\phi = -0.3$	0.114	0.157	0.240	0.194	0.241	0.269
		$\phi = -0.1$	0.115	0.157	0.213	0.213	0.242	0.277
		$\phi = 0.1$	0.106	0.163	0.226	0.197	0.248	0.260
		$\phi = 0.3$	0.119	0.160	0.229	0.199	0.241	0.271
		$\phi = 0.5$	0.116	0.192	0.254	0.193	0.271	0.316
	$MA(1)$	$\theta = -0.3$	0.115	0.170	0.238	0.202	0.273	0.251
		$\theta = -0.1$	0.100	0.185	0.250	0.200	0.239	0.283
		$\theta = 0.1$	0.113	0.165	0.244	0.193	0.260	0.269
		$\theta = 0.3$	0.121	0.165	0.222	0.203	0.234	0.254
		$\theta = 0.5$	0.121	0.161	0.235	0.197	0.244	0.275
	$ARMA$ $(1,1)$	$\phi, \theta = -0.3$	0.110	0.157	0.216	0.201	0.268	0.295
		$\phi, \theta = -0.1$	0.128	0.155	0.240	0.196	0.235	0.276
		$\phi, \theta = 0.1$	0.111	0.185	0.238	0.213	0.252	0.264
		$\phi, \theta = 0.3$	0.115	0.144	0.256	0.188	0.240	0.291
		$\phi, \theta = 0.5$	0.081	0.126	0.221	0.134	0.196	0.292
Setting 2	$AR(1)$	$\phi = -0.3$	0.208	0.360	0.605	0.539	0.662	0.709
		$\phi = -0.1$	0.160	0.345	0.614	0.525	0.664	0.694
		$\phi = 0.1$	0.205	0.340	0.598	0.520	0.625	0.692
		$\phi = 0.3$	0.190	0.357	0.608	0.520	0.647	0.701
		$\phi = 0.5$	0.188	0.356	0.588	0.486	0.648	0.748
	$MA(1)$	$\theta = -0.3$	0.188	0.343	0.612	0.534	0.653	0.736
		$\theta = -0.1$	0.182	0.352	0.603	0.507	0.629	0.695
		$\theta = 0.1$	0.183	0.357	0.622	0.503	0.636	0.691
		$\theta = 0.3$	0.197	0.339	0.619	0.537	0.624	0.718
		$\theta = 0.5$	0.185	0.374	0.623	0.485	0.637	0.728
	$ARMA$ $(1,1)$	$\phi, \theta = -0.3$	0.192	0.350	0.595	0.491	0.676	0.748
		$\phi, \theta = -0.1$	0.199	0.365	0.586	0.552	0.647	0.692
		$\phi, \theta = 0.1$	0.182	0.357	0.585	0.500	0.637	0.669
		$\phi, \theta = 0.3$	0.207	0.369	0.614	0.508	0.613	0.754
		$\phi, \theta = 0.5$	0.166	0.310	0.565	0.389	0.590	0.757
Setting 3	$AR(1)$	$\phi = -0.3$	0.083	0.084	0.092	0.123	0.096	0.101
		$\phi = -0.1$	0.050	0.080	0.080	0.078	0.084	0.074
		$\phi = 0.1$	0.052	0.068	0.063	0.084	0.079	0.086
		$\phi = 0.3$	0.070	0.101	0.094	0.094	0.108	0.100
		$\phi = 0.5$	0.072	0.113	0.117	0.153	0.138	0.129
	$MA(1)$	$\theta = -0.3$	0.079	0.105	0.094	0.101	0.114	0.095
		$\theta = -0.1$	0.061	0.059	0.073	0.096	0.077	0.072
		$\theta = 0.1$	0.052	0.074	0.066	0.082	0.086	0.072
		$\theta = 0.3$	0.056	0.090	0.079	0.104	0.104	0.085
		$\theta = 0.5$	0.085	0.092	0.106	0.120	0.117	0.115
	$ARMA$ $(1,1)$	$\phi, \theta = -0.3$	0.116	0.133	0.143	0.166	0.148	0.143
		$\phi, \theta = -0.1$	0.064	0.081	0.082	0.107	0.094	0.081
		$\phi, \theta = 0.1$	0.053	0.081	0.083	0.087	0.092	0.091
		$\phi, \theta = 0.3$	0.100	0.116	0.134	0.168	0.139	0.141
		$\phi, \theta = 0.5$	0.112	0.164	0.210	0.198	0.198	0.225
Setting 4	$AR(1)$	$\phi = -0.3$	0.069	0.198	0.282	0.334	0.376	0.345
		$\phi = -0.1$	0.047	0.106	0.174	0.279	0.267	0.242
		$\phi = 0.1$	0.054	0.131	0.193	0.282	0.260	0.249
		$\phi = 0.3$	0.077	0.179	0.266	0.324	0.355	0.324
		$\phi = 0.5$	0.132	0.288	0.466	0.450	0.523	0.504
	$MA(1)$	$\theta = -0.3$	0.062	0.163	0.227	0.302	0.297	0.278
		$\theta = -0.1$	0.055	0.113	0.183	0.268	0.224	0.230
		$\theta = 0.1$	0.040	0.095	0.199	0.239	0.261	0.232
		$\theta = 0.3$	0.059	0.107	0.193	0.269	0.260	0.245
		$\theta = 0.5$	0.063	0.132	0.239	0.291	0.287	0.285
	$ARMA$ $(1,1)$	$\phi, \theta = -0.3$	0.107	0.240	0.383	0.420	0.448	0.435
		$\phi, \theta = -0.1$	0.053	0.108	0.208	0.274	0.268	0.247
		$\phi, \theta = 0.1$	0.044	0.107	0.210	0.241	0.247	0.258
		$\phi, \theta = 0.3$	0.090	0.197	0.338	0.372	0.429	0.389
		$\phi, \theta = 0.5$	0.169	0.326	0.549	0.502	0.585	0.615

Table 2.9: Settings of treatments' slopes, cases of $C = 3$ and $C = 5$.

C	Setting	β_{tot}	Proportion of false H_0	β_A	β_B	β_C	β_D	β_E
$C = 3$	1	4	2/3	1	1	5		
	2	18	3/3	1	5	10		
	3	18	2/3	1	1	10		
	4	48	3/3	1	10	30		
$C = 5$	5	16	4/10	1	1	1	1	5
	6	44	7/10	1	1	1	5	10
	7	36	4/10	1	1	1	1	10
	8	134	7/10	1	1	1	10	30

Table 2.10: Estimated power for the alternation design solutions (using direct combining function on standardized test statistics), $C = 3$ treatments. Simulated difference in slope among the three treatments' effects; errors simulated from the standard normal distribution.

	t''_{MA_1}, $l_{grid} = 100$ Length of the series			t''_{MA_2}, $l_{grid} = 100$ Length of the series		
Setting	$n = 45$	$n = 75$	$n = 150$	$n = 45$	$n = 75$	$n = 150$
1	0.132	0.127	0.159	0.097	0.113	0.140
2	0.122	0.176	0.174	0.101	0.143	0.150
3	0.136	0.158	0.188	0.109	0.128	0.152
4	0.124	0.179	0.211	0.124	0.171	0.178

Table 2.11: Estimated power for the alternation design solutions (using direct combining function on standardized test statistics), $C = 5$ treatments. Simulated difference in slope among the five treatments' effects; errors simulated from the standard normal distribution.

	t''_{MA_1}, $l_{grid} = 100$ Length of the series			t''_{MA_2}, $l_{grid} = 100$ Length of the series		
Setting	$n = 75$	$n = 125$	$n = 250$	$n = 75$	$n = 125$	$n = 250$
5	0.075	0.097	0.113	0.071	0.084	0.099
6	0.108	0.115	0.161	0.079	0.108	.0144
7	0.092	.0126	0.138	0.079	0.120	0.114
8	0.106	0.116	0.153	0.86	0.116	0.129

Table 2.12: P-values of the tests on the difference between placebo and active drug effects; data from Baplu (2005).

	t''_{MA_1}, $l_{grid} = 100$	t''_{MA_2}, $l_{grid} = 100$	t_{BO}
P-values	0.0997	0.1430	0.1105

Chapter 3

Replicated multivariate single-case experiments

3.1 Introduction

As already pointed out in Chapter 2, single-case experiments are often used in clinical research. Statistical inference made on the single subject can assure internal validity, so conclusions on the effect difference of the treatments can be referred to the patient included in the study. In order to extend the single subject related results to a population of interest, more than one subject can be involved in the study, using replicated single-case designs. These are often much more consistent with the way in which consecutive patients are entered into clinical trials than the random sampling model underlying many group designs and standard statistical techniques. This chapter deals with the problem of extending single-case experiment results to a wider population level by combining together the results obtained on replicated single-case experiments.

We recall that by performing a single-case experiment the researcher usually aims to investigate the presence of a difference in the effects of the treatments considered in the study. In this setting valid inference sometimes cannot be made using parametric statistical procedures, and nonparametric tools often represent a valid alternative or complementary approach to analyze this kind of data (see Pesarin and Salmaso (2010)). In Chapter 2 we underlined that permutation methods, in particular, are an interesting solution to the problem, since they exploit the randomization of assignment of the measurement occasions to treatments in order to build the permutation distribution of the test statistic under the null hypothesis. Statistical inference is then based on the comparison of the test statistic from the observed data with this permutation distribution. In this chapter we consider and extend the permutation solution discussed in Chapter 2 as well; the principal idea of this univariate test is to smooth the single-case experiment time-series and then study the difference between the several treatments' effect working with the resulting smoothed processes, instead of with other classical statistics based on the original dataset. The method works well under a variety of data scenarios, is reliable under the null hypothesis and is powerful under the alternative. Through this technique we can assure the study's internal and statistical-conclusion validity.

In this chapter we will extend the above cited permutation solution in order to deal with the multivariate response case and with replicated single-case experiments, by using the nonparametric combination of dependent permutation tests (see Pesarin (2001) and Pesarin and Salmaso (2010)). In the second section we recall the single-case experiment theory and the permutation solution presented in Chapter 2; Section 3 is devoted to the presentation of our proposed extension. Section 4 shows the results of a simulation study in which the proposed permutation solutions are tested under a wide variety of data scenarios, and these methods are applied to a real experiment and mixed real and simulated data in Section 5. At the end, conclusions about the obtained results are discussed in Section 6.

3.2 A time-series permutation approach for single-case experiments

In single-case experiments we observe one subject over a period of time, administering different treatments during the study. The allocation of observation occasions to treatment conditions is done using randomization. Depending on which type of randomization we consider, we deal with different randomization designs, and in this thesis we present some solutions for alternation design, case in which any level of the independent variable can be present at each measurement occasion.

To extend the single-case results to a wider population of interest we can use replication; this is done by planning separated single-case experiments on several subjects, and then performing them simultaneously (simultaneous replication design) or sequentially (sequential replication design). We can introduce randomization into the study by simply applying the randomization schedules separately in the several single-case designs (see Onghena and Edgington (2005)). The complexity of the data we deal with in this case appears quite clear: considering the simpler case in which a univariate response in recorded in each single-case experiment, we handle several time-series, one for each subject, on which we wish to compare two or more treatments. When the response of interest is multivariate, the problem complexity increases further on as for each subject we register several time-series on which we want to test for the treatments' effect difference.

The permutation solution presented in Chapter 2 proposes the joint use of smoothing methods and the permutation theory in order to test for treatments' effect difference in univariate single-case experiments. Below we will adopt the same notation used in Chapter 2. We recall that the null and alternative hypotheses can be formalized as follows:

$$\begin{cases} H_0: & X_{A^j}(t) \overset{\mathcal{M}}{=} X_{A^l}(t) \qquad \forall j < l, \quad j,l = 1,\dots,C, \quad \forall t \in D \\ H_1: & \exists\, j,l \in \{1,\dots,C\}, \quad j \neq l \quad | \quad X_{A^j}(t) \overset{\mathcal{M}}{\neq} X_{A^l}(t), \end{cases},$$

with (t) and $\mathcal{M}$ underlining the attention on the model underlying the response time-process. The idea for the solution is to estimate the time functional shape of observations for the different treatments' sub-datasets, using a nonparametric smoother, and then to use a summary statistic which is able to measure the

difference among the several smoothed time-processes. As regards the choice of smoother, we propose the use of local regression and, for the choice of test statistic, a *MA* solution.

3.3 Extension via nonparametric combination of dependent permutation tests

As we said above, the multiplicity issue can arise in single-case experiments in two cases: at first, as very often happens in real data applications, when the dependent variable of the study is multivariate; then, when we are interested in asserting external validity of the analysis results, and we consider replications of single-case experiments. In actual fact both these problems are hard to solve using classical methods. This is essentially due to the fact that each single-case experiment with a univariate outcome is a time-series, and when we either consider a multivariate response or perform a replicated single-case experiment, it is like considering separated time-series. Of course a further problem is that these separated time-series actually are (in single-case experiments with a multivariate response) or may be (in replicated single-case experiments) correlated. In order to solve this problem, some solutions are proposed in Bulté and Onghena (2009), essentially related to the combination of the test statistics. We propose to use nonparametric combination of dependent permutation tests (see Pesarin (2001)); this way we can handle the multiplicity issue completely disregarding the dependence structure among the different tests.

In practice multiplicity can arise in both ways at the same time in single-case experiments. Replicated single-case experiments can indeed be performed when multivariate responses are observed for each subject. In this framework we propose to nonparametrically combine partial permutation tests, while seeking consistency in the design construction and the recording of data. As the name suggests, replicated single-case experiments can be seen as a replication of individual single-case experiments; each single-case experiment can then be performed recording a multivariate response. Hence it is natural to construct a combination procedure that works on two nested levels: firstly, at an internal level, we handle the multiplicity related to the multivariate nature of the response; secondly, at an external level, we cope with the multiplicity from replication of the individual single-case experiments. Suppose we perform a replicated single-case experiment on S subjects, for each of which we record values from a multivariate response $X_s = (X_s^1, \ldots, X_s^{p_s})$, for $s = 1, \ldots, S$. Hence for the lth component of the multivariate response and the sth subject we observe the univariate time-series $\boldsymbol{x}_s^l = (x_{s,1}^l, \ldots, x_{s,T}^l)$, with $s = 1, \ldots, S$ and $l = 1, \ldots, p_s$. Thus the index $l = 1, \ldots, P$ denotes the component of the multivariate response (and not the treatment condition as in the previous chapter), $s = 1, \ldots, S$ indicates the subjects, $\boldsymbol{x}_s^l$ for $l = 1, \ldots, p_s$ is the response of the lth variable for subject s. Formally the procedure we propose works as follows:

I) for all the individual single-case experiments, perform a partial permutation test, say T_s^l, on each component of the multivariate response $\boldsymbol{x}_s^l$, for $l = 1, \ldots, p_s$ and $s = 1, \ldots, S$, following the iterative procedure below:

   i. consider the original time-series of subject s and component of the multivariate response l, say $\boldsymbol{x} = (x_1, \ldots, x_T)$, and the C sub-series

$\boldsymbol{x}_{A^j} = (x_{A^j,1_{A^j}}, \ldots, x_{A^j,T_{A^j}})$. Compute the C smoothed processes $s_{A^j}(t; \boldsymbol{x}_{A^j})$, for $j = 1, \ldots, C$;

ii. compute the observed value for the partial test statistics $t'^{,obs}_m(\boldsymbol{x}) = \psi_m(\boldsymbol{x}) = \psi_m(s_{A^1}(t; \boldsymbol{x}_{A^1}), \ldots, s_{A^C}(t; \boldsymbol{x}_{A^C}))$, for $m = 1, \ldots, M$;

iii. according to the randomization scheme, perform a randomization of $\boldsymbol{x}$, obtaining the randomized time-series $\boldsymbol{x}^{*,1}$ and the C permuted sub-series $\boldsymbol{x}^{*,1}_{A^j}$. Compute the C permuted smoothed processes $s^{*,1}_{A^j}(t; \boldsymbol{x}^{*,1}_{A^j})$, for $j = 1, \ldots, C$;

iv. compute the value of the partial test statistics $t'^{,*,1}_m = \psi_m(\boldsymbol{x}^{*,1}) = \psi_m(s^{*,1}_{A^1}(t; \boldsymbol{x}^{*,1}_{A^1}), \ldots, s^{*,1}_{A^C}(t; \boldsymbol{x}^{*,1}_{A^C}))$, for $m = 1, \ldots, M$;

v. repeat steps iii and iv B times, obtaining the values of the partial test statistics $t'^{,*,i}_m = \psi_m(\boldsymbol{x}^{*,i}) = \psi_m(s^{*,i}_{A^1}(t; \boldsymbol{x}^{*,i}_{A^1}), \ldots, s^{*,i}_{A^C}(t; \boldsymbol{x}^{*,i}_{A^C}))$, for $i = 1, \ldots, B$, $m = 1, \ldots, M$;

vi. construct the permutation distributions of the partial test statistics under the null hypothesis from the vectors of values $t'^{,*}_m = (t'^{,*,1}_m, \ldots, t'^{,*,B}_m)$, for $m = 1, \ldots, M$;

vii. nonparametrically combine the partial permutation tests, obtaining the permutation distribution of the Multi-Aspect global test $T^l_s = t''_{MA} = \Psi(t'_1, \ldots, t'_M)$;

II) nonparametrically combine the partial permutation tests within each subject, obtaining a global subject related test $T'^{,s} = \Phi\left(T^1_s, \ldots, T^{p_s}_s\right)$, for $s = 1, \ldots, S$;

III) nonparametrically combine the global subject related permutation tests, obtaining a global test $T'' = \Phi\left(T'^{,1}, \ldots, T'^{,S}\right)$.

At the single subject - single component stage of the multivariate response level, we propose to perform the two *MA* permutation solutions presented in Chapter 2. We remark that, in order to deal with the implicit correlation structure present within each subject, synchronized randomizations of the treatments are performed at Step `I.iii.` of the above presented algorithm (for all the components of the multivariate response within each subject).

Furthermore, using a closed testing procedure (see Marcus et al. (1976)), we can also recover the information on the significance of partial and global subject related tests. Adjusted p-values can indeed be calculated to see which components of the multivariate response or which individual single-case experiments led to the eventual rejection of the null hypothesis. The above described combination scheme is presented schematically in Figure 3.1.

3.4 Power behavior

A simulation study was carried out to check the performance of the proposed solution, following a very similar scheme to the one used in the simulation study performed in Chapter 2. Our aim is to demonstrate the reliability and power of the proposed procedures in replicated single-case experiments and in the case

of multivariate response. Due to computational time restrictions, we considered the case of $C = 2$ treatments to be compared. The tested hypotheses are

$$\begin{cases} \bigcap_{s=1}^{S} H_0^s \\ \bigcup_{s=1}^{S} H_1^s \end{cases} \Leftrightarrow \begin{cases} \bigcap_{s=1}^{S} \left\{ X_{s,A^1}(t) \overset{\mathcal{M}}{=} X_{s,A^2}(t) \right\} & \forall t \in T \\ \bigcup_{s=1}^{S} \left\{ X_{s,A^1}(t) \overset{\mathcal{M}}{\neq} X_{s,A^2}(t) \right\}, & \end{cases},$$

where $X_{s,A^c}(t)$ indicates the p_c-dimensional multivariate response time-series related to the sth subject under treatment A^c, with $c = 1, 2$ and $s = 1, \ldots, S$. Notice that the computation order complexity is quite high, since at first a partial permutation test (and the related B permutations) is performed for each component of the multivariate response within each subject, then the procedure is repeated for all the subjects. Hence, the B permutations are performed in a threefold nested "*for*" loop. The reliability of the methods was measured through the estimation of the type I error, while the capability of the procedures to detect a difference among treatments' effect was deduced from the estimation of power. In the latter case different data scenarios were simulated under the alternative hypothesis. We considered $MC = 1000$ Monte Carlo replications and $B = 1000$ permutations. Exactly as in Chapter 2, with regard to the smoother, we performed a local polynomial regression with degree 2, span parameter $\lambda = 0.75$ and tricubic weighting. As regards the combining function, we propose to use Tippett's at both levels of combination (see scheme in Figure 3.1).

Under the null hypothesis, we considered four lenghts for the entire time series, $n = 30, 50, 70, 100$, and, as the underlying ARMA model, an $ARMA(1, 1)$ with several values for the autoregressive and moving-average parameters, $\phi = \theta = -0.3, -0.1, 0.1, 0.3, 0.5$; hence five $ARMA(1, 1)$ models are considered for each sample size, corresponding to the five possible values for $\phi = \theta$. Under the alternative hypothesis, we generated the difference in treatments' effect adding different deterministic parts to the different treatments' data generating processes for the same stochastic underlying ARMA models (see Chapter 2 for a more detailed description). We considered both a difference in levels and slopes between the treatments' effect. Several settings were considered under the alternative hypothesis. They differ in terms of number of false null partial hypotheses, considering both different numbers of subjects and/or variables for which we are under the alternative hypothesis. We maintained fixed differences in levels and slopes between the treatments' effect (under the alternative hypothesis), always working with a difference in levels of $\delta = 1$ (i.e. we simulated data under H_1 with respective means of the stationary processes equal to 0 for treatment A and to 1 for treatment B) and a difference in slopes of $\beta_{diff} = 49$ (meaning that we simulated data under H_1 with respective slopes $\beta_A = 1$ and $\beta_B = 50$). We considered the case of a replicated single-case experiment with $S = 3$ subjects for which a multivariate response of dimension $p_1 = p_2 = p_3 = 4$ was recorded. The settings are described in Table 3.1 in terms of subjects and variables simulated under the alternative hypothesis (for both cases of differences in levels and slopes). As regards the correlation structure of the errors, we considered the case of no correlation among the subjects and two levels of correlation between the pairs of variables within each single-case experiment, $\rho_{i,j} = 0.3, 0.8$ for $i < j$, $i, j = 1, \ldots, S$, where $\rho_{i,j}$ denotes the correlation coefficient between the ith and the jth component of the multivariate response. Hence, since normal errors have been used to simulate the data, we simulated

Table 3.1: Settings of simulated data under the alternative hypothesis - the symbol ✓ indicates the partial comparisons which have been simulated under the alternative hypothesis. X_s^i, $i = 1, \ldots, 4$, $s = 1, \ldots, 3$ indicates the ith component of the multivariate response for the sth subject.

Setting	*Var. under H_1 (per subj.)*	*Subj. under H_1*	Subj. 1				Subj. 2				Subj. 3			
			X_1^1	X_1^2	X_1^3	X_1^4	X_2^1	X_2^2	X_2^3	X_2^4	X_3^1	X_3^2	X_3^3	X_3^4
1	2	1	✓	✓										
2	2	2	✓	✓			✓	✓						
3	2	3	✓	✓			✓	✓			✓	✓		
4	4	2	✓	✓	✓	✓	✓	✓	✓	✓				

errors for each single-case experiment, at each observation occasion, from a multivariate normal distribution with a correlation matrix with values $\rho_{i,j}$, for $i < j$, $i, j = 1, \ldots, S$, for the elements out of the diagonal. Hence the considered settings in terms of $ARMA$ underlying models, sample size and number of false partial null hypotheses have been explored twice, considering the two correlation structures.

The results related to the behavior under the null and alternative hypotheses are reported in Table 3.2 and 3.3 for the two correlation structures respectively. The proposed methods showed to control the type I error at the nominal level in all the considered settings. Some of the simulation results, related to power of the proposed methods, when $\rho_{i,j} = 0.3$, are also displayed in Figure 3.2 for the cases of a simulated difference in levels and slopes respectively between the two treatments. Such results refer to data simulated from an $ARMA(1, 1)$ with autoregressive and moving-average parameters $\phi = \theta = 0.1$, with n and the distance from the global null hypothesis (a combination of the number of aspects and the number of subjects under the alternative hypothesis) varying. Notice that similar results were obtained for the two correlation structures, with a slight decrease of power for both procedures for $\rho_{i,j} = 0.8$, due to the larger correlation between the partial permutation tests. As we can see the power of both procedures reaches 1 already with a difference in levels of $\delta = 1$ and a low sample size ($n = 30$). In these cases the univariate version of the two MA solutions did not reach power values greater than 0.6 (see Chapter 2). This behavior, which can be found in the whole simulation study, suggests the capability of the nonparametric combination to better recognize that we are far from the null hypothesis when we increase the number of aspects and subjects under the alternative hypothesis. Again confirming the results obtained in Chapter 2, as regards the effect of the autocorrelation on the power behavior of the proposed methods, generally, the estimated rejection probabilities slightly decrease as the autocorrelation parameters of the ARMA model move, in absolute value, away from zero. Moreover, for both procedures, we can see that the power increases with the setting, i.e. as the number of aspects and subjects under the alternative hypotheses increases. It also has to be noticed, therefore, that the behavior difference of t''_{MA_2} and t''_{MA_1} that can be appreciated in the univariate version (t''_{MA_2} being more powerful than t''_{MA_1} with difference in levels, no clear difference with difference in slopes, see Chapter 2) is no longer present in the multivariate extension.

Table 3.2: Estimated rejection probabilities for the alternation design solutions, $C = 2$ treatments; for each subject the errors are simulated from the multivariate standard normal distribution with correlation 0.3 between all pairs of variables.

			t''_{MA_1}, $l_{grid} = 100$ Length of the series				t''_{MA_2}, $l_{grid} = 100$ Length of the series			
	Setting	*ARMA Par*	$n = 30$	$n = 50$	$n = 70$	$n = 100$	$n = 30$	$n = 50$	$n = 70$	$n = 100$
H_0		$\phi, \theta = -0.3$	0.014	0.011	0.010	0.012	0.019	0.027	0.023	0.020
		$\phi, \theta = -0.1$	0.010	0.017	0.018	0.013	0.021	0.022	0.027	0.021
		$\phi, \theta = 0.1$	0.013	0.013	0.006	0.015	0.014	0.022	0.018	0.025
		$\phi, \theta = 0.3$	0.013	0.008	0.016	0.016	0.023	0.021	0.030	0.020
		$\phi, \theta = 0.5$	0.014	0.013	0.016	0.014	0.019	0.024	0.026	0.024
H_1 - *Shift*	1	$\phi, \theta = -0.3$	0.993	0.986	0.994	0.991	0.987	0.989	0.986	0.981
		$\phi, \theta = -0.1$	0.986	0.991	0.995	0.988	0.984	0.986	0.992	0.986
		$\phi, \theta = 0.1$	0.992	0.994	0.994	0.987	0.986	0.988	0.985	0.982
		$\phi, \theta = 0.3$	0.987	0.985	0.986	0.988	0.984	0.987	0.980	0.982
		$\phi, \theta = 0.5$	0.988	0.991	0.985	0.984	0.987	0.983	0.988	0.988
H_1 - *Shift*	2	$\phi, \theta = -0.3$	0.989	0.993	0.994	0.993	0.990	0.990	0.990	0.989
		$\phi, \theta = -0.1$	0.987	0.993	0.997	0.993	0.991	0.991	0.993	0.986
		$\phi, \theta = 0.1$	0.991	0.995	0.989	0.995	0.989	0.985	0.992	0.986
		$\phi, \theta = 0.3$	0.988	0.990	0.994	0.996	0.988	0.989	0.991	0.987
		$\phi, \theta = 0.5$	0.993	0.995	0.989	0.994	0.989	0.989	0.990	0.989
H_1 - *Shift*	3	$\phi, \theta = -0.3$	0.994	0.995	0.993	0.998	0.991	0.990	0.994	0.997
		$\phi, \theta = -0.1$	0.992	0.993	0.995	0.993	0.992	0.995	0.993	0.992
		$\phi, \theta = 0.1$	0.996	0.997	0.990	0.996	0.994	0.994	0.995	0.995
		$\phi, \theta = 0.3$	0.994	0.994	0.997	0.994	0.990	0.992	0.995	0.990
		$\phi, \theta = 0.5$	0.992	0.989	0.992	0.996	0.992	0.993	0.988	0.990
H_1 - *Shift*	4	$\phi, \theta = -0.3$	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = -0.1$	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000
		$\phi, \theta = 0.1$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = 0.3$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = 0.5$	0.999	1.000	1.000	1.000	0.999	1.000	1.000	1.000
H_1 - *Slope*	1		0.978	0.978	0.979	0.981	0.985	0.988	0.983	0.990
	2		0.989	0.986	0.990	0.980	0.992	0.994	0.990	0.992
	3		0.996	0.992	0.990	0.996	0.996	1.000	0.994	0.996
	4		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

3.5 Two applications

3.5.1 The Baplu (2005) experiment

We applied the proposed MA solutions to mixed real and simulated data: the real part comes from a single-case experiment carried out at the Virga Jesse Hospital in Hasselt, Belgium (see Baplu (2005)). The experiment is the same considered in Chapter 2. The data refer to a 17-year old adolescent who had concentration problems. The treatment which was tried on the patient was methylphenidate (Ritalin). In this chapter we consider the single outcome measures, recorded twice a day, and not the summed scores. Hence we deal with scores on a scale from 0 (never or rarely) to 3 (very often) for each of the following items: a) does not follow through on instructions and fails to finish work, b) fails to give close attention to details or makes careless mistakes in schoolwork, c) has difficulty sustaining attention in tasks or play activities.

As we wish to show how the proposed permutation solution works in the case of replicated multivariate single-case experiments, we simulated the outcomes of two other patients, following the same setup that can be found in Baplu (2005). We simulated the two other multivariate single-case experiments under the alternative hypothesis on all the items, with several levels for the difference in the effects of the two treatments. The data for the real and simulated subjects

Table 3.3: Estimated rejection probabilities for the alternation design solutions, $C = 2$ treatments; for each subject the errors are simulated from the multivariate standard normal distribution with correlation 0.8 between all pairs of variables.

			t''_{MA_1}, $l_{grid} = 100$ Length of the series				t''_{MA_2}, $l_{grid} = 100$ Length of the series			
	Setting	*ARMA Par*	$n = 30$	$n = 50$	$n = 70$	$n = 100$	$n = 30$	$n = 50$	$n = 70$	$n = 100$
H_0		$\phi, \theta = -0.3$	0.025	0.026	0.028	0.033	0.036	0.026	0.025	0.044
		$\phi, \theta = -0.1$	0.031	0.022	0.037	0.030	0.043	0.031	0.046	0.034
		$\phi, \theta = 0.1$	0.038	0.036	0.029	0.033	0.049	0.044	0.056	0.049
		$\phi, \theta = 0.3$	0.028	0.031	0.031	0.029	0.029	0.046	0.040	0.049
		$\phi, \theta = 0.5$	0.024	0.033	0.038	0.037	0.037	0.050	0.035	0.047
H_1 - *Shift*	1	$\phi, \theta = -0.3$	0.982	0.983	0.983	0.980	0.973	0.972	0.976	0.976
		$\phi, \theta = -0.1$	0.980	0.978	0.980	0.991	0.973	0.976	0.983	0.974
		$\phi, \theta = 0.1$	0.970	0.981	0.988	0.987	0.973	0.984	0.980	0.977
		$\phi, \theta = 0.3$	0.978	0.982	0.981	0.985	0.975	0.984	0.970	0.980
		$\phi, \theta = 0.5$	0.980	0.977	0.975	0.987	0.970	0.969	0.974	0.977
H_1 - *Shift*	2	$\phi, \theta = -0.3$	0.994	0.995	0.989	0.994	0.987	0.990	0.987	0.994
		$\phi, \theta = -0.1$	0.989	0.987	0.986	0.993	0.982	0.989	0.984	0.989
		$\phi, \theta = 0.1$	0.984	0.989	0.986	0.989	0.984	0.980	0.980	0.990
		$\phi, \theta = 0.3$	0.990	0.990	0.988	0.991	0.990	0.990	0.991	0.989
		$\phi, \theta = 0.5$	0.982	0.980	0.983	0.982	0.982	0.983	0.983	0.988
H_1 - *Shift*	3	$\phi, \theta = -0.3$	0.995	0.996	0.995	0.995	0.987	0.990	0.993	0.987
		$\phi, \theta = -0.1$	0.993	0.995	0.993	0.992	0.993	0.994	0.990	0.996
		$\phi, \theta = 0.1$	0.992	0.988	0.994	0.993	0.988	0.991	0.992	0.991
		$\phi, \theta = 0.3$	0.986	0.993	0.989	0.998	0.990	0.988	0.988	0.995
		$\phi, \theta = 0.5$	0.985	0.993	0.991	0.995	0.992	0.991	0.987	0.996
H_1 - *Shift*	4	$\phi, \theta = -0.3$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = -0.1$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = 0.1$	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = 0.3$	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\phi, \theta = 0.5$	0.995	0.999	1.000	1.000	0.998	0.998	1.000	1.000
H_1 - *Slope*	1		0.980	0.979	0.979	0.982	0.985	0.991	0.984	0.989
	2		0.985	0.988	0.991	0.985	0.991	0.994	0.993	0.991
	3		0.997	0.994	0.990	0.996	0.994	1.000	0.994	0.996
	4		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

are displayed in Figure 3.3. We remark that, as in the real experiment no score greater than 2 was recorded, we did not allow that value for the simulated experiments either.

Table 3.4 reports the results of the application of the two MA solutions to the data. It reports the adjusted partial p-values to test for effect difference on each dimension of the multivariate response in each single-case experiment, the global p-values for the several single-case experiments, and the final p-value for the global test. A visual inspection of Figure 3.3 might suggest that the effect of Rilatine is better in terms of concentration scores as compared to placebo, in particular in the graphics related to the second simulated experiment. The obtained results, reported in Table 3.4, suggest that both MA solutions are able to recognize a global difference in the effects of the two treatments. Notice how the permutation solutions and the use of nonparametric combination are able to extract more from the extra information due to the replication of the single-case experiment on more patients. Furthermore, it can also be seen that t''_{MA_2} seems to be better able to recognize the effect differences. This result is coherent with the considerations made after the simulation study: difference in levels, which might be conjectured looking at the graphics, is the case where in the simulation study t''_{MA_2} was shown to lead to higher powers.

Table 3.4: Application to mixed real (from Baplu (2005)) and simulated data; p-values of the tests on the effect difference.

	t''_{MA_1}, $l_{grid} = 100$			t''_{MA_2}, $l_{grid} = 100$		
Subject	Partial	Sub-global	Global	Partial	Sub-global	Global
	0.248			0.267		
1	0.497	0.247		0.549	0.267	
	0.248			0.378		
	0.872			0.584		
2	0.749	0.301	*0.0119*	0.111	0.010	*0.0011*
	0.302			0.011		
	0.364			0.328		
3	0.364	0.006		0.215	0.000	
	0.006			0.001		

3.5.2 The Yelland et al. (2009) experiment

We also applied the proposed permutation tests to a real study: an experiment based on a replicated single-case experiment to assess the efficacy of a certain drug for chronic neuropathic pain, presented in Yelland et al. (2009). Neurophatic pain is a condition characterized by lesions or diseases affecting some areas of the nervous system. The function of such nerves is affected in a way that it sends pain messages to the brain. Neuropathic pain is often described as burning, stabbing or like an electric shock. This kind of disease can be very difficult to treat; in particular determining the best treatment for individual patients is a very challenging issue nowadays, as individual responses might be difficult to predict (see Yelland et al. (2009) for a more in depth discussion). The drug administered in the study is an antidepressant which is often used to treat neuropathic pain. The study we consider here refers to a replicated single case experiment in which a multivariate response is recorded in order to compare the active drug and placebo's effects on curing this disease. Patients came from two Australian public hospitals in Port Kembla and in Brisbane (Australia). After an open-label dose-finding period of 2-3 weeks, only for the patients who were not already on the active drug, the 12 week lasting experiment started: the patients were observed for three 4-week periods. During these periods, for two weeks the patients were treated with the active drug and during the other two with placebo. The order of the two treatments was randomized separately in the three periods. Hence, in total 8 randomizations were possible. Some descriptive information was recorded for the patients at their entry in the study, like sex, age and other variables related to the severity of chronic pain. Moreover, several outcome variables were considered related to the pain, the capability of sleeping and a general evaluation of the two treatments.

In this thesis we have at our disposal the data on only eight patients among the ones enrolled in the study and who finished the 12 weeks lasting period, hence less than the original sample size presented in Yelland et al. (2009). We considered a bivariate response consisting in a daily estimation of pain in the marked area and a daily estimation of sleep interference, both on a visual scale from 0 to 10. The daily outcomes were recorded only during the second week of each treatment within each 4-week period in order to allow for washout (the effects of the active drug wear off within two days). Figures 3.4 and 3.5 display

the data: on the two columns we find the bivariate response, while the eight patients are represented on the rows. It can be noticed how the several patients respond quite differently to the both the treatments.

We applied the two proposed permutation solutions to the data. We remind that we constructed the two procedures in order to test for a difference in the effects of the two treatments, hence a two-sided alternative is considered. Table 3.5 summarizes the results of the study on our eight patients: we report the partial p-values for the test on each patient-response combination, along with the global patients' related p-values and the overall p-values, the latter ones to assert if the active drug and placebo have a different effect globally on the bivariate response and on the eight considered patients. We remark that, since only 8 randomizations are possible, the minimum p-value turns out to be 0.125: hence the results must be read in terms of how close are the obtained p-values to such minimum value. The analysis suggests that the active drug and placebo have a different effect on patients 2 and 4. Less strong conclusions in the same direction can be done also for patients 1, 3, 6 and 7 (see Figures 3.4 and 3.5). As regards the choice of the best treatment, we can assert that in patients 2 and 4, the ones with higher evidence of difference between the two treatments, the active drug looks to improve both the pain and sleep scores. The same can be said for patient 7. On the contrary, an opposite situation can be deduced from the displayed data of patient 6 which shows better scores under the placebo treatment. Globally the overall permutation tests suggest a difference in the effect between the drug and placebo. Looking at the other information we had available for the patients, we can underline that patients giving better responses to the active drug (patients 2, 4 and 7, respectively 72, 78 and 96 years old) are in average quite older than patient 6 (45 years old), who showed a better effect for the placebo treatment. Moreover it is also worth noticing that scores of patient 6 are quite high compared to the ones of the other patients, fact that, together with the young age (this patient is also the youngest of the eight considered), could suggest a particular situation for this patient.

We must mention the fact that other combining strategies can be used to analyze this data, which could lead to more powerful solutions for the global permutation test. In particular a trick could be that of looking at the whole replicated single-case experiment as just one big unique study, instead of as the union of several smaller studies on the single patients. By doing this the whole permutation space becomes the space of all possible combination of the 8 possible randomizations per subject, which turns out to have a cardinality of 8^8, much larger than the only 8 possible randomizations that we considered before. We performed such alternative solution, again using Tippett combining function to combine the partial subject related tests. We applied this alternative solution getting a p-value for the global test of 0.058 for t''_{MA_1} and of 0.113 for t''_{MA_2}. These results suggest with higher evidence the presence of a difference in the effects of the two treatments, especially in the case of t''_{MA_1}.

In conclusion, the active drug shows to have a different effect with respect to placebo. The analysis suggests that patient specific decisions should then be taken regarding the administration of this treatment for chronic neuropathic pain. This also confirms the results obtained in Yelland et al. (2009).

Table 3.5: Real application from Yelland et al. (2009); p-values of the tests on the effect difference.

	t''_{MA_1}, $l_{grid} = 100$			t''_{MA_2}, $l_{grid} = 100$		
Subject	Partial	Sub-global	Global	Partial	Sub-global	Global
1	0.250 0.250	0.250	*0.125*	0.250 0.250	0.250	*0.125*
2	0.125 0.625	0.125		0.125 0.875	0.125	
3	0.250 1.000	0.250		0.250 1.000	0.250	
4	0.125 0.125	0.125		0.125 0.125	0.125	
5	0.875 0.875	0.625		0.875 0.875	0.750	
6	0.375 0.250	0.250		0.375 0.250	0.250	
7	0.250 0.250	0.375		0.250 0.250	0.375	
8	0.875 0.875	0.750		0.875 0.875	0.750	

3.6 Main results

In this chapter we presented an extension, via nonparametric combination, of a time-series permutation solution to test for treatments' effect in single-case experiments with alternation designs. The aim of our research was to propose a technique that can be used generally in replicated single-case experiments when a multivariate response is also observed.

The use of nonparametric combination of dependent permutation tests allows us to successfully handle this problem. A simulation study showed that the proposed extension is reliable under the null hypothesis and powerful under the alternative. In general a good feature of the proposed solutions is the increase in their power as the data moves further away from the null hypothesis.

We also applied the new techniques to a real experiment and mixed real and simulated data, in which we show the flexibility and usefulness of the methods which can be used to analyze complicated replicated single-case studies with multivariate responses, even when the length of the single-cases time-series is short.

Several further developments could be made using the tools presented here. For instance, we could explore how to improve the performance of the method changing the combining functions that are used at the different levels of the combination scheme. The performance of the method could also be explored within many other data scenarios.

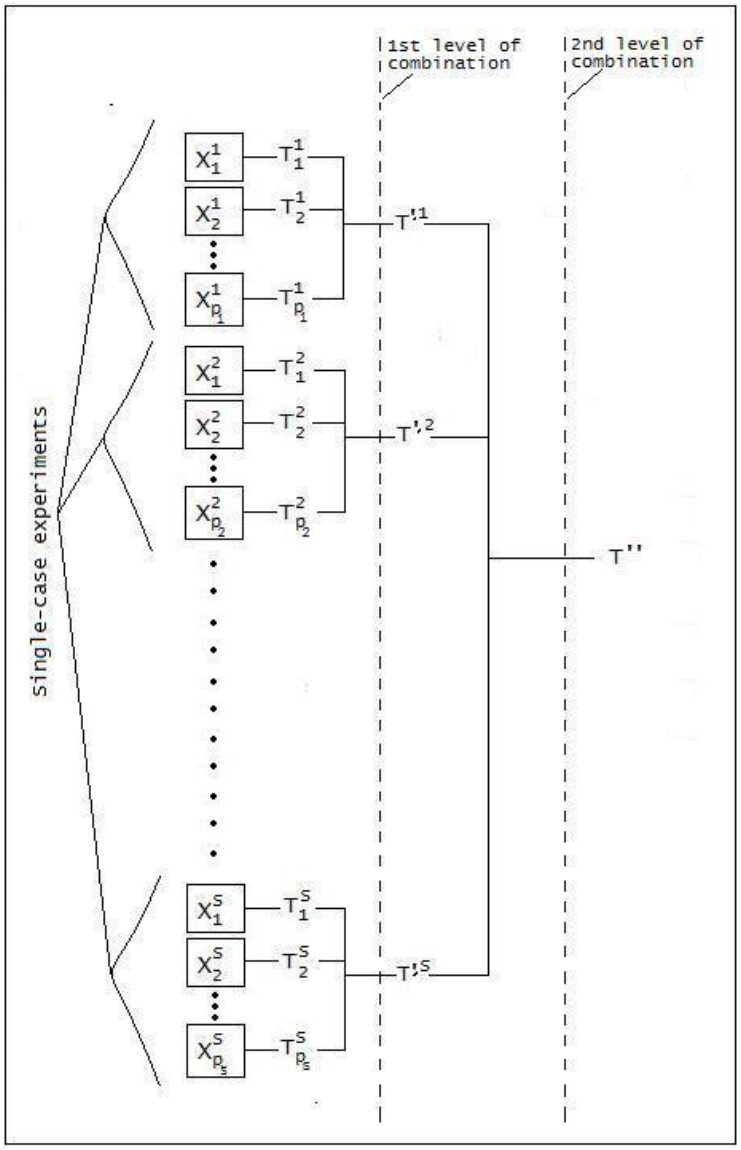

Figure 3.1: Combination scheme for replicated single-case experiments with multivariate response.

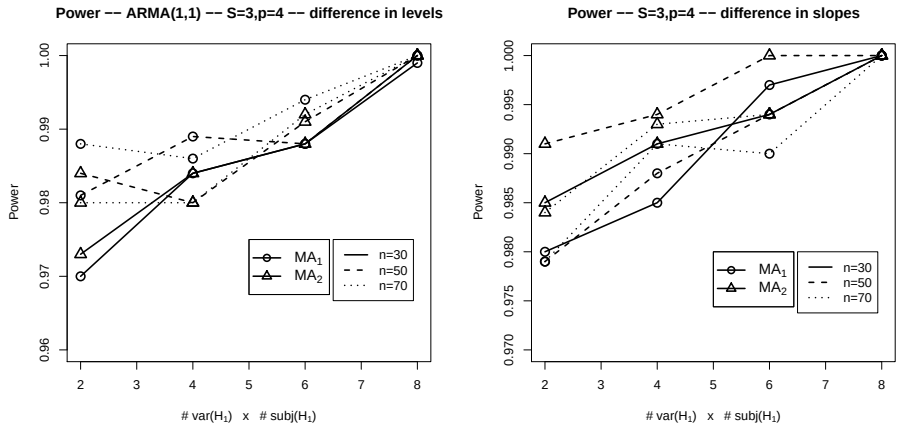

Figure 3.2: Simulation study results. On the left, power functions for the two considered permutation solutions in the case of a simulated difference in levels, an $ARMA(1,1)$ model for the errors with parameters $\phi, \theta = 0.1$, with n and the distance from the global null hypothesis (a combination of number of aspects and number of subjects under the alternative hypothesis) varying; on the right, power functions for the two considered permutation solutions in the case of a simulated difference in slopes, with n and the distance from the global null hypothesis varying.

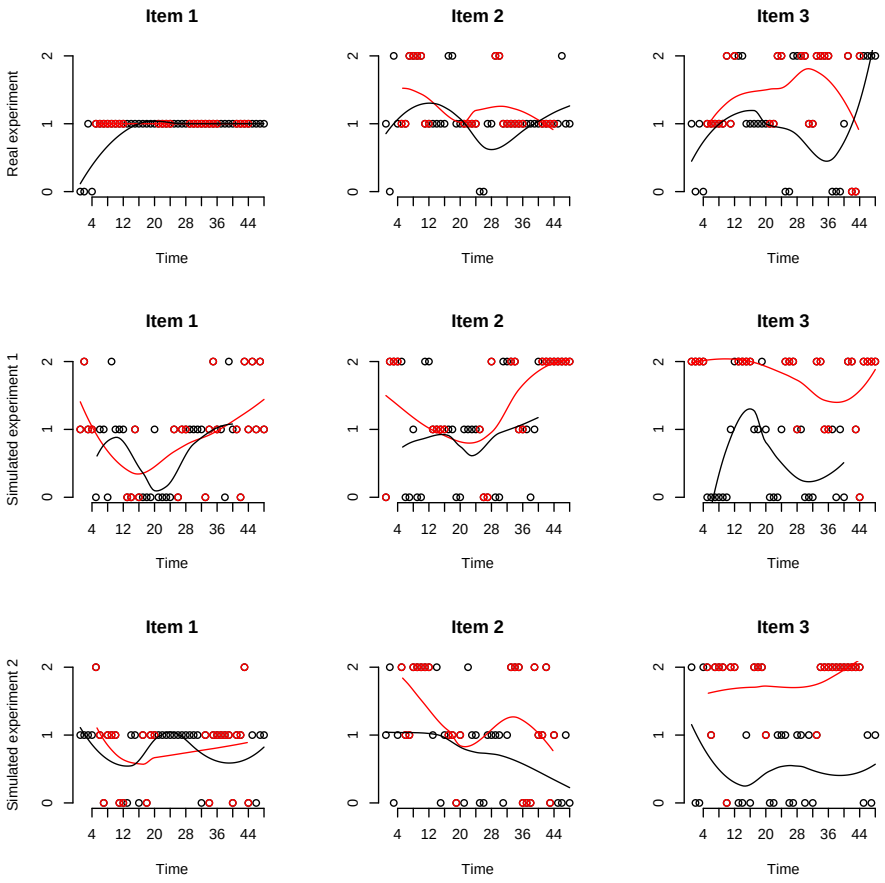

Figure 3.3: real (from Baplu (2005)) and simulated data: scores and smoothed time sub-processes.

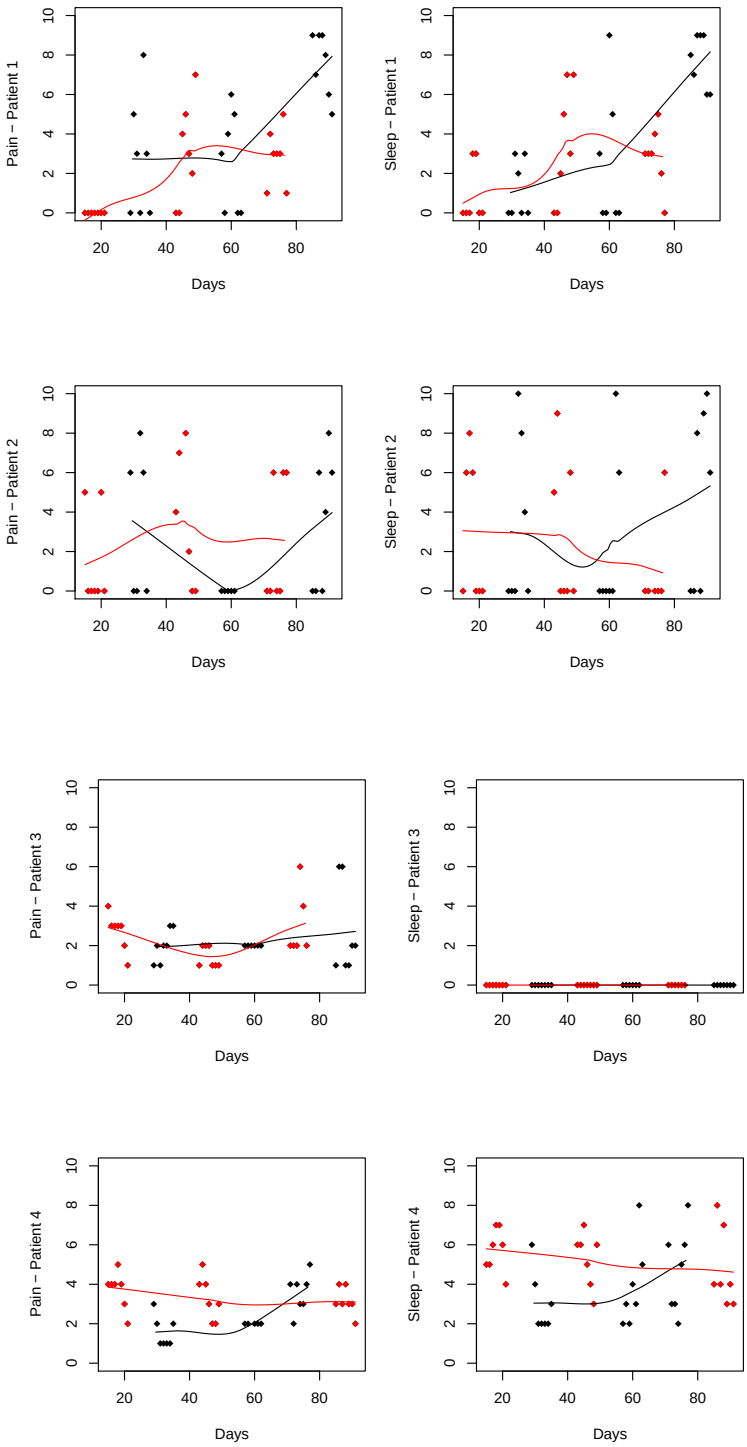

Figure 3.4: Real data from Yelland et al. (2009). The bivariate response is always displayed in the two columns, the first, second, third and fourth patients are reported in the four rows. Black circles refer to the active drug treated observations, red circles to the placebo ones

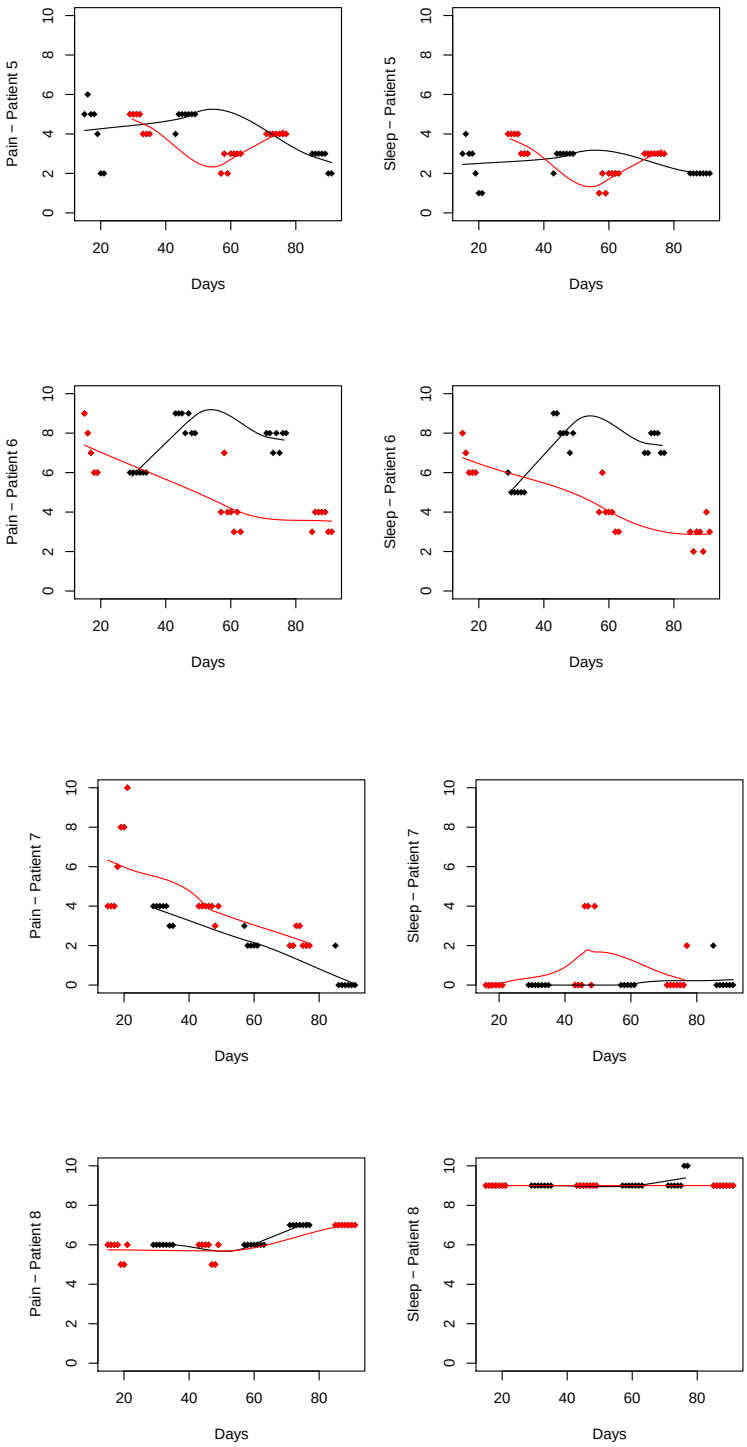

Figure 3.5: Real data from Yelland et al. (2009). The bivariate response is always displayed in the two columns, the fifth, sixth, seventh and eighth patients are reported respectively in the four rows. Black circles refer to the active drug treated observations, red circles to the placebo ones

Chapter 4

Univariate CUB models

4.1 Introduction

Analysis of ordinal data is nowadays receiving a growing theoretical and applied interest in many fields (see Agresti (1999, 2002, 2010); Agresti and Natarajan (2001)). For instance we often need this kind of analysis in marketing, when we wish to study the preferences of consumers about a set of products (market segmentation), or in clinical studies, when researchers wish to rate different treatments or drugs which are measured through ordinal categorical scores. With reference to these examples, the evaluation of the product or the rating of the treatments are processes which depend on specific subject (consumer or patient) and object (product or drug) characteristics. Formally, the rating issue arises when the subject is asked to express on a Likert scale a level of satisfaction or a degree of evaluation for the object of interest. In this chapter we propose a permutation technique to test for covariate influence on rating ordinal responses, working within a particular parametric framework for this kind of data.

In this sense several regression techniques have been proposed in order to analyze categorical data. For instance discriminant analysis performs a classification analysis for categorical responses, seeing the possible values that the outcome can take as possible categories and working on the likelihoods coming from the estimation of the models for the multivariate vector of covariates in the several categories. Indeed the covariates are here assumed to be realizations of normal random variables and a Bayes argument is used to construct the conditional probability of having a specific multivariate realization of the independent variable given the category the subject belongs to. Hence in this case the classical link between a linear predictor and the distribution of the categorical response does not exist. Another parametric modeling framework is based on the General Linear Models (GLM) paradigm introduced by McCullagh and Nelder (1989) and, specifically, by McCullagh (1980) for ordinal data. In this case the covariates are considered as fixed given values and some aspect of the distribution of the ordinal response is modeled as a function of the ordinal response. In the same vein, a new approach is represented by CUB models, which have been introduced by Piccolo (2003), D'Elia and Piccolo (2005) and subsequently generalized by Piccolo and D'Elia (2008) and Iannario and Piccolo (2009) (the acronym CUB is derived from the presence of ***C***ovariates in the

mixture of discrete ***U***niform and shifted ***B***inomial distributions). Again the ordinal response's distribution is modeled on the basis of the values of the covariates, which are considered to be fixed and not realization of a random vector. Briefly, CUB models are generated by a class of discrete probability distributions and allow to analyze ordinal data by taking into account the presence of two intrinsically continuous quantities (*feeling* and *uncertainty*) pertaining to the response. The rationale of such proposal is developed according to a psychological mechanism which explains the selected choice as the result of a personal liking/disliking (feeling) towards the object under judgment and an inherent indecision (uncertainty) generated by multiple alternatives. Formally the model is defined as a mixture of two discrete distributions where feeling and uncertainty components are explained by a shifted Binomial and Uniform random variables, respectively.

The definition of the two distributions contributing to the mixture comes from the following considerations. As far as the first component is concerned, the choice of a particular score may be interpreted as result of the pairwise comparisons between the possible scores, so that each comparison generates a dichotomous choice (Bernoulli experiment). Consider, for instance, a scale from 1 to m, with $m = 4$, and suppose that a subject is asked to rate a particular item (and let Y denote the response random variable). Then, if the subject's response is $y = 3$, he/she preferred the score 3 with respect to the other possible scores (1, 2 and 4), that is he/she chose 3 in the pairwise comparisons 1 versus 3, 2 versus 3 and 3 versus 4. If we assume that the random variables describing the comparisons are (approximately) independent Bernoulli distributions with parameters $(1 - \xi)$ and ξ, as the probability that each comparison is lost and won by the first score, respectively, then a given sequence of failure/success has (approximately) a probability of $(1-\xi)^{y-1}\xi^{m-y}$. A combinatorial argument immediately proves that the probability of a given choice is $\binom{m-1}{y-1}(1-\xi)^{y-1}\xi^{m-y}$. In addition, a heuristic argument may justify the use of this distribution since it is able in an effective way to map a continuous latent variable into a discrete set of values $1, \ldots, m$ and to allow different skewness and shapes (see Piccolo (2003)). A formal justification for the selection of a Uniform distribution as a building block for modeling uncertainty is its simplicity (absence of parameters) and the property of being a random variable which maximizes entropy over all discrete distributions with a finite support. In fact, we are not saying that a portion of respondents belongs to completely lazy people subset (who choose a category only by a chance mechanism) but we are assuming that each respondent manifests a propensity to act according such extreme behavior. The weight of this propensity is $(1 - \pi)$ and it is related to the second parameter of the model.

It is of course of interest to include some covariates of interest in order to model the distribution of the respondents' choices. Hence the main generalization of CUB model consists of the inclusion of the dependence of one or both parameters on some covariates. In fact, interpretation and fitting of CUB models generally improve when we relate feeling and uncertainty to subjects' characteristics since this allows to explain the different patterns of responses and the presence of clusters among respondents. It is worth of interest to check the significance of such relationship within an inferential framework, and thus to be able to identify the covariates that really influence the ordinal response while controlling the inferential errors. Moreover, given the pronounced importance

of ordinal data analysis in applied fields, it is surely of interest to specify a testing procedure which is adequate also for small or moderate sample sizes. More precisely, in this chapter we aim at presenting a permutation solution to test for the effect of a covariate on a rating response within the CUB modeling framework. Several types of permutation tests have been proposed in order to compare different models, and a comprehensive review of the literature is given by Anderson and Ter Braak (2003). Permutation strategies can be divided in two groups: permutation of raw data and permutation of residuals. The solution we propose belongs to the former class. Within this group of methods there are three important approaches: the first one is the unrestricted permutation of raw data (see Manly (1997)), which is an exact solution only in the case of one factor of interest. Restricted permutations of raw data can also be performed (see Good (1994); Edgington (1995)), and it is an exact test when we do not deal with interaction terms. A third solution is the synchronized permutation test (see Salmaso (2003); Basso et al. (2007)). Simulation studies (see Gonzalez and Manly (1998)) show that no method is uniformly the best over all kinds of design. Our solution belongs to the family of restricted permutation tests.

Section 2 is devoted to the formalization of the CUB model and the summary of the main inferential parametric results. The proposed permutation test is presented in Section 3, where the iterative procedure is described and a stepwise procedure for the choice of a suitable CUB model while controlling for the multiplicity issue is also proposed. Section 4 reports the results of a simulation study on the proposed solution. The applicability, in terms of minimum sample size, of our proposal is discussed in Section 5. Section 6 presents the application of the proposed test to a real dataset. At the end the conclusions are summarized in Section 7.

4.2 Formalization of the model

We consider that respondents are asked to select a given category in a set of m given and ordered alternatives related to satisfaction, agreement, attraction, evaluation, perception, worry, and so on. An independent and identically distributed (*i.i.d.*) sample of size n is assumed to be drawn from one or more populations of interest. Thus, we are collecting ratings of n people with respect to a definite item, object, sentence, question, and so on, and the observed sample will be denoted by $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$.

This process defines a discrete random variable Y whose support is $\{1, 2, \ldots, m\}$, for a given and known m. On the basis of the logic and psychological hypotheses listed in the previous section, we assume that such random variable has the following CUB probability distribution:

$$Pr\left(Y = y\right) = \pi \binom{m-1}{y-1} (1-\xi)^{y-1} \xi^{m-y} + (1-\pi) \left(\frac{1}{m}\right), \quad y = 1, 2, \ldots, m,$$

where $\pi \in (0, 1]$ and $\xi \in [0, 1]$. With such constraints, this distribution is always well defined and Iannario (2010) proved its identifiability for any $m > 3$. Notice that the parametric space of such random variable is the left open unit square, that is :

$$\Omega(\pi, \xi) = \{(\pi, \xi) : \ 0 < \pi \leq 1; \, 0 \leq \xi \leq 1\}.$$

Parameters ξ and π are related to feeling and uncertainty, respectively. As a matter of fact, the CUB distribution is positively (negatively) skewed when its mode is less (greater) than the midrange $(m+1)/2$ and thus low (high) ratings have higher probabilities: this happens for $\xi > 1/2$ ($\xi < 1/2$, respectively). As a consequence, $(1-\xi)$ denotes a measure of *feeling* towards the object/service/item. On the contrary, the indecision of the respondents increases directly with the weight of the Uniform distribution in the mixture; thus, $(1-\pi)$ is a direct measure of the *uncertainty* that characterizes the ordinal choice. In addition, since there is a one-to-one correspondence among a point in the unit square and a CUB model, it is possible to achieve effective interpretations of estimated results in terms of feeling and uncertainty, respectively, as shown by Iannario (2008a) and Corduas et al. (2010), for instance.

As said in the previous section, one of the most relevant and simplest generalization stems from the inclusion of subjects' covariates. The aim is to use such covariates for explaining feeling and/or uncertainty parameters and for improving model fitting. Hence, the general formulation of a CUB (p, q) model (with p covariates to explain uncertainty and q covariates to explain feeling) is expressed by the *stochastic component*:

$$Pr\left(Y = y|\boldsymbol{z_i}, \boldsymbol{w_i}\right) = \pi_i \binom{m-1}{y_i-1}(1-\xi_i)^{y_i-1}\xi_i^{m-y_i} + (1-\pi_i)\left(\frac{1}{m}\right), \quad (4.1)$$

with $y = 1, 2, \ldots, m$ and *two systematic components*:

$$\pi_i = \frac{1}{1+e^{-\boldsymbol{z_i}\,\boldsymbol{\beta}}}; \qquad \xi_i = \frac{1}{1+e^{-\boldsymbol{w_i}\,\boldsymbol{\gamma}}}; \qquad i = 1, 2, \ldots, n, \quad (4.2)$$

where $\boldsymbol{z_i}$ and $\boldsymbol{w_i}$ are the covariates row-vectors for explaining π_i and ξ_i, respectively, and $\boldsymbol{\psi} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ is the vector of parameters associated to the covariates. Here the logistic formulation has been chosen since it is the simplest and common mapping among the real line and the unit interval; however, other convenient links (as probit, log-log complimentary, etc.) might be introduced if necessary.

In the current literature, the relationship has been usually established between covariates and ratings through the expectation of the response (as in the GLM framework). Instead, notice that in CUB models this link is straightly expressed through the parameters since:

$$E[Y] = \pi(m-1)\left(\frac{1}{2} - \xi\right) + \frac{m+1}{2},$$

and thus different parameter vectors $(\pi, \xi)'$ can generate the same expectation.

Inference on CUB models has been mainly developed in a parametric framework. Given a sample of observed ordinal data and row-vectors of covariates $(y_i, \boldsymbol{z_i}, \boldsymbol{w_i})$, for $i = 1, \ldots, n$, the log-likelihood function for the parameter vector $\boldsymbol{\psi} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ of a general CUB (p, q) model is defined by

$$\ell\left(\boldsymbol{\psi}\right) = \sum_{i=1}^{n} \log\left[\frac{1}{1+e^{-\boldsymbol{z_i}\,\boldsymbol{\beta}}}\left(\binom{m-1}{y_i-1}\frac{e^{-\boldsymbol{w_i}\,\boldsymbol{\gamma}(y_i-1)}}{(1+e^{-\boldsymbol{w_i}\,\boldsymbol{\gamma}})^{m-1}} - \frac{1}{m}\right) + \frac{1}{m}\right].$$

In order to compute maximum likelihood estimates of the parameters with an EM algorithm, Piccolo (2006) derived a statistical procedure and Iannario and Piccolo (2009) made available an effective `R` software for its implementation.

Some measures for testing the adequacy of the estimated models (asymptotic significance of parameters, log-likelihood and fitting measures, simulations, graphical displays, and so on) are also currently available.

4.3 A permutation test on the covariates

4.3.1 The permutation procedure

We propose a nonparametric solution for testing the significance of the covariate coefficients of the CUB model, based on the permutation of the tested covariates. Indeed, the parametric solutions for this testing problem are usually based on asymptotic holding assumptions on the distribution of the maximum likelihood statistics under H_0, as Wald and likelihood ratio tests, for instance. Such conditions are not valid for small sample size and thus a suitable nonparametric solution is needed.

Let us consider a general CUB (p, q) model. Test on the adequacy of the estimated model and test on the comparison between nested models can be considered as a unique inferential problem, where under the null hypothesis one or more coefficients are equal to zero. Indeed, testing the adequacy of a CUB (p, q) model consists in comparing the general model with $p + q$ covariates with the simplest model with no covariates. Let us formalize the problem of interest: let us consider the CUB model where the response variable Y follows the mixture distribution as described in section 2 (equations 4.1-4.2). If we denote by Ψ the parametric space, then $\boldsymbol{\psi} \in \Psi$ and the general null hypothesis of interest can be written as

$$H_0 : \quad \boldsymbol{\psi} \in \Psi_0,$$

with $\Psi_0 \subset \Psi$. In Ψ_0 some elements of $\boldsymbol{\beta}^* = (\beta_1, \ldots, \beta_p)'$ and/or some elements of $\boldsymbol{\gamma}^* = (\gamma_1, \ldots, \gamma_q)'$ are equal to zero. The alternative hypothesis can be represented as follows

$$H_1 : \quad \boldsymbol{\psi} \in \Psi_1,$$

where $\Psi_1 = \Psi - \Psi_0$.

In the present section, we refer to the full model as the model specified under the alternative hypothesis, that is the CUB (p, q) model. Let us describe the permutation test: suppose to observe a sample of data described by the row vectors $(y_i, \boldsymbol{z}_i, \boldsymbol{w}_i)$, for $i = 1, \ldots, n$, generated under H_0 by a CUB (p_1, q_1) model (null model), with $p_1 \leq p$, $q_1 \leq q$ and $p_1 + q_1 < p + q$. When not all the covariates are tested let us assume that the non tested covariates are categorical. Moreover let us assume that the observed frequencies of each combination of values of the non tested covariates are greater than one (all-cell-permutation condition), i.e. there are at least 2 respondents characterized by each possible combination of values of non tested covariates. As we will see, in presence of non tested covariates this condition is essential for the applicability of the proposed method.

In other words, we are testing $p_2 = p - p_1$ covariates for π and $q_2 = q - q_1$ covariates for ξ. Without loss of generality, let us assume to test the last p_2 covariates for π and the last q_2 covariates for ξ. Hence, the tested hypotheses can be written as

$$H_0 : \beta_{p_1+1} = \cdots = \beta_p = \gamma_{q_1+1} = \cdots = \gamma_q = 0 \tag{4.3}$$

versus

$$H_1 : \exists r \in \{p_1+1, p_1+2, \ldots, p\} : \beta_r \neq 0 \text{ or } \exists s \in \{q_1+1, q_1+2, \ldots, q\} : \gamma_s \neq 0. \tag{4.4}$$

It is helpful to organize the sample information (dataset) as a matrix like the following:

$$\left(\begin{array}{c|ccc|ccc|ccc|ccc} y_1 & z_{11} & \cdots & z_{1p_1} & z_{1,p_1+1} & \cdots & z_{1p} & w_{11} & \cdots & w_{1q_1} & w_{1,q_1+1} & \cdots & w_{1q} \\ y_2 & z_{21} & \cdots & z_{2p_1} & z_{2,p_1+1} & \cdots & z_{2p} & w_{21} & \cdots & w_{2q_1} & w_{2,q_1+1} & \cdots & w_{2q} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ y_n & z_{n1} & \cdots & z_{np_1} & z_{n,p_1+1} & \cdots & z_{np} & w_{n1} & \cdots & w_{nq_1} & w_{n,q_1+1} & \cdots & w_{nq} \end{array}\right)$$

where the first column corresponds to the vector $\boldsymbol{y}$ of observed values of the response and the distinction between non tested and tested covariates for π and for ξ is highlighted.

Let us denote with $\boldsymbol{Z_1}$ and with $\boldsymbol{W_1}$ the sub-matrices of the dataset corresponding to the non tested covariates for π and ξ respectively (second and fourth block of columns), and similarly with $\boldsymbol{Z_2}$ and $\boldsymbol{W_2}$ the sub-matrices corresponding to the tested covariates (third and fifth block of columns).

The starting point for the construction of any permutation test is the identification of exchangeable units in the dataset under the null hypothesis; after that is done and a suitable test statistic is chosen, permutations of such quantities can be performed and the distribution of the tests statistic under the null hypothesis can be estimated. Under our null hypothesis the rows of the tested covariates $\boldsymbol{Z_2}$ and $\boldsymbol{W_2}$ are exchangeable within each block of rows having the same combination of values for the non tested covariates: indeed under the null hypothesis the tested covariates are not included in the model so rows having the same combinations of values for the non tested covariates determine responses from CUB models having the same values for the uncertainty and feeling parameters, while rows having different combinations of values determine responses coming from different CUB models. At this point, after having chosen a test statistic, its permutation distribution can be calculated through suitable (constrained if needed, see Remark 1 below) permutations in the dataset. Let us assume, without loss of generality, a suitable test statistic t is chosen such that H_0 is rejected for large values of t. Let us set also a large number B of permutations that can coincide either with the total number of possible permutations or with a lower number, depending on the choice to consider the whole permutation space or a random sample of permutations for computational convenience (see Pesarin and Salmaso (2010)). The procedure consists of the following steps:

i. Calculate the observed value of the test statistic as function of the observed dataset $t_o = t(\boldsymbol{y}, \boldsymbol{Z_1}, \boldsymbol{Z_2}, \boldsymbol{W_1}, \boldsymbol{W_2})$.

ii. Permute the rows of $\boldsymbol{Z_2}$ and $\boldsymbol{W_2}$, keeping fixed the remaining elements of the dataset. When there is at least one non tested covariate, constrained permutations within blocks have to be performed (see also Remark 1 below). Let us denote with ${}_1\boldsymbol{Z_2^*}$ and ${}_1\boldsymbol{W_2^*}$ the results of this permutation, i.e. the permuted matrices.

iii. Calculate the value of the test statistic corresponding to the permuted dataset.

iv. Repeat Steps `ii` and `iii` B times, yielding the permutation distribution of the test statistic. The value of t corresponding to the b-th permutation

is ${}_b t^* = t(\boldsymbol{y}, \boldsymbol{Z_1}, {}_b\boldsymbol{Z_2^*}, \boldsymbol{W_1}, {}_b\boldsymbol{W_2^*})$, $b = 1, 2, \ldots, B$, where ${}_b\boldsymbol{Z_2^*}$ and ${}_b\boldsymbol{W_2^*}$ represent the result of the b-th permutation on X_2 and W_2 respectively.

v. Calculate the p-value λ of the test as usual according to the permutation distribution of t, i.e. $\lambda = \sum_{b=1}^{B} I_{[t_0,\infty)}({}_b t^*)/B$ where $I_{[t_0,\infty)}(t) = 1$ if $t \geq t_0$ and 0 otherwise.

The idea of performing constrained permutations within blocks is displayed in Figure 4.1, in which we consider the particular case of two dichotomous non tested covariates, w_1 and w_2, and two tested covariates, w_3 and w_4, for ξ. We can see that is this case we have four possible blocks, one for each possible combination of values of the two non tested covariates. Within each of the four blocks then, we can permute the rows of tested covariates, as indicated in the scheme.

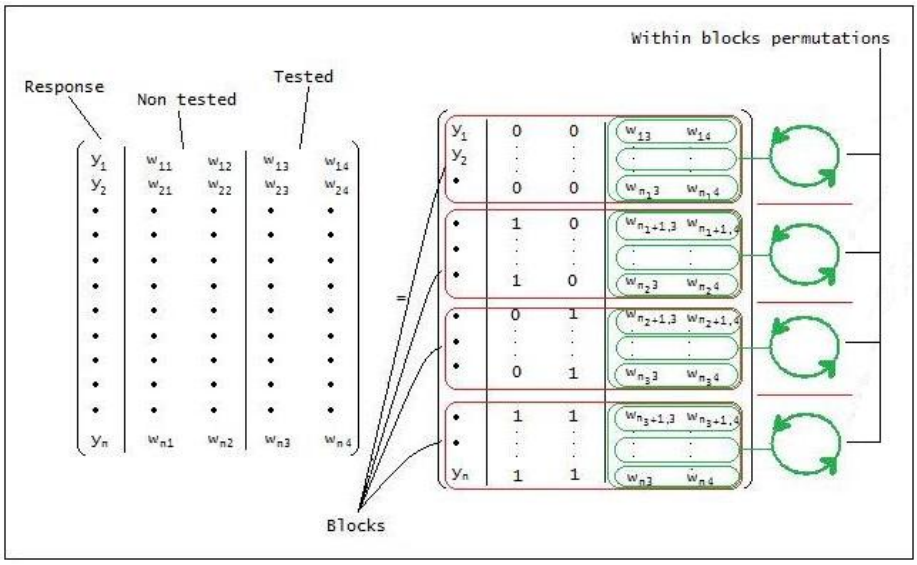

Figure 4.1: Permutation scheme. Case of two dichotomous non tested covariates, w_1 and w_2, and two tested covariates, w_3 and w_4, for ξ.

To perform the described procedure, many possible test statistics can be chosen. We decided to use the following classical parametric likelihood based test statistics:

- The classical likelihood-ratio statistic $t_{lrt} = 2\left[\ell\left(\hat{\boldsymbol{\psi}}\right) - \ell\left(\hat{\boldsymbol{\psi}}_0\right)\right]$, where $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\psi}}_0$ represent the maximum likelihood estimator of $\boldsymbol{\psi}$ under the full model and null model, respectively. Consider that at each permutation only the full model needs to be estimated.

- The Wald-type test $t_{wald} = (|\frac{\hat{\beta}_{p_1+1}}{se(\hat{\beta}_{p_1+1})}|, \ldots, |\frac{\hat{\beta}_p}{se(\hat{\beta}_p)}|, |\frac{\hat{\gamma}_{q_1+1}}{se(\hat{\gamma}_{q_1+1})}|, \ldots, |\frac{\hat{\gamma}_q}{se(\hat{\gamma}_q)}|) \cdot \boldsymbol{c}$, where $\boldsymbol{c} \in \Re_+^{p_2+q_2}$ is a suitable vector of weights and $se(\cdot)$ denotes the standard error of an estimator. Hence this test statistic is a linear combination of the partial test statistics on the single tested coefficients and

each single test statistic consists in the absolute value of the standardized estimator.

- The Non Parametric Combination (NPC-type) test $t_{npc} = \varphi_s(\lambda_1, \ldots, \lambda_{p_2+q_2}; \boldsymbol{c})$, where λ_k is the p-value of the k-th partial test related to a covariate ($k = 1, \ldots, p_2 + q_2$), $\boldsymbol{c}$ a vector of weights and $\varphi_s(\cdot)$ a suitable function according to the NPC of dependent permutation tests theory (see Pesarin and Salmaso (2010)). Notice that the Wald-type test can be considered a particular case of the NPC-type test, when the so-called direct combination is applied.

Remark 1. As noticed above, when there is at least one non tested covariate (i.e. when the null model include not only the constant) under the null hypothesis not all the rows of the tested covariates are exchangeable, since the probability distribution $Pr(Y = y|\boldsymbol{z}, \boldsymbol{w})$ changes according to the values of the non-tested covariates. Anyhow, the same probability distribution can be found in blocks of rows with the same combination of values for the non-tested covariates, and the rows of the tested covariates are exchangeable within these blocks.
Remark 2. As already said, in order to determine the permutation distribution of the test statistic, only the rows of the tested covariates have to be permuted. Furthermore, the permutations have to be synchronized, so that the dependence among the tested covariates for the same respondent can be taken into account.
Remark 3. Even if the proposed method could be considered a semi-parametric procedure, since it combines parametric and nonparametric tools, it can be properly classified as a permutation approach, because in this framework the parametric techniques taken into account can be seen just as a general way to manipulate the data and calculate the values of a test statistic. This is surely a full nonparametric method because no assumption is done and no asymptotic theory is applied to explicitly describe the distribution of the test statistic.
Remark 4. The proposed method differs from the so-called model-based permutation approaches used in linear models, where the permuted elements consist in the error terms. Conversely, in order to calculate or estimate the permutation distribution of the test statistic, in this nonlinear model we need to permute some of the covariate vectors.

We remind that the procedure is defined in general for categorical or discrete non tested covariates. However it can be applied also in the case of continuous non tested covariates after a suitable class transformation or categorization of them. Anyway, in presence of non tested covariates, regardless of the nature of them, the all-cell-permutation condition has to be satisfied. Table 4.1 displays the cases in which the permutation solution can be applied and the constraints to be satisfied according to the situations. Finally, it should be considered that the procedure is quite general and many other test statistics can be chosen for its application.

4.3.2 A stepwise procedure for covariate selection

In practice, it is useful to establish a suitable procedure or a set of rules to find the best model for the data. Hence, in principle, many tests need to be performed in order to decide for a suitable model. Usually, researcher may work in two directions: (i) starting from a set of covariates, he/she may specify the full CUB (p, q) model, including all the covariates, and go back (backward stepwise

Table 4.1: Applicability of the permutation solution.

Covariates under H_0	**Covariates**	
	Categorical or discrete	*Continuous*
None	Always applicable	Always applicable
One or more	All-cell-permutation condition	All-cell-permutation condition, after transformation

procedure) removing non significant covariates until he finds a parsimonious but well fitting model; (ii) otherwise, he/she may follow a forward stepwise procedure, starting from the simplest model (the one with only the constant covariate), and adding one covariate at a time until a satisfactory model for the data has been found.

We propose a forward stepwise procedure based on permutation inference in order to find a suitable model. Our choice of a forward procedure mainly comes from one reason: in the general case the permutation solution performs constrained permutations within blocks; the number of blocks increases in geometrical progression with the number of non-tested covariates, raising the risk that the all-cell-permutation condition is not satisfied. Hence the solution we propose consists in starting from the simplest model with only the constant covariate, and adding one covariate at each step. At every step significance tests on the covariates not already included in the model are performed and the covariate related to the more significant coefficient is included in the model. Choosing a suitable procedure is important also because different procedures of model selection can lead to different final models. Hence, a joint use of multiple permutation tests and closed testing methods is a suitable way to solve the problem of multiplicity and provide an efficient procedure for the model selection.

More specifically, let r be the number of covariates under study and s the number of covariates included in the model ($s = 1, \ldots, r$). Hence when all the covariates affect both feeling and uncertainty the model is a CUB (r, r). The procedure works as follows:

i. Fit the null CUB $(0, 0)$ model ($s = 0$).

ii. Perform $r - s$ permutation tests to test the null model against each of the possible CUB models obtained adding one of the $r - s$ available covariates. To perform this step, the Wald-type or the NPC-type permutation tests defined in the previous section can be applied.

iii. Apply the NPC methodology (see Pesarin and Salmaso (2010)) to combine the single permutation tests (partial tests) in step (ii) and obtain a global test on the joint nullity of the parameters versus the significance of some of them.

iv. If the null hypothesis of the global test at step (iii) cannot be rejected, no tested covariate should be added to the model. Then the selection procedure ends. If the global null hypothesis at step (iii) is rejected then adjust the p-values of the partial tests. To do it, apply the closed testing non

parametric procedure proposed by Pesarin and Salmaso (2010). Specify the new null model adding the covariate related to the lowest adjusted p-value. Hence s has increased by one.

v. Starting from the new null model, repeat steps (ii), (iii) and (iv) on the remaining covariates.

vi. The selection procedure ends when $s = r$ or when the null hypothesis of the global test is not rejected.

It should be noted that the expression "add a covariate" means the inclusion of a variable in the model as a covariate of feeling, uncertainty or both. Hence three tests have to be performed to decide how many and which parameters are affected by the covariate. Thus we propose to test the inclusion of a covariate in the model in the following way:

1. Perform the partial permutation tests, $t_{(1,0)}$ (for π) and $t_{(0,1)}$ (for ξ) say, needed to verify if the covariate affects the parameters π or ξ.

2. Combine together the two tests of step (1), obtaining a global test. If the global test is not significant, conclude that the covariate has no influence on the response; if the combined test is significant, adjust the two partial p-values and include the covariate in the model according to the following rules: a) if the adjusted p-value of $t_{(1,0)}$ is less than the significance level (α), include the covariate for π; b) if the adjusted p-value of $t_{(0,1)}$ is less than α, include the covariate for ξ.

This way we are taking care of the multiplicity issue at a inner level, while deciding which parameters the covariate affects, and also at an outer level at each step of the procedure.

4.4 Power behavior

A Monte Carlo simulation study has been carried out in order to study the power behavior of the permutation test on covariates for finite sample size. The study is divided in two main parts: at a first step we considered the CUB $(1, 1)$, the CUB $(1, 0)$ and the CUB $(0, 1)$ models as the full model and the CUB $(0, 0)$ model as null hypothesis, in order to check the behavior of the permutation tests in the simplest possible cases. At a second step we extend the simulation study considering more complicated CUB models under the alternative hypothesis (up to the CUB $(3, 3)$ model).

In general in the simulation study the performance of the permutation tests has been compared with a parametric competitor in terms of rejection rates both under the null and alternative hypotheses for different settings. We considered the case of dichotomous covariates, which is a common case for several subjects' characteristics represented through dummy variables (gender, marital or employment status, qualitative membership, and so on). Actually, by using a dichotomous covariate W taking values in $\{0, 1\}$, we are comparing two populations which correspond to two different CUB models under the alternative hypothesis. Let us suppose that we wish to test if the covariate affects the feeling towards the object, then in this case we are testing the null hypothesis

$$H_0: \quad \gamma_1 = 0, \qquad \Leftrightarrow \qquad H_0: \quad \xi = \xi_{(0)} = \left[1 + e^{-\gamma_0}\right]^{-1}$$

Table 4.2: Settings for ξ in the simulation study.

$\xi_{(0)}$	γ_0	δ_ξ	$\xi_{(1)} = \xi_{(0)} + \delta_\xi$	γ_1	hypothesis
0.10	-2.1972	0.00	0.10	0.0000	H_0
0.10	-2.1972	0.20	0.30	1.3499	
0.10	-2.1972	0.40	0.50	2.1972	H_1
0.10	-2.1972	0.50	0.60	2.6027	

against the alternative

$$H_1: \quad \gamma_1 \neq 0, \qquad \Leftrightarrow \qquad H_1: \begin{cases} \text{if} \quad w = 0, & \xi = \xi_{(0)} = [1 + e^{-\gamma_0}]^{-1} \\ \text{if} \quad w = 1, & \xi = \xi_{(1)} = [1 + e^{-\gamma_0 - \gamma_1}]^{-1} \end{cases} .$$

Of course, a corresponding scheme applies for π. To test the CUB $(1,1)$ full model against a CUB $(0,0)$ null model, the same reasoning has to be done on both the uncertainty and feeling parameters, and the null tested hypothesis is:

$$H_0: \ \beta_1 = \gamma_1 = 0.$$

We performed the simulation study by generating data for different values of π and ξ in the compared populations or equivalently for different values of β_1 and γ_1. The same reasoning holds for more complicated CUB models.

In this context, it is immediate to derive formulas for a convenient re-parametrization Iannario (2008b). For simplicity for the case the covariate affects the feeling towards the object they are given by:

$$\gamma_0 = \log\left(\frac{\xi_{(0)}}{1 - \xi_{(0)}}\right), \qquad \gamma_1 = \log\left[\frac{\xi_{(1)}\left(1 - \xi_{(0)}\right)}{\xi_{(0)}\left(1 - \xi_{(1)}\right)}\right].$$

Regarding the first part of the simulation study, Tables 4.2 and 4.3 show the considered settings in the simulation study for ξ and π, respectively. The "distance" between the values of the parameters ξ and π for the two populations, under the alternative hypothesis, can be expressed by the values δ_ξ and δ_π, respectively. We remark that the case considered under the null hypothesis of a CUB $(0,0)$ model is characterized by $\pi = 0.8$ and $\xi = 0.1$, which corresponds to high feeling and moderate uncertainty, typical of many real applications. This is also the reason for starting from $\pi_0 = 0.8$ and adding negative δ_π to obtain the several values for π_1 (see Table 4.3). Figure 4.2 displays the simulation settings in the parametric space $\Omega(\pi, \xi)$, allowing for them an interpretation in terms of "distance" from the null hypothesis of a CUB $(0,0)$ model: larger points denote greater distance from the null model in terms of number of non-zero coefficients.

In the simulation study we considered the permutation solutions t_{lrt}, t_{wald} and t_{npc} with Tippett combining function (see Pesarin and Salmaso (2010)) and the parametric likelihood-ratio test, hereafter labeled as $t_{\text{par-lrt}}$. The test based on the statistic t_{npc} with Tippett combination, performs the Wald-type tests on the single parameters and combines the related p-values. Table 4.7 reports the estimated rejection probabilities of the compared tests on the parameters

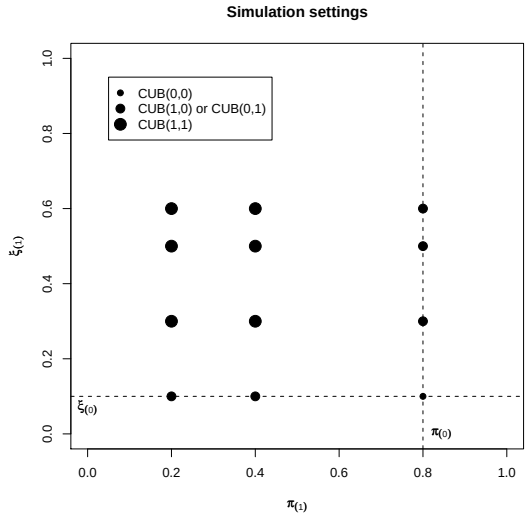

Figure 4.2: Settings for the simulation study, first part. CUB $(0,0)$ null model vs CUB $(1,1)$, CUB $(1,0)$ and CUB $(0,1)$ alternative models.

Table 4.3: Settings for π in the simulation study.

$\pi_{(0)}$	β_0	δ_π	$\pi_{(1)} = \pi_{(0)} + \delta_\pi$	β_1	hypothesis
0.80	1.3863	0.00	0.80	0.0000	H_0
0.80	1.3863	-0.40	0.40	-1.7918	H_1
0.80	1.3863	-0.60	0.20	-2.7726	

Table 4.4: Estimated rejection probabilities for $m = 7$, $n = 50$ and $\alpha = 0.05$.

		t_{lrt}	t_{wald}	t_{npc}	$t_{\text{par-lrt}}$
$\delta_\xi = 0$ $(\xi_{(0)} = \xi_{(1)} = 0.1)$	$\delta_\pi = 0$ $(\pi_{(0)} = \pi_{(1)} = 0.8)$	0.041	0.038	0.051	0.070
	$\delta_\pi = -0.4$ $(\pi_{(0)} = 0.8,\ \pi_{(1)} = 0.4)$	0.478	0.390	0.390	0.334
	$\delta_\pi = -0.6$ $(\pi_{(0)} = 0.8,\ \pi_{(1)} = 0.2)$	0.790	0.514	0.514	0.676
$\delta_\xi = 0.2$ $(\xi_{(0)} = 0.1,\ \xi_{(1)} = 0.3)$	$\delta_\pi = 0$ $(\pi_{(0)} = \pi_{(1)} = 0.8)$	0.887	0.874	0.874	0.790
	$\delta_\pi = -0.4$ $(\pi_{(0)} = 0.8,\ \pi_{(1)} = 0.4)$	0.754	0.646	0.626	0.806
	$\delta_\pi = -0.6$ $(\pi_{(0)} = 0.8,\ \pi_{(1)} = 0.2)$	0.780	0.564	0.567	0.857
$\delta_\xi = 0.4$ $(\xi_{(0)} = 0.1,\ \xi_{(1)} = 0.5)$	$\delta_\pi = 0$ $(\pi_{(0)} = \pi_{(1)} = 0.8)$	0.999	0.996	0.996	0.999
	$\delta_\pi = -0.4$ $(\pi_{(0)} = 0.8,\ \pi_{(1)} = 0.4)$	0.952	0.900	0.894	0.971
	$\delta_\pi = -0.6$ $(\pi_{(0)} = 0.8,\ \pi_{(1)} = 0.2)$	0.934	0.853	0.851	0.970
$\delta_\xi = 0.5$ $(\xi_{(0)} = 0.1,\ \xi_{(1)} = 0.6)$	$\delta_\pi = 0$ $(\pi_{(0)} = \pi_{(1)} = 0.8)$	0.999	0.998	0.998	0.999
	$\delta_\pi = -0.4$ $(\pi_{(0)} = 0.8,\ \pi_{(1)} = 0.4)$	0.981	0.942	0.946	0.992
	$\delta_\pi = -0.6$ $(\pi_{(0)} = 0.8,\ \pi_{(1)} = 0.2)$	0.962	0.909	0.902	0.985

γ_1 and β_1 for $m = 7$, $n = 50$ and working at a nominal level of $\alpha = 0.05$. $B = 1000$ random permutations and $MC = 1000$ Monte Carlo replications have been considered. Notice that under the null hypothesis ($\delta_\pi = \delta_\xi = 0$) all the permutation solutions control the type I error at the nominal level, while the parametric counterpart does not. Under the alternative hypothesis, permutation inference seems to be more powerful than the parametric one under a CUB $(1, 0)$ or a CUB $(0, 1)$ model; instead opposite situation can be appreciated when under the alternative hypothesis we specify a CUB $(1, 1)$ model. We have to remark that a proper comparison between permutation methods and parametric test is not possible, because the different solutions are not of the same size (the actual levels of the tests are different). In general t_{lrt} is the most powerful permutation solution among the ones considered in this study.

Table 4.8 reports the same simulation results for $n = 100$. By increasing the sample size, we observe a better performance of the parametric solution under the null hypothesis, even if the estimated probability of type I error still exceeds the nominal level α. In general all the estimated rejection probabilities increase with respect to the case of $n = 50$.

In the second part of the simulation study models with more than one covariate for π and ξ (up to the CUB $(3, 3)$ model) are considered under the alternative hypothesis, to study how the number of covariates affects the inferential results and the performance of the tests. The fact that dichotomous covariates are considered means that every significant covariate introduced in

Table 4.5: Estimated rejection probabilities for $m = 7$, $n = 100$ and $\alpha = 0.05$.

		t_{lrt}	t_{wald}	t_{npc}	$t_{\text{par-lrt}}$
$\delta_\xi = 0$ ($\xi_{(0)} = \xi_{(1)} = 0.1$)	$\delta_\pi = 0$ ($\pi_{(0)} = \pi_{(1)} = 0.8$)	0.05	0.048	0.053	0.059
	$\delta_\pi = -0.4$ ($\pi_{(0)} = 0.8$, $\pi_{(1)} = 0.4$)	0.793	0.776	0.776	0.658
	$\delta_\pi = -0.6$ ($\pi_{(0)} = 0.8$, $\pi_{(1)} = 0.2$)	0.976	0.871	0.871	0.959
$\delta_\xi = 0.2$ ($\xi_{(0)} = 0.1$, $\xi_{(1)} = 0.3$)	$\delta_\pi = 0$ ($\pi_{(0)} = \pi_{(1)} = 0.8$)	0.992	0.992	0.992	0.986
	$\delta_\pi = -0.4$ ($\pi_{(0)} = 0.8$, $\pi_{(1)} = 0.4$)	0.976	0.967	0.948	0.985
	$\delta_\pi = -0.6$ ($\pi_{(0)} = 0.8$, $\pi_{(1)} = 0.2$)	0.988	0.932	0.929	0.993
$\delta_\xi = 0.4$ ($\xi_{(0)} = 0.1$, $\xi_{(1)} = 0.5$)	$\delta_\pi = 0$ ($\pi_{(0)} = \pi_{(1)} = 0.8$)	1.000	1.000	1.000	1.000
	$\delta_\pi = -0.4$ ($\pi_{(0)} = 0.8$, $\pi_{(1)} = 0.4$)	0.999	0.995	0.996	1.000
	$\delta_\pi = -0.6$ ($\pi_{(0)} = 0.8$, $\pi_{(1)} = 0.2$)	0.995	0.989	0.989	0.998
$\delta_\xi = 0.5$ ($\xi_{(0)} = 0.1$, $\xi_{(1)} = 0.6$)	$\delta_\pi = 0$ ($\pi_{(0)} = \pi_{(1)} = 0.8$)	1.000	1.000	1.000	1.000
	$\delta_\pi = -0.4$ ($\pi_{(0)} = 0.8$, $\pi_{(1)} = 0.4$)	1.000	0.998	0.998	1.000
	$\delta_\pi = -0.6$ ($\pi_{(0)} = 0.8$, $\pi_{(1)} = 0.2$)	1.000	0.995	0.994	1.000

Table 4.6: Simulation settings: models considered under the alternative hypothesis and marginal and global effects of the covariates on π and ξ.

Settings	Model under H_1	$\delta_{\pi,(z_1)}$	$\delta_{\xi,(w_i)}$, $i = 1, \ldots, 3$	$\delta_{\xi,((w_1),(w_2),(w_3))}$
1	CUB $(0, 2)$	–	0.05	0.12
	CUB $(0, 3)$	–	0.05	0.21
	CUB $(1, 2)$	-0.6	0.05	0.12
	CUB $(1, 3)$	-0.6	0.05	0.21
2	CUB $(0, 2)$	–	0.10	0.26
	CUB $(0, 3)$	–	0.10	0.46
	CUB $(1, 2)$	-0.6	0.10	0.26
	CUB $(1, 3)$	-0.6	0.10	0.46
3	CUB $(0, 2)$	–	0.20	0.52
	CUB $(0, 3)$	–	0.20	0.76
	CUB $(1, 2)$	-0.6	0.20	0.52
	CUB $(1, 3)$	-0.6	0.20	0.76

the model determines two sub-populations to be compared; hence for instance when two covariates are introduced in the model four sub-populations are actually compared, identified by each combination of values taken by the covariates. Depending on which parameter the covariates influence, the compared sub-populations can be identified by CUB models with different values of the uncertainty or feeling parameter; for instance if two covariates are introduced for the feeling parameter we are actually dealing with four sub-populations having four different values of the feeling parameter $\xi_{(0,0)}, \xi_{(1,0)}, \xi_{(0,1)}, \xi_{(1,1)}$. In general a measure of the distance between the compared populations can be given by the values $\delta_{\pi,((z_1),\ldots,(z_p))} = \pi_{((z_1=1),\ldots,(z_p=1))} - \pi_{((z_1=0),\ldots,(z_p=0))}$ and $\delta_{\xi,((w_1),\ldots,(w_q))} = \xi_{((w_1=1),\ldots,(w_q=1))} - \xi_{((w_1=0),\ldots,(w_q=0))}$, which in a way summarize the joint influence of the covariates on the response. In this simulation study (without loss of generality) we worked with a fixed marginal effect of the only covariate considered for the uncertainty parameter $\delta_{\pi,(z_1)} = \pi_{(1)} - \pi_{(0)} = -0.6$ and with three fixed values for the marginal effect of the covariates considered for the feeling parameter $\delta_{\xi,(w_1)} = \xi_{(1,0,0)} - \xi_{(0,0,0)} = \delta_{\xi,(w_2)} = \xi_{(0,1,0)} - \xi_{(0,0,0)} = \delta_{\xi,(w_3)} = \xi_{(0,0,1)} - \xi_{(0,0,0)} = 0.05, 0.1, 0.2$ in the three considered settings. More precisely in each setting the marginal effects of the covariates are simulated all equal, hence $\delta_{\xi,(w_i)} = 0.05$ for $i = 1, \ldots, 3$ in the first setting, $\delta_{\xi,(w_i)} = 0.1$ for $i = 1, \ldots, 3$ in the second setting and $\delta_{\xi,(w_i)} = 0.2$ for $i = 1, \ldots, 3$ in the third setting. The null CUB $(0, 0)$ model has been kept with the same values $\pi = 0.80$ and $\xi = 0.10$ in all the settings. The considered scenarios are reported in Table 4.6, in terms of model considered under the alternative hypothesis and marginal and joint effects of the covariates on the π and ξ parameters.

Again we compare the permutation solutions t_{lrt}, t_{wald} and t_{npc} and the parametric likelihood-ratio test, hereafter labeled as $t_{\text{par-lrt}}$, considering a number of $B = 1000$ permutations and $MC = 1000$ replications. Tables 4.7 and 4.8 summarize the obtained results, in terms of estimated rejection probabilities of the compared tests, respectively for sample size $n = 50$ and $n = 100$. Notice how the power of all the procedures increases as the sample size n increases. It has also to be underlined that the parametric solution does not control the

Table 4.7: Estimated rejection probabilities for $m = 7$, $n = 50$ and $\alpha = 0.05$.

True model	Tested models		t_{lrt}	t_{wald}	t_{npc}	$t_{\text{par-lrt}}$
	H_0	H_1				
H_0	CUB $(0,0)$	CUB $(3,3)$	0.059	0.049	0.064	0.119
H_1 – Setting 1 $(\delta_{\pi,(z_1)} = -0.6, \delta_{\xi,(w_i)} = 0.05\, \forall i = 1,\dots,3)$	CUB $(0,0)$	CUB $(0,2)$	0.235	0.229	0.220	0.501
		CUB $(0,3)$	0.257	0.275	0.232	0.671
		CUB $(1,2)$	0.507	0.064	0.195	0.919
		CUB $(1,3)$	0.394	0.044	0.111	0.891
H_1 – Setting 2 $(\delta_{\pi,(z_1)} = -0.6, \delta_{\xi,(w_i)} = 0.1\, \forall i = 1,\dots,3)$	CUB $(0,0)$	CUB $(0,2)$	0.688	0.686	0.646	0.740
		CUB $(0,3)$	0.806	0.816	0.681	0.849
		CUB $(1,2)$	0.437	0.167	0.188	0.743
		CUB $(1,3)$	0.468	0.213	0.333	0.761
H_1 – Setting 3 $(\delta_{\pi,(z_1)} = -0.6, \delta_{\xi,(w_i)} = 0.2\, \forall i = 1,\dots,3)$	CUB $(0,0)$	CUB $(0,2)$	0.985	0.983	0.958	0.992
		CUB $(0,3)$	0.986	0.981	0.965	0.999
		CUB $(1,2)$	1	0.999	1	1
		CUB $(1,3)$	0.999	0.999	1	1

type I error when the sample size is low ($n = 50$); a better performance can be appreciated for $n = 100$. As a matter of fact an unusual power behavior can be observed for the estimated power of all the procedures: when a significant covariate is simulated for the uncertainty parameter π we can observe a decrease of the estimated rejection probabilities. This problem appears to be more serious for the permutation solutions based on the nonparametric combination of partial permutation tests (t_{wald} and t_{npc}). This strange behavior is probably due to the strange behavior of the partial test on the parameter associating the covariate to the feeling component, which had already been noticed in the first part of the study: we think that a possible reason for this is that when we introduce more uncertainty in a part of the observations (one of the two sub-samples identified by the uncertainty associated covariate) we are actually making more difficult for the tests to recognize the influence of the other covariates on the feeling parameter. Hence when we consider the likelihood-ratio test (both parametric and permutation versions) we loose some power but we are still able to see the improvement of the model due to the introduction of the covariates, but when we take into consideration the partial tests the problem becomes more serious. Anyway it has to be noticed that, for fixed values of $\delta_{\pi,(z_1)}$, the power of all the procedures still increases as we move away from the null hypothesis regarding the covariates influencing the feeling parameter (as $\delta_{\xi,(w_i)}$ increases, hence as we go from Setting 1 to Setting 3).

We can conclude that the parametric solution should be carefully used for low sample sizes, since asymptotic results do not hold in this case and the test does not control the type I error. Increasing the sample size, on one hand the performances of the parametric test under the null hypothesis improve, and on the other hand they tend to converge to the ones of the permutation tests also under the alternative hypothesis.

Table 4.8: Estimated rejection probabilities for $m = 7$, $n = 100$ and $\alpha = 0.05$.

True model	Tested models		t_{lrt}	t_{wald}	t_{npc}	$t_{\text{par-lrt}}$
	H_0	H_1				
H_0	CUB $(0,0)$	CUB $(3,3)$	0.039	0.035	0.0370	0.063
H_1 – Setting 1 $(\delta_{\pi,(z_1)} = -0.6, \delta_{\xi,(w_i)} = 0.05\,\forall i = 1,\dots,3)$	CUB $(0,0)$	CUB $(0,2)$	0.455	0.466	0.420	0.476
		CUB $(0,3)$	0.626	0.635	0.528	0.785
		CUB $(1,2)$	0.874	0.393	0.589	0.959
		CUB $(1,3)$	0.810	0.294	0.571	0.964
H_1 – Setting 2 $(\delta_{\pi,(z_1)} = -0.6, \delta_{\xi,(w_i)} = 0.1\,\forall i = 1,\dots,3)$	CUB $(0,0)$	CUB $(0,2)$	0.972	0.968	0.915	0.965
		CUB $(0,3)$	0.999	0.997	0.965	0.998
		CUB $(1,2)$	0.853	0.623	0.718	0.978
		CUB $(1,3)$	0.866	0.628	0.708	0.991
H_1 – Setting 3 $(\delta_{\pi,(z_1)} = -0.6, \delta_{\xi,(w_i)} = 0.2\,\forall i = 1,\dots,3)$	CUB $(0,0)$	CUB $(0,2)$	1	1	1	1
		CUB $(0,3)$	1	1	1	1
		CUB $(1,2)$	0.946	0.926	1	0.987
		CUB $(1,3)$	0.954	0.939	1	0.998

4.5 Minimum sample size for the application of the permutation test

Before performing a test of significance, every researcher should make sure that it controls the I type error, and consists in a powerful solution for the testing problem. Obviously the application of a testing procedure is not possible when the minimum possible p-value of the tests is greater than the significance level α, according to the distribution of the test statistic.

Permutation tests in fact can present this problem when the number of possible permutations is low: when all the possible permutations are considered, the p-value of a general permutation test which rejects the null hypothesis for high values of the test statistic t is given by

$$\lambda = \frac{\#\ (t^* \geq t^{oss})}{B},$$

where t^* are the values of the test statistic computed on the permuted samples, t^{oss} is the observed value of t and B is the total number of distinct permutations of the dataset. This quantity can vary from the minimum and the maximum values

$$\lambda_{min} = \frac{1}{B} \leq \lambda \leq \frac{B}{B} = 1 = \lambda_{max}.$$

Hence a necessary condition for the applicability of the method is $\lambda_{min} = 1/B < \alpha$, i. e. $B > 1/\alpha$. If we consider the usual 0.05 value for α, we obtain that the minimum number of total distinct permutations to have a powerful test is $B = 21$.

Let us adapt the above presented rule to the permutation test we are considering in this work. Let us indicate with n_j, $j = 1,\dots,M$ the number of subjects belonging to the jth block corresponding to a given combination of values/levels of the non-tested covariates; hence M indicates the number of blocks. The total number of distinct permutations is then given by

$$B = \prod_{j=1}^{M} n_j!.$$

Table 4.9: Minimum sample sizes (and the optimal partition of subjects in the blocks) for $\alpha = 0.05$ and several values for M.

M	n_{min}	$n_j,\ j = 1, \ldots, M$
1	$n = 4$	$n_1 = 4\,(B = 24)$
2	$n = 6$	$n_1 = 2, n_2 = 4\,(B = 48)$
3	$n = 7$	$n_1 = 2, n_2 = 2, n_3 = 3\,(B = 24)$
4	$n = 9$	$n_1 = 2, n_2 = 2, n_3 = 2, n_4 = 3\,(B = 48)$
$M \geq 5$	$n = 2M$	$n_j = 2,\ j = 1, \ldots, M\,(B = 2^M)$

Table 4.10: Minimum sample sizes (and the optimal partition of subjects in the blocks) for $\alpha = 0.10$ and several values for M.

M	n_{min}	$n_j,\ j = 1, \ldots, M$
1	$n = 4$	$n_1 = 4\,(B = 24)$
2	$n = 5$	$n_1 = 2, n_2 = 3\,(B = 12)$
3	$n = 7$	$n_1 = 2, n_2 = 2, n_3 = 3\,(B = 24)$
4	$n = 2M = 8$	$n_j = 2,\ j = 1, \ldots, M\,(B = 2^4 = 16)$
$M \geq 5$	$n = 2M$	$n_j = 2,\ j = 1, \ldots, M\,(B = 2^M)$

We remind that the all-cell-permutation condition needs to be satisfied in order to apply the permutation solution. This means that $\forall j,\ n_j \geq 2$. Hence the considered test does not need great samples sizes in the several blocks: for $\alpha = 0.05$, $M \geq 5$ the all-cell-permutation condition allows the application of the test, since $B \geq 2^5 = 32 > 21$. Tables 4.9 and 4.10 report the minimum sample sizes (and the optimal partition of subjects in the blocks) for several values for M respectively for two values of the significance level $\alpha = 0.05, 0.10$. Notice that the the minimum required sample sizes appear to be quite low. As one could have expected, they increase as M increases, i.e. as the number of non-tested covariates (or number of values that the non-tested covariates can take) increases. In real case applications a common problem consists in testing the influence of a covariate on the response in presence of two, three non-tested dichotomous covariates; this corresponds to $M = 4, 8$ respectively.

4.6 Real data application

Real data have been analyzed with the proposed permutation solution and they refer to a statistical survey on Passito, a typical Italian wine produced in the North-East of Italy (see Arboretti Giancristofaro et al. (2011)). A sample of 386 consumers living in Veneto region (Italy) has been asked to give an opinion about Passito by considering different points of view. Several covariates were also recorded. In this analysis we considered five response variables, representing satisfaction about five aspects of the wine:

- y_1: level of liking;
- y_2: satisfaction about aroma;
- y_3: satisfaction about the sweet taste;

- y_4: satisfaction about alcohol percentage;

- y_5: satisfaction about intensity of taste.

These responses are measured on a scale from 1 (maximum dissatisfaction) to 7 (maximum satisfaction). Three dichotomous covariates were considered in this analysis: Age (coded as 0 if the subject was ≤ 25 years old, and 1 other wise), Gender (0 for females and 1 for males), Residence (coded as 0 for East Veneto, and 1 for West Veneto).

We performed the proposed forward stepwise procedure in order to find a suitable model for each of the responses. The proposed t_{wald} permutation solution was used at step (ii) of the procedure; regarding the NPC methodology at step (iii), the Tippett combining function (see Pesarin and Salmaso (2010)) has been used for the combination of covariate related partial tests. Table 4.11 reports the obtained final models for the five responses of interest, with the estimates of the β and γ parameters associated to the added covariates, the estimates of the π and ξ parameters in the two groups identified by the tested covariate and the p-values of the permutation tests on the goodness-of-fit of the final models (i.e. the tests comparing the obtained models with the simplest model without any covariate). The estimates of the uncertainty parameter suggest, for all the responses, that we are in a very low uncertainty situation, i.e. the respondents seem to be quite sure about their ratings.

Table 4.11: Real case application results. Final models obtained from the proposed stepwise procedure: estimates of the parameters and p-values for the goodness-of-fit tests.

Response	Age		Gender		Residence		p-value
	π	ξ	π	ξ	π	ξ	
Level of liking	– – – – –	– – – – –	– – – – –	$\hat{\gamma}_1 = -0.616$ $\hat{\xi}_{(0)} = 0.589$ $\hat{\xi}_{(1)} = 0.437$	– – – – –	– – – – –	0.000
Satisfaction about aroma	– – – – –	– – – – –	$\hat{\beta}_1 = 1.112$ $\hat{\pi}_{(0)} = 0.563$ $\hat{\pi}_{(1)} = 0.797$	$\hat{\gamma}_1 = -0.318$ $\hat{\xi}_{(0)} = 0.430$ $\hat{\xi}_{(1)} = 0.354$	– – – – –	– – – – –	0.002
Satisfaction about the sweet taste	– – – – –	$\hat{\gamma}_1 = 0.35322$ $\hat{\xi}_{(0)} = 0.396$ $\hat{\xi}_{(1)} = 0.483$	– – – – –	$\hat{\gamma}_2 = -0.409$ $\hat{\xi}_{(0)} = 0.396$ $\hat{\xi}_{(1)} = 0.303$	$\hat{\beta}_1 = -2.213$ $\hat{\pi}_{(0)} = 0.888$ $\hat{\pi}_{(1)} = 0.465$	– – – – –	0.001
Satisfaction about the alcohol percentage	– – – – –	– – – – –	– – – – –	– – – – –	– – – – –	$\hat{\gamma}_1 = 0.275$ $\hat{\xi}_{(0)} = 0.494$ $\hat{\xi}_{(1)} = 0.562$	0.000
Satisfaction about the intensity of the taste	– – – – –	– – – – –	$\hat{\beta}_1 = 1.715$ $\hat{\pi}_{(0)} = 0.368$ $\hat{\pi}_{(1)} = 0.764$	– – – – –	– – – – –	– – – – –	0.001

All the considered covariates appear to affect some responses: results suggest that men are less uncertain on rating the considered responses, and their feeling is greater than that of women. Greater uncertainty and less feeling is observed for respondents living in West Veneto in comparison with East Veneto. We can also say that younger respondents are more satisfied about the sweet taste. At the end, it can be noticed that the goodness-of-fit tests give a strong evidence of the influence of the added covariates on each of the considered responses. The low p-values confirm the powerfulness of the permutation solution, already shown by the simulation study: thanks to the high sample size, even not so

extreme differences of the parameters values in the groups identified by the covariates are very well recognized by the permutation tests.

4.7 Main results

In this chapter we proposed a permutation solution to test for the effects of covariates on ordinal response variables within the CUB models framework. The nonparametric strategy is based on the constrained synchronized permutation of the tested covariates vectors. The aim of the work is to provide a competitive and well performing alternative to the standard parametric solution based on the asymptotic theory, in order to deal with real data problems with small sample size. The solution is applicable even with continuous covariates, after a suitable class transformation (very common practice in the analysis of statistical surveys). Moreover, a nonparametric forward stepwise procedure based on the permutation solution has been proposed to analyze real data problems.

A simulation study was performed to analyze the power behavior of the proposed solution and to compare it with a parametric counterpart (the classical likelihood-ratio test). Several test statistics have been implemented in the permutation solution. The results show that all the considered permutation tests control the type I error even for small sample size, while this is true for the parametric solution only for large sample sizes. The permutation solutions present high powers; they show to be competitive with respect to the parametric counterpart, even if a proper comparison cannot be done as the actual significance levels of the several tests are not the same. Regarding the permutation methods, even for a low sample size ($n = 50$), we can notice fast increase of the power as, for fixed values of $\delta_{\pi,(z_1)}$, the difference of the values of ξ in the compared populations under H_1 grows. Moreover we can notice that, more clearly the two permutation solutions based on the nonparametric combination of partial permutation test on the single regression parameters, but in general all the testing procedures present a decrease of the power when a covariate for the uncertainty parameter is introduced in the model. The performances of all the methods improve as we increase the sample size: the actual type I error rate of the parametric solution gets closer to the nominal level, the powers of all the tests increase and they tend to converge towards each other.

We also discussed the problem of which is the minimum sample size needed to perform the proposed permutation test. We presented the minimum sample sizes and the best partitions of subjects in the M blocks of covariates' combinations.

A real problem was analyzed through the proposed permutation solutions: a statistical survey on Passito wine has been considered and a forward stepwise procedure has been used to select a suitable model for some responses of interest in the study, by considering some explanatory covariates. Meaningful results have been obtained.

Both the simulation study and the real case application suggested that the proposed solution is well performing even when the sample size is not high and asymptotic theory cannot be applied. The use of this permutation tool is then recommended in presence of small sample sizes, and its usefulness is anyway confirmed even for problems with higher number of observations.

Chapter 5

Multivariate CUB models

5.1 Introduction

As pointed out in the previous chapter, analysis of ordinal data can be faced in many fields; for example in marketing to study the preferences of consumers about a set of products, or in clinical studies to rate treatments or drugs. Normally the ordinal response (rating) which is given to products or treatments is a process which depends on specific subject (consumer or patient) and object (product or drug) characteristic. In this area a new approach is represented by CUB models, which are generated by a class of discrete probability distributions, to model the data taking into account two intrinsically continuous quantities (feeling and uncertainty) pertaining to the response; these are modeled as a *shifted Binomial* and an *Uniform* random variable respectively. Let us consider the CUB model where the response variable Y follows the mixture distribution as described in the previous chapter (equations 4.1-4.2). Moreover let us use the same notation, hence we observe the sample $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ of responses and $\boldsymbol{z_i}$ and $\boldsymbol{w_i}$, with $i = 1, \ldots, n$ of subjects' covariate vectors for explaining π_i and ξ_i, the uncertainty and feeling parameters respectively. Therefore $\boldsymbol{\psi} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ denotes the vector of parameters associated to the covariates.

Inference on CUB models has been mainly developed in a parametric framework, via maximum likelihood and asymptotic theory (see Piccolo (2006) and Iannario and Piccolo (2009)) and, in the previous chapter we proposed a nonparametric approach by means of a permutation solution to test for the effect of covariates on a rating response within the CUB modeling framework. The method is based on the constrained permutation of raw data and it basically performs a test on the comparison between nested CUB models. As it is shown, our permutation test is a competitive alternative to the classical parametric test when the sample size is not high.

In this chapter we extend the above presented permutation solution to the case in which a multivariate ordinal response is observed. The procedure is described in Section 2, where we explain how nonparametric combination of dependent permutation tests (see Pesarin (2001) and Pesarin and Salmaso (2010)) is used to end up with a global tool for comparing several nested CUB models at the same time on several aspects of the multivariate response. In Section 3 a simulation study is carried out in order to explore the performances of the

multivariate test. In Section 4 the method is applied to two real datasets. The first real application is the same presented in Chapter 4, here analyzed from a multivariate point of view. The second real dataset regards the evaluation of the Ski School of Sesto Pusteria in the Trentino Alto Adige region (Italy). The conclusions of the chapter are summarized in Section 5.

5.2 A permutation test for multivariate responses

We propose a nonparametric solution for the test of significance for the coefficients of the covariates of the CUB model when a multivariate response is observed. We recall that in Chapter 4 the permutation test we proposed is based on the constrained permutations of the tested covariates. Now, we wish to test the global influence of one or more covariates on a multivariate response. Moreover, for each partial inferential problem (i.e. influence of the tested covariates on each component of the multivariate response) we are testing the influence of the same covariates. The global alternative hypotheses is usually defined when at least one of the partial null hypotheses is false.

We propose to solve the problem in the permutation framework, using nonparametric combination of dependent permutation tests (see Pesarin (2001) and Pesarin and Salmaso (2010)). In this way we can provide both a global measure of the significance of specific covariates on all the considered responses and partial adjusted tests for the single model comparisons. Let a multivariate response $\mathbf{Y}^{'} = (Y^1, \ldots, Y^d)$ be of interest, and let a set of its realizations $(\boldsymbol{y}^1, \ldots, \boldsymbol{y}^d)$, with $\boldsymbol{y}^l = (y_1^l, y_2^l, \ldots, y_n^l)'$, $l = 1, \ldots, d$, be observed on n respondents. Moreover let the subject related covariates $\boldsymbol{z_i}$ and $\boldsymbol{w_i}$, with $i = 1, \ldots, n$ be observed on the same set of respondents. Formally the procedure we propose works as follows:

i. Set the null and the alternative models that need to be compared (say $\mathcal{M}^{l,0}$ and $\mathcal{M}^{l,1}$ respectively), treating separately all the components of the multivariate response Y^l, $l = 1, \ldots, d$.

ii. For each component of the multivariate response l, $l = 1, \ldots, d$, consider the observed data $(y_i^l, \boldsymbol{z_i}, \boldsymbol{w_i})$, with $i = 1, \ldots, n$ and perform one of the permutation tests proposed in Bonnini et al. (2011) to compare $\mathcal{M}^{l,0}$ and $\mathcal{M}^{l,1}$ (t^l say). In order to maintain the dependence due to the fact that for each test the responses come from the same n subjects on the multivariate response, synchronized permutations have to be performed on the several tests.

iii. Consider the d separated tests t^l, $l = 1, \ldots, d$ and combine them in the global test t to test the global null hypothesis of interest, using the nonparametric combination of dependent permutation tests.

iv. If the global test t is significant, correct the partial tests t^l, $l = 1, \ldots, d$, for multiplicity, using a closed testing nonparametric procedure (see Pesarin (2001) and Pesarin and Salmaso (2010)), obtaining the adjusted p-values p^l_{adj}, $l = 1, \ldots, d$, and conclude that the tested covariates influence the response on those aspects where the adjusted p-values p^l_{adj} are significant. If the global test t is not significant, conclude that the tested covariates do not influence the multivariate response.

We remark that the permutations tests t^l, $l = 1, \ldots, d$, can be chosen among the ones proposed for the univariate case. Moreover measures of the significance of specific domain related models can be also tested.

5.3 Power behavior

A Monte Carlo simulation study has been carried out in order to check the performances of the proposed multivariate method. In Chapter 4 a simulation study was presented for the univariate permutation test. The study shows that the method is both reliable under the null and powerful under the alternative hypothesis.

5.3.1 How to simulate the multivariate response

We must simulate multivariate responses. A multivariate version of the CUB model has not been defined yet in the literature, hence a first problem we had to deal with when we started the simulation study was how to simulate the data from a multivariate CUB model. We started from the fact that the permutation approach we propose to test for covariates significance on a multivariate response works performing partial tests on the single aspects of the response variable, in a way as they had marginal CUB distributions. Hence, as a first step in the simulation mechanism, we decided to simulate the data in a way such that the several components had marginally a CUB distribution. Moreover we wanted to be able to induce some correlation structure among the several components of the multivariate response; a common instrument used to describe the dependence between random variables is the copula (see Nelsen (2006)). The basic idea is to apply the probability integral transform to the single components and then specify the dependence among the resulting uniform random variables, instead of among the original ones. Formally, let us assume we deal with a multivariate random vector $\boldsymbol{Y} = (Y^1, \ldots, Y^d)$. Therefore we can define $\mathcal{C} : [0,1]^d \to [0,1]$ as a d-dimensional copula if $\mathcal{C}$ is a joint cumulative distribution function of a d-dimensional random vector on the unit cube $[0,1]^d$ with uniform marginals. Let $F_1, \ldots, F_d$ be the marginal cumulative distribution functions of the components of $\boldsymbol{Y}$; the idea is that by applying probability integral transform to each component of $\boldsymbol{Y}$, the random vector

$$\left(U^1, \ldots, U^d\right) = \left(F_1(Y^1), \ldots, F_d(Y^d)\right)$$

has uniform margins and it takes values on the unit cube $[0,1]^d$, hence it is a copula. At this point one can define the copula of $(Y^1, \ldots, Y^d)$ as the joint cumulative distribution function of $(U^1, \ldots, U^d)$, $\mathcal{C}(u^1, \ldots, u^d)$ say. Hence by specifying $\mathcal{C}$ we can model the dependence structure between the components of $\boldsymbol{Y}$, while the marginal cumulative distribution functions $F_1, \ldots, F_d$ contain all information on the marginal distributions.

Our idea is to use this concept to simulate a multivariate response with CUB marginals and a specified dependence structure. Indeed, the other way around, it is also possible to write

$$(Y^1, \ldots, Y^d) = \left(F_1^{-1}(U^1), \ldots, F_d^{-1}(U^d)\right),$$

which can be used to simulate from $(Y^1, \ldots, Y^d)$ in copula models. However copulas theory does not share the same results for continuous and discrete data; all the results holding in the continuous case follow from Sklar's theorem (see Sklar (1959)): it assures that, given the joint cumulative distribution function of a continuous random vector, this can be written as a function of the marginal cumulative distribution functions, and that such function (called *copula*) is unique. Formally the theorem says that in the continuous case the following representation (Sklar's representation):

$$Pr\left(Y^1 \leq y^1, \ldots, Y^d \leq y^d\right) = H\left(y^1, \ldots, y^d\right) = \mathcal{C}\left(F_1(y^1), \ldots, F_d(y^d)\right)$$

is unique given H, in the sense that there is only one possible specification for $\mathcal{C}$. When we pass to the discrete case in particular identifiability issues arise, due to the lack of uniqueness of Sklar's representation in this case. Copulas models for discrete data are anyway valid constructions and, as suggested by Genest and Neslehová (2007), they are helpful in the context of simulation; the identifiability problem, indeed concerns the estimation field and not the simulation one.

All we need is to be able to simulate from a copula and to invert the cumulative distribution function for a CUB model. Consider the general case of a multivariate response of dimension d. The simulating procedure works according to the following two steps:

1) simulate a sample from a multivariate copula, according to a pre-specified correlation structure, getting $\boldsymbol{u_i} = \left(u_i^1, \ldots, u_i^d\right)$, with $i = 1, \ldots, n$;

2) consider each multivariate element of the simulated sample, $\boldsymbol{u_i}$, and transform it into the final element through the inverse cumulative distribution function of a CUB model, i.e. $\boldsymbol{y_i} = \left(y_i^1, \ldots, y_i^d\right) = \left(F_1^{-1}(u_i^1), \ldots, F_d^{-1}(u_i^d)\right)$.

We remark that at any combination (i, l), with $i = 1, \ldots, n$ and $l = 1, \ldots, d$ a specific CUB model has to be considered while inverting the cumulative distribution function $F_l(y)$, according to the parameters values $\boldsymbol{\beta}^l$ and $\boldsymbol{\gamma}^l$ chosen for that component of the multivariate response and the values taken for the covariates of interest $(\boldsymbol{x_i}, \boldsymbol{w_i})$ by the ith subject. In general the cumulative distribution function of a CUB model, for the ith subject and the lth component of the multivariate response, with uncertainty and feeling parameters $\pi_i^l = 1/\left(1 + e^{-\boldsymbol{z_i}\boldsymbol{\beta}^l}\right)$ and $\xi_i^l = 1/\left(1 + e^{-\boldsymbol{w_i}\boldsymbol{\gamma}^l}\right)$ can be derived as follows:

$$\begin{aligned}
F_l(y) = \Pr\left(Y \leq y | \boldsymbol{z_i}, \boldsymbol{w_i}\right) &= \sum_{h=1}^{y} \left[\pi_i^l \binom{m-1}{h-1} (1-\xi_i^l)^{h-1} {\xi_i^l}^{m-h} + (1-\pi_i^l)\left(\frac{1}{m}\right)\right] \\
&= \pi_i^l \sum_{h=1}^{y} \left[\binom{m-1}{h-1} (1-\xi_i^l)^{h-1} {\xi_i^l}^{m-h}\right] + (1-\pi_i^l)\left(\frac{y}{m}\right) \\
&= \pi_i^l \sum_{h=1}^{y} \left[\binom{m-1}{h-1} (1-\xi_i^l)^{h-1} {\xi_i^l}^{((m-1)-(h-1))}\right] + (1-\pi_i^l)\left(\frac{y}{m}\right) \\
&= \pi_i^l \sum_{h=0}^{y-1} \left[\binom{m-1}{h} (1-\xi_i^l)^{h} {\xi_i^l}^{((m-1)-h)}\right] + (1-\pi_i^l)\left(\frac{y}{m}\right).
\end{aligned}$$

We can recognize that $F_l(y)$ is actually a mixture of two cumulative distribution functions: the one of a binomial distribution $Bi(m-1,(1-\xi_i^l))$ calculated in $y-1$ and the one of a discrete uniform distribution. At this point the inversion of such function becomes straightforward: we build an algorithm that uses the quantile functions of the cumulative distribution functions and takes care of the fact that we are dealing with a mixture of such functions, then we can numerically invert such function.

We used an Archimedean copula (see Nelsen (2006), p. 116), mainly because it allows to model the dependence with only one parameter. The use of this copula family with discrete data can be found in the literature, for instance in Pfeifer and Neslehová (2004). A copula is said to be Archimedean if its joint cumulative distribution function can be represented as

$$\mathcal{C}\left(u^1,\ldots,u^d\right)=\phi\left(\phi^{-1}(u^1)+\cdots+\phi^{-1}(u^d)\right),$$

where ϕ is called generator. In particular, following an example that can be found in Pfeifer and Neslehová (2004), we chose the so-called Frank's copula, which is determined by the generator $\phi(t)=-\left[\log(1-(1-exp(-\theta))exp(-t))\right]/\theta$, where θ is the parameter regulating the dependence induced among the d components of the multivariate response. We remark that this is just one possible choice and that in principle other copulas could be used.

A last remark refers to the general term *dependence* that has been used while talking about what copulas models for discrete data can handle. This choice has been done since a more specific definition for this dependence cannot be given in our case. In the discrete case the copula alone cannot characterize the dependence between the several components of the final multivariate vector (see Genest and Neslehová (2007) for a more in depth discussion): indeed concordance measures are margin-dependent in the discrete case, so for instance the classical correlation coefficients calculated on couples of components of the multivariate vector can vary according to the marginal models. One helpful property anyway keeps on holding in our case, i.e. increasing the value of the dependence parameter θ of the copula, also the dependence induced in the final discrete responses increases. Hence, in this simulation study, we will disregard which kind of dependence we will have in the final multivariate discrete vector, taking into consideration only the fact that the setting differing in terms of θ differ somehow in terms of simulated dependence within the multivariate response.

5.3.2 Settings and results

In this simulation study we considered similar settings as the ones used in Chapter 4. Working in terms of rejection rates, the reliability of the method under the null hypothesis has been checked, and its power under alternative hypotheses has been studied. The aim of this study is to confirm the coherence of the tests performances in the univariate and multivariate cases; we considered only two permutation solutions (hereafter t_{lrt}, t_{wald}), as a parametric counterpart is not available for the multivariate problem of interest. We considered only the CUB $(0,1)$ model and not the CUB $(1,0)$ and CUB $(1,1)$ under the alternative hypothesis, as in Chapter 4 the power behavior in the latter two cases has already been studied and, once proved the coherence of the tests performances in the

univariate and multivariate cases for the CUB $(0,1)$, we can extend the results to the general CUB (p,q) model; this is essentially due to the fact that the power of the global permutation solution depends on the power behavior of the partial tests. We considered the case of dichotomous covariates. We remark that the CUB $(0,1)$ scenario with a dichotomous covariate, besides being a very common and simple structure, is also helpful to interpret the feeling parameter γ, which in this case can be seen as a measure of the difference in feeling between the two sub-populations identified by the covariate (see Chapter 4).

As regards the simulation settings, since the increase of the power functions, in the univariate case, as the data move away from the null hypothesis has already been shown in Chapter 4, we decided to choose a fixed value for the "distance" between the values of the parameters ξ for the two populations determined by the covariate (hereafter δ_ξ), and instead change the number of components of the multivariate response which we simulate under the alternative hypothesis, in order to see if the estimated rejection probabilities increase as we increase it (in a way in fact this can be seen as the data was moving away from the global null hypothesis).

Moreover several values for the dependence parameter θ have been considered in order to check the behavior of the power functions for different levels of dependence among the components of the multivariate response. As already underlined above, the dependence simulated in the copula is then not transmitted in the exact same way to the final simulated responses (see Pfeifer and Neslehová (2004) for a more in depth discussion). The general idea is that the dependence present in the final data does not depend on the underlying copula alone and it is instead influenced by the marginal distributions as well. Therefore in our particular case some correlation is surely introduced in the final responses when we simulate all the components under the alternative hypothesis, i.e. all marginally following a CUB distribution with the same significantly influent covariate. This means that we are not able to quantify the dependence among the response components; anyway, at the same time we can say that increasing the dependence parameter θ and letting fixed all the other parameters in the simulation setting, we get an increase of the dependence in the final response, no matter which is the impact, on the final correlation, of the introduction of a significant covariate.

Table 5.1 shows the considered simulation settings for $d = 2, 3$ dimensions of the multivariate response: the table must be read in terms of number of components simulated under the alternative hypothesis. Settings for 1 to 3 refer to a bivariate simulated response, while settings from 4 to 7 to the case $d = 3$. Hence in the first and in the fourth setting we are simulating the data under the global null hypothesis that the covariate does not influence any of the components. We kept fixed the value of $\delta_\xi = 0.2$. We set two values for the sample size ($n = 50, 100$) and three values for the dependence parameter ($\theta = 0, 5, 10$). The feeling parameter was set to $\xi = 0.1$ for the components simulated under the null hypothesis (CUB $(0,0)$ model), and to $\xi_{(0)} = 0.1$, $\xi_{(1)} = 0.3$, in the two sub-groups identified by the covariate, for the components simulated under the alternative hypothesis (CUB $(0,1)$ model). The uncertainty parameter was set always equal to $\pi = 0.9$ (low uncertainty situation, very often happening in real applications).

Tables 5.2 and 5.3 report the estimated rejection probabilities of the compared tests (partial adjusted and global permutation tests) on the parameter γ_1

Table 5.1: Simulation settings for $d = 2, 3$, each cell indicating under which hypothesis the specific component is simulated for the specific setting. The symbol $-$ indicates that the component is not considered in that setting.

Setting	Y^1	Y^2	Y^3
1	H_0	H_0	$-$
2	H_1	H_0	$-$
3	H_1	H_1	$-$
4	H_0	H_0	H_0
5	H_1	H_0	H_0
6	H_1	H_1	H_0
7	H_1	H_1	H_1

for $m = 7$ and working at a nominal level of $\alpha = 0.05$ respectively for $n = 50$ and $n = 100$. A number of $B = 1000$ permutations and $MC = 1000$ Monte Carlo replications have been considered. Regarding the partial adjusted tests, we can notice how their power increases under the alternative partial hypotheses as the sample size increases. Moreover, the obtained results show how the global permutation test controls the type I error when the global null hypothesis is true. Such test also turns out to be a powerful solution as one of the partial null hypothesis is not true. In general we can also notice that its power increases as the sample size increases, for all the considered settings, reaching the value one as a well performing test is expected to do. Therefore, a power increase can also be registered while increasing the number of false partial null hypotheses (hence passing from setting 2 to 3 and from setting 5 to 7). It also has to be underlined that when more than one partial null hypothesis is false (hence in settings 3, 6 and 7), the power decreases as the copula's dependence parameter θ increases, again confirming an expected behavior (see Pesarin and Salmaso (2010)).

5.4 Real case applications

5.4.1 The Passito survey

We applied the proposed multivariate solution to the Passito survey data, that we analyzed already in Chapter 4. In such previous analysis, univariate permutation tests were performed for each of the wine aspects considered in the study. In this chapter, instead, we apply the multivariate test in order to check if the final the final models obtained by the stepwise procedure in the previous analysis globally fit well the data.

We remind that the sample size was of 386 Passito consumers living in Veneto region (Italy). Five response variables, on a scale from 1 (maximum dissatisfaction) to 7 (maximum satisfaction), were considered: level of liking (y_1), satisfaction about aroma (y_2), satisfaction about the sweet taste (y_3), satisfaction about alcohol percentage (y_4), satisfaction about intensity of taste (y_5). The three dichotomous covariates were age (coded as 0 if the subject was ≤ 25 years old, and 1 other wise), gender (0 for females and 1 for males) and residence

Table 5.2: Estimated rejection probabilities for the partial adjusted permutation tests on the single components of the multivariate response ($t_{\text{lrt}}^{Y^l}$ and $t_{\text{wald}}^{Y^l}$, with $l = 1, 2, 3$) and of the global solution ($t_{\text{lrt}}^{\text{glob}}$ and $t_{\text{wald}}^{\text{glob}}$), settings 1 to 7 and $\theta = 0, 5, 10$. The results refer to sample size $n = 50$, nominal level of $\alpha = 0.05$, $B = 1000$ permutations and $CMC = 1000$ replications. Estimates in bold indicate quantities which should be smaller than α, as cases under the (partial or global) null hypotheses.

Setting	$\theta = 0$				$\theta = 5$				$\theta = 10$			
	t_{lrt}		t_{wald}		t_{lrt}		t_{wald}		t_{lrt}		t_{wald}	
	$t_{\text{lrt}}^{Y^{1,2,3}}$	$t_{\text{lrt}}^{\text{glob}}$	$t_{\text{wald}}^{Y^{1,2,3}}$	$t_{\text{wald}}^{\text{glob}}$	$t_{\text{lrt}}^{Y^{1,2,3}}$	$t_{\text{lrt}}^{\text{glob}}$	$t_{\text{wald}}^{Y^{1,2,3}}$	$t_{\text{wald}}^{\text{glob}}$	$t_{\text{lrt}}^{Y^{1,2,3}}$	$t_{\text{lrt}}^{\text{glob}}$	$t_{\text{wald}}^{Y^{1,2,3}}$	$t_{\text{wald}}^{\text{glob}}$
1	**0.032** **0.031**	**0.057**	**0.027** **0.031**	**0.058**	**0.035** **0.032**	**0.057**	**0.035** **0.029**	**0.056**	**0.029** **0.030**	**0.048**	**0.031** **0.035**	**0.048**
2	0.923 **0.052**	0.923	0.912 **0.052**	0.923	0.918 **0.050**	0.921	0.916 **0.025**	0.921	0.914 **0.046**	0.919	0.905 **0.050**	0.917
3	0.949 0.956	0.995	0.946 0.948	0.995	0.947 0.947	0.984	0.936 0.936	0.985	0.941 0.944	0.973	0.934 0.938	0.974
4	**0.024** **0.015** **0.020**	**0.056**	**0.027** **0.015** **0.019**	**0.060**	**0.016** **0.022** **0.021**	**0.050**	**0.016** **0.014** **0.023**	**0.053**	**0.019** **0.025** **0.031**	**0.057**	**0.019** **0.023** **0.027**	**0.057**
5	0.905 **0.026** **0.025**	0.907	0.889 **0.026** **0.024**	0.910	0.918 **0.030** **0.025**	0.923	0.907 **0.025** **0.026**	0.924	0.910 **0.032** **0.044**	0.918	0.887 **0.026** **0.038**	0.917
6	0.923 0.915 **0.041**	0.994	0.913 0.903 **0.045**	0.994	0.935 0.923 **0.040**	0.980	0.929 0.916 **0.045**	0.980	0.928 0.924 **0.056**	0.978	0.917 0.903 **0.050**	0.978
7	0.948 0.939 0.950	0.999	0.942 0.926 0.948	0.999	0.956 0.953 0.955	0.991	0.949 0.937 0.950	0.991	0.940 0.945 0.933	0.980	0.930 0.933 0.929	0.979

Table 5.3: Estimated rejection probabilities for the partial adjusted permutation tests on the single components of the multivariate response ($t_{\text{lrt}}^{Y^l}$ and $t_{\text{wald}}^{Y^l}$, with $l = 1, 2, 3$) and of the global solution ($t_{\text{lrt}}^{\text{glob}}$ and $t_{\text{wald}}^{\text{glob}}$), settings 1 to 7 and $\theta = 0, 5, 10$. The results refer to sample size $n = 100$, nominal level of $\alpha = 0.05$, $B = 1000$ permutations and $CMC = 1000$ replications. Estimates in bold indicate quantities which should be smaller than α, as cases under the (partial or global) null hypotheses.

Setting	$\theta = 0$				$\theta = 5$				$\theta = 10$			
	t_{lrt}		t_{wald}		t_{lrt}		t_{wald}		t_{lrt}		t_{wald}	
	$t_{\text{lrt}}^{Y^{1,2,3}}$	$t_{\text{lrt}}^{\text{glob}}$	$t_{\text{wald}}^{Y^{1,2,3}}$	$t_{\text{wald}}^{\text{glob}}$	$t_{\text{lrt}}^{Y^{1,2,3}}$	$t_{\text{lrt}}^{\text{glob}}$	$t_{\text{wald}}^{Y^{1,2,3}}$	$t_{\text{wald}}^{\text{glob}}$	$t_{\text{lrt}}^{Y^{1,2,3}}$	$t_{\text{lrt}}^{\text{glob}}$	$t_{\text{wald}}^{Y^{1,2,3}}$	$t_{\text{wald}}^{\text{glob}}$
1	**0.028** **0.037**	**0.062**	**0.031** **0.039**	**0.063**	**0.026** **0.030**	**0.050**	**0.026** **0.029**	**0.051**	**0.032** **0.025**	**0.047**	**0.029** **0.027**	**0.047**
2	0.997 **0.064**	0.997	0.996 **0.062**	0.997	0.998 **0.050**	0.998	0.998 **0.048**	0.998	1.000 **0.045**	1.000	1.000 **0.043**	1.000
3	0.998 0.998	1.000	0.999 0.998	1.000	1.000 1.000	1.000	0.999 1.000	1.000	1.000 1.000	1.000	1.000 1.000	1.000
4	**0.015** **0.024** **0.028**	**0.060**	**0.013** **0.025** **0.027**	**0.060**	**0.017** **0.028** **0.021**	**0.055**	**0.016** **0.027** **0.021**	**0.053**	**0.023** **0.027** **0.019**	**0.057**	**0.020** **0.028** **0.020**	**0.058**
5	**0.997** **0.028** **0.042**	0.997	0.995 **0.028** **0.039**	0.997	0.999 **0.032** **0.028**	0.999	0.999 **0.035** **0.031**	0.999	0.998 **0.035** **0.035**	0.998	0.998 **0.036** **0.032**	0.998
6	0.997 0.999 **0.060**	1.000	0.998 0.999 **0.059**	1.000	1.000 1.000 **0.052**	1.000	1.000 1.000 **0.050**	1.000	0.998 0.997 **0.054**	0.999	0.999 0.997 **0.053**	0.999
7	1.000 0.999 0.999	1.000	1.000 0.999 0.999	1.000	1.000 1.000 1.000	1.000	1.000 1.000 0.999	1.000	1.000 1.000 1.000	1.000	1.000 1.000 1.000	1.000

(coded as 0 for East Veneto, and 1 for West Veneto). The final models obtained in Chapter 4 suggest that all the covariates globally influence the multivariate response. We want to check such result by means of our multivariate test: thus the questions we want to answer are a) if the three covariates globally influence the multivariate response and b) which responses are actually affected by which covariate(s) while taking care of the multiplicity issue.

We applied the multivariate permutation solution, implementing the t_{wald} test at the first step of the procedure, i.e. performing the partial permutation tests on the univariate responses. The reason for choosing t_{wald} instead of t_{lrt} is that afterwards that solutions will turn useful to answer to the second question we want to answer, i.e. which responses are actually affected by which covariate(s). For the same reason also t_{npc} would have been a suitable choice. In practice the analysis was conduced as follows:

- we performed the global permutation test for the significance of the 3 covariates on the multivariate response;
- since this produced a significant result, we corrected the partial tests on the single responses, taking care of the multiplicity issue;
- since all the partial tests suggested the influence of the three covariates on the single responses, we corrected the sub-partial Wald type tests related to the single response/covariate combinations, taking care of the multiplicity issue.

The results are showed in Table 5.4, where we report the adjusted p-values of the sub-partial Wald type tests for each response/covariate combination, the adjusted p-values of the partial tests on the single responses, and the global p-value for the multivariate test. The analysis confirms almost all the results obtained in Chapter 4. The global test suggests a significant effect of the three covariates on the multivariate response. The partial adjusted p-values on the single responses tell us that the rating of all the considered items is globally affected by the three covariates. In the central cells of the table we reported only the significant adjusted p-values of the sub-partial Wald type tests for each response/covariate combination. It is worth noticing that such results suggest final models, obtained taking care of the multiple nature of the response, that coincide for almost all the items with the ones obtained by the stepwise procedure proposed in Chapter 4.

5.4.2 The S.E.S.T.O. survey

The S.E.S.T.O. (Statistical Evaluation of a Skischool from Tourists' Opinions) survey is the first Italian survey on the evaluation (by parents) of ski courses for young children (up to 14 years old) and it is a pilot study performed in the Ski School of Sesto, in the Dolomites near Bolzano in the north of Italy. Several customer satisfaction variables towards different aspects of ski teaching have been evaluated in a rating scale 1-10. A multivariate response has been considered in the study, related to five aspects of the customer satisfaction: "Easy Learning", "Helpful Teacher", "Fun", "Involvement" and "General Satisfaction". Moreover the dichotomous covariate "First presence in Sesto" for parameter ξ has been included in the analysis, to verify if the families who were in Sesto for the first

Table 5.4: First real case application results. Partial adjusted p-values of the sub-partial Wald type tests for each response/covariate combination and of the partial tests on the single responses. Global p-value to test the joint effect of the three covariates on the multivariate response.

Response	Age - p-value for		Gender - p-value for		Residence - p-value for		Global p-value
	β_1	γ_1	β_2	γ_2	β_3	γ_3	
Level of liking (p-value = 0.008)	– – – – –	– – – – –	– – – – –	0.001	– – – – –	– – – – –	0.001
Satisfaction about aroma (p-value = 0.008)	– – – – –	– – – – –	– – – – –	0.044	– – – – –	– – – – –	
Satisfaction about the sweet taste (p-value = 0.001)	– – – – –	0.036	– – – – –	0.010	0.001	– – – – –	
Satisfaction about the alcohol percentage (p-value = 0.008)	– – – – –	– – – – –	– – – – –	0.012	– – – – –	– – – – –	
Satisfaction about the intensity of the taste (p-value = 0.080)	– – – – –	– – – – –	0.001	– – – – –	– – – – –	– – – – –	

time presented a different feeling toward the ski courses than the others. The sample size is $n = 96$ children.

The global p-value 0.001 leads to the rejection of the global null hypothesis at $\alpha = 0.05$ hence the tested covariate affects the feeling. Table 5.5 reports the adjusted p-values of the partial permutation tests on the single components of the response, together with the maximum likelihood estimates, for the univariate CUB models, of the γ parameters and the feeling parameters in the two groups identified by the covariate ($\hat{\xi}_{(0)}$ and $\hat{\xi}_{(1)}$). According to the adjusted p-values of the partial tests, to be in Sesto for the first time has no influence on the easy of learning but it positively affects the feeling of the respondents towards the helpfulness of the teacher (adjusted p-value equal to 0.009), the fun of the children (adjusted p-value equal to 0.013), the involvement of the children (adjusted p-value equal to 0.015) and also the general satisfaction (adjusted p-value equal to 0.001).

To conclude, one could point out that the response variable "general satisfaction" is probably inherently correlated to the other variables. As a last remark it is worth underlining that this fact anyway does not affect the results we obtained. Indeed the dependence among the responses, and thus among the permutation tests performed on each response, is implicitly taken into account by our global procedure.

5.5 Main results

In this chapter we presented an extension of a permutation solution to test for covariate influence on an ordinal response, working within the CUB models framework. Such extension allows us to deal with a multivariate response, which is a very common output while analyzing this kind of data. The method basically works implementing the permutation solution proposed in Chapter 4 separately on each component of the multivariate response, anyway taking into account the

Table 5.5: Second real case application results. Feeling parameters estimates for the several univariate models and partial adjusted permutation tests on the significance of the "First presence in Sesto" covariate on the the components of the multivariate response.

Response	First presence in Sesto, feeling parameters estimates	Part. adj. p-values
Easy Learning	$\hat{\gamma}_0 = -2.16244,\ \hat{\gamma}_1 = -0.44946$ $\Updownarrow$ $\hat{\xi}_{(0)} = 0.103\ \hat{\xi}_{(1)} = 0.068$	0.152
Helpful Teacher	$\gamma_0 = -2.79602,\ \hat{\gamma}_1 = -1.66700$ $\Updownarrow$ $\hat{\xi}_{(0)} = 0.058\ \hat{\xi}_{(1)} = 0.011$	0.009
Fun	$\hat{\gamma}_0 = -2.53541,\ \hat{\gamma}_1 = -1.51638$ $\Updownarrow$ $\hat{\xi}_{(0)} = 0.073\ \hat{\xi}_{(1)} = 0.017$	0.013
Involvement	$\hat{\gamma}_0 = -2.32239,\ \hat{\gamma}_1 = -0.9357$ $\Updownarrow$ $\hat{\xi}_{(0)} = 0.089\ \hat{\xi}_{(1)} = 0.037$	0.015
General Satisfaction	$\hat{\gamma}_0 = -2.29823,\ \hat{\gamma}_1 = -0.95986$ $\Updownarrow$ $\hat{\xi}_{(0)} = 0.091\ \hat{\xi}_{(1)} = 0.037$	0.001

dependence among such components by performing synchronized permutations and then nonparametrically combining the partial tests.

The method's performances have been checked through a simulation study were the cases of $C = 2$ and $C = 3$ dimensions of the multivariate response have been considered. Several settings have been explored, which differ from each other in terms of number of partial components under the alternative hypothesis. The results have shown the very good behavior of the global permutation solution, which is reliable under the global null hypothesis and powerful under the alternative especially for low sample sizes. Its power increases (reaching the value one) as we increase the sample size and as the dependence among the components of the multivariate outcome decreases. Moreover a power increase can be observed while increasing the number of false partial null hypothesis.

The permutation test has also been applied to two real datasets. The obtained results show the usefulness of the method and its coherence with its univariate counterpart.

We can conclude that the proposed permutation solution is useful in order to test for the influence of covariates on a multivariate ordinal response, while working in the CUB models framework.

Chapter 6

Conclusions

6.1 Original results

This thesis presents some innovative results about the use of permutation methods in real applications. In particular two main real contexts have been considered in this work: testing for difference in treatments' effects while performing single-case experiments and studying covariates influence on ordinal responses within a regression framework. The two application fields are quite different, and quite far away from each other are the proposed solutions to the problems, even if they are both belonging to the permutation approach. The difference between the presented approaches is due to the substantial difference in the structures of the data coming from the two applications. In the case of single-case experiments we do not deal with a classical random sample from some population of interest, as the recorded values come instead from one single subject which is observed during a certain period of time. Hence we handle a vector of values of length n which in practice can be referred to as a time-series. Within this time-series we can distinguish among values coming from different treatments, the effects of which we wish to compare at a single-subject (and not at a wider population) level. In the second application instead the data comes from a sample of respondents, making us handle the usual *i.i.d.* sample of size n from a population of interest. Here again we aim to make some comparisons, but in this case such comparisons are referred to some sub-populations which are identified by the different values that the related covariates can take. Due to this fundamental difference the permutation solutions we proposed for the two problems are quite different regarding the permutation mechanism. Permutation tests are constructed permuting the quantity which is exchangeable in the data under the null hypothesis, this way estimating the null distribution of some test statistic and comparing to it the observed value of such statistic. Hence the permutation strategy can be different from case to case, according to which is the exchangeable quantity in the problem of interest. In our two applications the exchangeable quantities are different, since in single-case data we permute the treatments sequence on time, within the single subject, whereas in the regression problem for ordinal data we permute the rows of tested covariates among subjects. Hence one main contribute on this thesis is that we showed how permutation techniques can be a flexible instrument to deal with

very different real problems and data scenarios.

More specifically important results have been reached in both the real application issues. Regarding single-case experiments we proposed a permutation procedure which is able to solve the problem with good results in a wide range of possible scenarios. We improved an already existing permutation procedure, exploiting the implementation of a MA solution: this way the method is able to catch several effects' difference besides the elementary case in which the treatments' effects are characterized by a stationary processes and they differ only in the average level. The case of possible trends in the treatments' effects is indeed considered. The method is innovative also in the sense that it proposes a joint use of time-series smoothing techniques and the permutation strategy: this idea is proposed in this thesis and, by means of a simulation study, it shows to give competitive results. We think that the strength of this mixed procedure comes from the its ability of wasting much less information about the shape of the time-series on time than other (also permutation) methods using classical summary statistics of the data (sample mean, median, slope and intercept estimates for trends, etc.).

Another innovative contribute is the possibility of making inference on single-case experiments when a multivariate response is observed, for which no global inferential errors controlling procedure is currently available in the literature. Using the nonparametric combination of dependent permutation tests, a solution is proposed which allows to assert a difference of treatments' effects globally on several outcomes of interest, while controlling for the multiplicity issue. The reader can understand how this results has quite some importance thinking about the fact that multivariate responses are much more likely in clinical studies than univariate outcomes.

This work proposes also a solution to extend the results obtained in single-case experiments to a wider population of interest. Also on this aspect, no procedure which takes care of the multiplicity issue is currently available in the literature. The problem is quite hard to solve because multiple time-series have to be analyzed at the same time. Again by nonparametrically combining the univariate permutation tests we are able to solve the problem. The idea of a two step procedure, which takes care of the way the experiment is conducted and the data are recorded, is an innovative proposal that allows the analyst to end up with a global result at a population level but also to recover single-subjects related information, while taking care of the multiplicity issue. We obtain in this way a double advantage, as we can make both individual and global conclusions.

Regarding the regression problem dealing with ordinal responses, this thesis contributes to give more applicability space to a very interesting class of regression models (CUB models), which is getting growing interest from statistical research papers in the last years. We believe that CUB models are a very handy solution to analyze ordinal data: they are flexible and allow a nice interpretation of the phenomenon under observation. As low sample size is a common situation while analyzing real data, we think that a well performing alternative to the likelihood based solution is an important contribute for the application of this kind of model. The permutation test we propose shows a good behavior when the sample size is not high, which suggests that its use should be recommended in real applications when less than 150 respondents are included in the study. Therefore an innovative proposal can be found in the stepwise procedure we present, which is able to make a covariate selection while controlling for the

multiplicity issue at the several steps.

At the end, this work proposes a global solution for the not yet solved problem of dealing with an ordinal multivariate response, which is a very common situation in real applications. In this sense the permutation strategy turns out to be an interesting and innovative alternative to parametric inference not only for low sample sizes, but in general for analyzing survey data.

6.2 Further developments

Some further developments can be surely of interest for both the main topics in this thesis. The use of nonparametric combination to construct MA solutions and to extend the proposal to multivariate tools brings with it a wide range of possible combinations of test statistics and combining function that could be adopted. Also the possibility of constructing combination procedures that work in more levels, bring us to the use of specific decisions and leave a lot of space for other possible alternatives. Thus a lot can be discussed about possible further developments of the solutions we proposed and of the study of their performances.

Regarding the single-case experiment problem many variable pieces are composing the puzzle: first of all the local regression as smoother at the starting step of the procedure is just our choice and, as already mentioned in Chapter 2, other possible techniques could be chosen. Also the test statistics composing the MA solution are our particular choices and other solutions are of course possible, both in terms of different statistics and of number of partial tests which are combined together. Also in the extended solution for multivariate replicated single-case experiments specific choices have been made while building the multivariate test. We proposed a two steps procedure which works on first level global tests on the subjects and then it combines them in a second level final global test, but the definition of the two levels could be changed and different partial steps could be proposed. In this sense, as we already pointed out previously, we believe that further developments are possible starting from the method we proposed in the second and third chapter of this thesis. Thus it may be of interest to study performances of other smoothers for the time subprocesses, or even simply change the choice of smoothing parameters for the local regression solution. Again, the role played by the use of different partial test statistics in the MA solution could also be studied. Further developments could be of interest regarding the data scenarios on which the method performances have been explored. Regarding the extended procedure for multivariate replicated experiments, we could explore how changing the combining functions (that are used at the different levels of the combination scheme) the procedure performances change.

Our proposal to test for covariates effect in CUB models is also characterized by specific choices, which could be changed. At first the choice of the test statistics could be changed: other measures of the goodness of fit of the model could be studied, as, for instance, some kind of residuals. MA solutions could also be tried, which combine together the partial tests we studied in this work. Therefore the combining function used in the proposed univariate test to combine the Wald type partial statistics on the single parameters could also be changed in order to check if better behaviors can be achieved. The choice of the combining

functions could be studied more in depth also for the multivariate procedure proposed in Chapter 5. Finally it could be of interest to check which can be the reason that makes the power of all the procedures (also the parametric test) decrease when we have covariates affecting the uncertainty parameter of the model.

Concluding we hope that this thesis can be a starting point for future new research on the topics faced here. We also wish that this work can be an inspiration and a motivation for other researchers to consider the possibility of using permutation techniques in order to solve real complex problems.

Appendix A

R codes

A.1 R codes for single-case experiments

This section of the appendix describes the R codes needed to perform the permutation solutions to test for difference in treatments' effect in univariate as well as multivariate replicated single-case experiments. These codes are available upon request. They are contained in the files listed below:

- `assignments.txt`, `quantity.txt`, `selectdesign.txt`: functions related to single-case experiments, needed to perform the randomization test. See Bulté and Onghena (2009) for a more detailed description;
- `combine.txt`, `FWEminP.r`, `t2p.r`: functions related to the nonparametric combination of randomization tests. See Basso et al. (2010) for a more detailed description;
- `perm_test_MA1.R`, `perm_test_MA1.R`, `perm_test_SP_MA1.R`, `perm_test_SP_MA1.R`: functions performing the two Multi Aspect randomization tests. The first two functions perform the univariate solution, while the second two implement the general procedure for multivariate replicated single-case experiments.
- `toy_example_singlecase.R`: R script containing the code to simulate data and perform the randomization tests according to a toy example described below.

The file `toy_example_singlecase.R` contains a code to simulate a toy example and see how the R functions work. The toy example considers a replicated single-case study with a multivariate response, with $S = 8$ subjects and $p = 4$ dimensions of the multivariate response, and it refers to a particular case considered in the simulation study presented in Chapter 3. The length of the time-series is set to $n = 50$ and, as underlying ARMA model, an $ARMA(1, 1)$ with autoregressive and moving-average parameters $\phi, \theta = 0.5$ was chosen. In the first part of `toy_example_cub.R`, data are generated under the alternative hypothesis with a difference in levels between the treatment effects, following a similar scheme as the one used in the simulation study. We maintained fixed differences in levels between the treatment effects (under the alternative hypothesis), always working with a difference of $\delta = 3$, and we simulated the data

under the alternative hypothesis for only some dimensions of the multivariate response and some subjects. The setting is described in Table A.1 in terms of subjects and variables simulated under the alternative hypothesis.

Table A.1: Toy example; setting of simulated data, the symbol ✓ indicates the partial comparisons which have been performed under the alternative hypothesis. X_i^s, $i = 1, \ldots, 4$, $s = 1, \ldots, 8$ indicates the ith component of the multivariate response for the sth subject.

Subject	*Var. under H_1 (for each subj.)*	Responses, $s = 1, \ldots, 8$			
		X_1^s	X_2^s	X_3^s	X_4^s
1	1	✓			
2	2	✓	✓		
3	3	✓	✓	✓	
4	4	✓	✓	✓	✓
5	0				
6	1				✓
7	2			✓	✓
8	3		✓	✓	✓

As regards the correlational structure of the errors, we considered the case of no correlation among the subjects and correlation between the pairs of variables within each single-case experiment set at $\rho_{i,j} = 0.3$ for $i < j$, $i, j = 1, \ldots, S$. Normal errors were used to simulate the data. In the second part of the file the multivariate permutations test is applied on the data providing the final global solution. The code to obtain partial adjusted results is also available.

A.2 *R* codes for CUB models

This section of the appendix describes the `R` codes needed to perform the permutation solutions to test for covariates' effect in both univariate and multivariate CUB models. Again these codes are available upon request and they are contained in the files listed below:

- `cubR_perm.R`, `perm_test_functions_cub.r`: functions related to CUB models, needed to perform the randomization test. In particular `cubR_perm.R` modifies the code proposed by Iannario and Piccolo (2009) for the implementation of CUB models, in a way such that it can be recalled by the function performing the permutation test. `perm_test_functions_cub.r` contains the functions for simulate data from a univariate or multivariate cub model;
- `combine.txt`, `FWEminP.r`, `t2p.r`: functions related to the nonparametric combination of randomization tests. See Basso et al. (2010) for a more detailed description;
- `perm_test_blocks.R`: function performing the univariate permutation tests;
- `toy_example_cub.R`: `R` script containing the code to simulate data and perform the permutation tests according to a toy example described below.

The code `toy_example_cub.R` simulates a toy example and can be uses to see how the `R` functions work. The toy example considers the case of a multivariate ordinal response $\boldsymbol{Y}$, with $c = 3$ outcomes, and it refers to a particular case considered in the simulation study presented in Chapter 5 (setting 6). The sample size is set to $n = 50$, the number of values that the response can take to $m = 7$ and the value for the dependence parameter of the Frank copula to $\theta = 5$. The feeling parameter is set to $\xi = 0.1$ for the third component, simulated under the null hypothesis (CUB $(0, 0)$ model), and to $\xi_{(0)} = 0.1$, $\xi_{(1)} = 0.3$, in the two sub-groups identified by the covariate, for the first and second components, simulated under the alternative hypothesis (CUB $(0, 1)$ model). The uncertainty parameter is set always to $\pi = 0.9$. In the first part of `toy_example_cub.R`, data are generated according to the scheme described above. In the second part of the file the permutations tests are applied on the single components of the multivariate response and they are then combined together to obtain the final global solution as well as the partial adjusted tests.

Bibliography

N. Agay, E. Yechiam, Z. Carmel, and Y. Levkovitz. Non-specific effects of methylphenidate (Ritalin) on cognitive ability and decision-making of ADHD and healthy adults. *Psychopharmacology*, 210:511–519, 2010.

A. Agresti. Modeling ordered categorical data: recent advances and future challenges. *Statistics in Medicine*, 18:2191–2207, 1999.

A. Agresti. *Categorical Data Analysis*. J. Wiley & Sons, New York, 2002.

A. Agresti. *Analysis of Ordinal Categorical Data*. J. Wiley & Sons, Hoboken, NJ, 2010.

A. Agresti and R. Natarajan. Modeling clustered ordered categorical data: a survey. *International Statistical Review*, 69:345–371, 2001.

M. Anderson and C. Ter Braak. Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation*, 73(2):79–86, 2003.

R. Arboretti Giancristofaro, V. Boatto, S. Bonnini, L. Corain, L. Salmaso, T. Tempesta, and D. Vecchiato. Preferences and consumption habits of wine consumers: a survey on Passito's Italian wine. *Food Quality and Preference*, 2011. To appear.

A. Baplu. Ritalin, the smart drug: a double-blind placebo controlled N-of-1 experiment in an adolescent with concentration problems. *Unpublished masterś thesis, Katholieke Universiteit Leuven, Belgium*, 2005.

D. Basso, M. Chiarandini, and L. Salmaso. Synchronized permutation tests in replicated $I \times J$ designs. *Journal of Statistical Planning and Inference*, 137(8): 2564–2578, 2007.

D. Basso, F. Pesarin, L. Salmaso, and A. Solari. *Permutation Tests for Stochastic Ordering and ANOVA: Theory and Applications with R*. Springer, New York, 2010.

S. Bonnini, D. Piccolo, L. Salmaso, and F. Solmi. Permutation inference for a class of mixture models. *Communication in Statistics: Theory and Methods*, 2011. Forthcoming.

I. Bulté and P. Onghena. An R package for single-case randomization tests. *Behavior Research Methods*, 40:467–478, 2008.

I. Bulté and P. Onghena. Randomization tests for multiple-baseline designs: an extension of the SCRT-R package. *Behavior Research Methods*, 41:477–485, 2009.

M. Corduas, M. Iannario, and D. Piccolo. A class of statistical models for evaluating services and performances. In M. Bini, P. Monari, D. Piccolo, and L. Salmaso, editors, *Statistical methods for the evaluation of educational services and quality of products.* Contribution to Statistics, pages 99–117. Springer, New York, 2010.

A. Crutchley and J. A. Temlett. Methylphenidate (Ritalin) use and abuse. *South African Medical Journal*, 89:1076–1079, 1999.

J.R. de Jong, J.W.S. Vlaeyen, P. Onghena, M.E.J.B. Goossens, M. Geilen, and H. Mulder. Fear of Movement/(Re)injury in Chronic Low Back Pain Education or Exposure In Vivo as Mediator to Fear Reduction. *Clinical Journal of Pain*, 21:9–17, 2005.

A. D'Elia and D. Piccolo. A mixture model for preference data analysis. *Computational Statistics & Data Analysis*, 49(8):917–934, 2005. Also available at `http://www.jstatsoft.org/v27/i08`.

P. Diggle, P. Heagerty, K. Liang, and S. Zeger. *Analysis of Longitudinal Data (Second edition).* Oxford University Press, USA, New York, 2002.

E. Edgington. *Randomization Tests.* CRC Press, 1995.

E.S. Edgington. Randomized single-subject experimental designs. *Behavior Research and Therapy*, 34:567–574, 1996.

E.S. Edgington and P. Onghena. *Randomization Tests,* 4^{th} *edition.* Chapman & Hall, London, 2007.

R. A. Fisher. *The Design of Experiments.* Oliver & Boyd, Edinburgh, 1935.

C. Genest and J. Neslehová. A Primer on Copulas for Count Data. *ASTIN Bulletin*, 37:475–515, 2007.

L. Gonzalez and B. Manly. Analysis of variance by randomization with small data stes. *Environmetrics*, 9(1):53–65, 1998.

P. Good. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses.* Springer, 1994.

B.S. Gorman and D.B. Allison. Statistical alternatives for single-case designs. In R.D. Franklin, D.B. Allison, and B.S. Gorman, editors, *Design and Analysis of Single-Case Research*, pages 159–214. Lawrence Erlbaum Associates, Mahwah, NY, 1996.

M. Iannario. A class of models for ordinal variables with covariates effects. *Quaderni di Statistica*, 10:53–72, 2008a.

M. Iannario. Dummy variables in CUB models. *Statistica*, LXVIII:179–200, 2008b.

M. Iannario and D. Piccolo. A program in `R` for CUB models inference, Version 2.0. `http://www.dipstat.unina.it/CUBmodels1/`, 2009.

J.R. Levin, L.A. Marascuilo, and L.J. Hubert. N = nonparametric randomization tests. In T.S. Kratochwill, editor, *Single Subject Research: Strategies for Evaluating Change*, pages 167–196. Academic, New York, 1978.

B. Manly. *Randomization, Bootstrap and Monte Carlo Methods in Biology.* Chapman and Hall, London, 1997.

R. Marcus, E. Peritz, and K.R Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63:655–660, 1976.

P. McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*, 42:109–142, 1980.

P. McCullagh and J.A. Nelder. *Generalized Linear Models,* 2^{nd} *edition.* Chapman and Hall, London, 1989.

R.B. Nelsen. *An Introduction to Copulas,* 2^{nd} *edition.* Springer, New York, 2006.

P. Onghena. Single-case designs. *Encyclopedia of Statistics in Behavioral Science*, 4:1850–1854, 2005.

P. Onghena and E. S. Edgington. Customization of pain treatments: single-case design and analysis. *Clinical Journal of Pain*, 21:56–68, 2005.

F. Pesarin. *Multivariate Permutation Tests with Applications in Biostatistics.* Wiley, Chichester, 2001.

F. Pesarin and L. Salmaso. *Permutation Tests for Complex Data: Theory, Applications and Software.* Wiley, Chichester, 2010.

F. Pesarin and L. Salmaso. A new characterization of weak consistency of permutation tests. *Journal of Statistical Planning and Inference*, 2011. Submitted.

D. Pfeifer and J. Neslehová. Modeling and generating dependent risk processes for irm and dfa. *ASTIN Bulletin*, 34:333–360, 2004.

D. Piccolo. On the moments of a mixture of uniform and shifted binomial random variables. *Quaderni di Statistica*, 5:85–104, 2003.

D. Piccolo. Observed information matrix for MUB models. *Quaderni di Statistica*, 8:33–78, 2006.

D. Piccolo and A. D'Elia. A new approach for modelling consumers' preferences. *Food Quality and Preference*, 19:247–259, 2008.

L. Salmaso. Synchronized permutation tests in replicated 2^k factorial designs. *Communications in Statistics: Theory and Methods*, 32(7):1419–1437, 2003.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

R. Tannock, R. J. Schachar, R. P. Carr, and G. D. Logan. Dose-response effects of methylphenidate on academic performance and overt behavior in hyperactive children. *Pediatrics*, 84:648–657, 1989.

M.M. ter Kuile, I. Bulté, P.T.M. Weijenborg, A. Beekman, R. Melles, and P. Onghena. Therapist-aided exposure for women with lifelong vaginismus: a replicated single-case design. *Journal of Consulting and Clinical Psychology*, 77:149–159, 2009.

B..T. Todman and P. Dugard. *Single-Case and Small-n Experimental Designs: A Practical Guide to Randomization Tests.* Lawrence Erlbaum Associates, Mahwah, NJ, 2001.

M.J. Yelland, C.J. Poulos, P.I. Pillans, G.M. Bashford, C.J. Nikles, J.M. Sturtevant, N. Vine, C.B. Del Mar, P.J. Schluter, M. Tan, J. Chan, F. Mackenzie, and R. Brown. N-of-1 randomized trials to assess the efficacy of gabapentin for chronic neuropathic pain. *Pain Medicine*, 10:754–761, 2009.

Francesca Solmi

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 049 827 4174
e-mail: solmi@stat.unipd.it

Current Position

Since January 2009; (expected completion: April 2012)
PhD Student in Statistical Sciences, University of Padova.
Thesis title: New permutation methodologies to deal with the multiplicity issue: multiple comparisons and multiple tests with applications to single- case experiments and to regression analysis.
Supervisor: Prof. Fortunato Pesarin
Co-supervisors: Prof. Luigi Salmaso and Prof. Patrick Onghena.

Research interests

- Permutation testing
- Nonparametric combination of permutation tests
- Applied statistics to the biomedical field

Education

February 2007 – July 2008
Master degree (*laurea specialistica/magistrale***) in Statistics and Computing Science** .
University of Padova, Faculty of Statistical Sciences
Title of dissertation: "Permutation methods: the multiplicity issue and complex applications in the biomedical field"
Supervisors: Prof. Fortunato Pesarin and Prof. Luigi Salmaso
Final mark: 106/110

September 2004 – February 2007
Bachelor degree (*laurea triennale***) in Statistics and Management**.
University of Padova, Faculty of Statistical Sciences

Title of dissertation: "The non-inferiority problem in medical statistics"
Supervisor: Prof. Fortunato Pesarin
Final mark: 110/110.

Visiting periods

November 2009 – october 2010
Faculty of Psychology and Educational Sciences, Katholieke Universiteit of Leuven
Leuven, Belgium
Supervisor: Prof. Patrick Onghena

Work experience

December 2009 – now
Johnson & Johnson Medical S.p.A., Pomezia (Roma), Italy.
Statistical consultant.

September 2008 – January 2009
Faculty of Statistical Sciences - University of Padova, Padova, Italy.
Tutor.

June 2002 – December 2008
Bertocco Lorella Srl, Padova, Italy.
Waitress.

January 2003 – August 2004
Zanellato Andrea Sas, Padova, Italy.
Employee.

October 2001 – November 2001
Wall Street Institute, Padova, Italy.
Employee.

June 2001 – September 2001
Omnitel Pronto Italia Spa, Padova, Italy.
Costumer care.

Awards and Scholarship

2009
PhD scholarship (University of Padova and Fondazione Cassa di Risparmio di Padova e Rovigo).

2011
Best Contributed Paper Award at the International Conference in Honor of the late Jagdish Srivastava, Optimal designs of Experiments - Theory and Application, Vienna (Austria), September 25th-30th 2011.

Computer skills

- Operative System: Windows
- Statistical softwares: R, Matlab, Eviews, Minitab, SAS, NPCtest, Logxact
- Markup Languages: LaTeX, HTML
- Programming: Javascript, PHP
- Other skills: Parellel computing

Language skills

Italian: native; English: fluent (written/spoken);

Publications

Articles in journals

Arboretti Giancristofaro, R., Bolzan, M., Bonnini, S., Corain, L., Solmi, F. (2011). Advantages of closed testing method for multiple comparison procedures. *Communications in Statistics, Simulation and Computation*, accepted for publication.

Bonnini, S., Piccolo, D., Salmaso, L., Solmi, F. (2011). Permutation inference for a class of mixture models. *Communications in Statistics, Theory and Methods* **12(6)**, accepted for publication.

Arboretti Giancristofaro, R., Bonnini, S., Solmi, F. (2008). Non-parametric two-stage active control testing method for non-inferiority tests. *Quaderni di Statistica* **10**, 73–98.

Arboretti Giancristofaro, R., Solmi, F., Salmaso, L., Grego, F., Maturi, C. (2007). Nonparametric methods for the comparison of two techniques of surgical operation to repair abdominal aortic aneurysms. *JP Journal of Biostatistics* **1**, 109–122.

Articles submitted

Bonnini, S., Solmi, F. (2011). Advances in CUB models. *Statistics and Computing*, submitted.

Bonnini, S., Salmaso, L., Solmi, F. (2011). Nonparametric inference via permutation tests for CUB models. In *Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, submitted.

Solmi, F., Onghena, P., Salmaso, L., Bulté, I. (2011). A permutation solution to test for treatment effects in alternation design single-case experiments. *Methodology*, submitted.

Solmi, F., Onghena, P., Salmaso, L., Bulté, I. (2011). Extensions of permutation solutions to test for treatment effects in replicated single-case alternation experiments with multivariate responses. *British Journal of Mathematical and Statistical Psychology*, submitted.

Working papers
Solmi, F., Onghena, P., Salmaso, L., Bulté, (2011). A permutation solution to test for treatment effects in alternation design single-case experiments. *Working Paper n. 19, Department of Management and Engineering, University of Padova, Italy.*

Conference presentations

Arboretti Giancristofaro, R., Bonnini, S., Solmi, F. (2011). Advances in CUB models with application to the evaluation of natural parks in the dolomites. (talk) *International Conference in Honor of the late Jagdish Srivastava, Optimal designs of Experiments − Theory and Application*, Vienna, Austria, September 25th-30th.

Bonnini, S., Salmaso, L., Solmi, F. (2011). Nonparametric inference via permutation tests for CUB models. (talk, Bonnini, S.) *8th Scientific Meeting of the CLassification and Data Analysis Group of the Italian Statistical Society*, Pavia, Italy, September 7th-9th.

Bonnini, S., Salmaso, L., Solmi, F. (2011). Statistical analysis of the perceived quality and customer satisfaction of a ski school: the Sesto Survey 2010. (talk, Bonnini, S.) *IES Conference - Innovation and Society, Statistical methods for service evaluation*, Florence, Italy, May 30th- June 1st.

Arboretti Giancristofaro, R., Bonnini, S., Corain, L., Solmi, F. (2010). A comparison on FWE-type multiple comparison procedures. (talk, Corain, L.) *Joint Statistical Meeting 2010*, Vancouver, Canada, July 31st- August 5th.

Arboretti Giancristofaro, R., Boatto, V, Bonnini, S., Solmi, F., Tempesta, T. (2010). Customer satisfaction survey on Passito wine with the application of a new approach to modeling discrete choices. (talk, Bonnini, S.) *Joint Statistical Meeting 2010*, Vancouver, Canada, July 31st- August 5th.

Solmi, F. (2010). Single-case experiments: a permutation solution for alternation designs. (poster) *Statistics for Complex Problems: The Multivariate Permutation Approach and Related Topics*, Padova, Italy, June 14th-15th.

Arboretti Giancristofaro, R., Bonnini, S., Corain, L., Solmi, F. (2009). A comparative study on multiple comparison procedures. (talk, Bonnini, S.) *Conference of the Italian Statistical Society*, Pescara, Italy, September 23th-25th.

Corain, L., Salmaso, L., Solmi, F. (2009). Multiple comparison procedures for treatment classification within ANOVA layout. (talk) *XVI Join Meeting of the French Society of Classification*, Grenoble, France, September 2nd-4th.

Bonnini, S., Corain, L., Salmaso, L., Solmi, F. (2009). Advantages of closed testing method for multiple comparison procedures. (talk, Bonnini, S.) *Conference "Multivariate Methods and Models for Evaluating Public Service"*, Rimini, Italy, May 25th−26th.

Grego, F., Solmi, F., Maturi, C., Salmaso, L., Pesarin, F., Deriu, G.P. (2007). Metodi non parametrici applicati al confronto tra due tecniche chirurgiche di esclusione di aneurisma alláorta addominale. (talk, Salmaso, L.) *VI Congresso Nazionale della Società Italiana di Biometria*, Rimini, Italy, 20th−22th June.

Other Interests

- Dog training, I am a member of the kennel club *Gruppo Cinofilo Educativo Amatori* of Albignasego (PD) Italy.
- Skiing in San Martino di Castrozza (TN) - Italy (www.sanmartino.com).
- Riding my motorbike.

References

Prof. Patrick Onghena
Faculty of Psychology and Educational Sciences
Katholieke Universiteit of Leuven
Dekenstraat 2, 3000 Leuven, Belgium
Phone: +32 16 326100 or +32 16 326137
e-mail: patrick.onghena@ppw.kuleuven.be