Journal of Statistical Software

September 2012, Volume 50, Issue 12. *http://www.jstatsoft.org/*

nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments

Kimihiro Noguchi
University of California, Davis

Yulia R. Gel
University of Waterloo

Edgar Brunner
University of Göttingen

Frank Konietschke
University of Göttingen

Abstract

Longitudinal data from factorial experiments frequently arise in various fields of study, ranging from medicine and biology to public policy and sociology. In most practical situations, the distribution of observed data is unknown and there may exist a number of atypical measurements and outliers. Hence, use of parametric and semiparametric procedures that impose restrictive distributional assumptions on observed longitudinal samples becomes questionable. This, in turn, has led to a substantial demand for statistical procedures that enable us to accurately and reliably analyze longitudinal measurements in factorial experiments with minimal conditions on available data, and robust nonparametric methodology offering such a possibility becomes of particular practical importance. In this article, we introduce a new R package **nparLD** which provides statisticians and researchers from other disciplines an easy and user-friendly access to the most up-to-date robust rank-based methods for the analysis of longitudinal data in factorial settings. We illustrate the implemented procedures by case studies from dentistry, biology, and medicine.

Keywords: nonparametric, longitudinal data, factorial design, **nparLD**, R.

1. Introduction

Longitudinal data are measurements collected from the same experimental units, usually referred to as subjects or individuals, over time. Such data are widely encountered in biology,

medicine, public policy, sociology, and psychology, and often arise in the following factorial settings:

- One homogeneous group of subjects is observed repeatedly at $s = 1, \ldots, t$ time points.
- One homogeneous group of subjects is observed repeatedly at $s = 1, \ldots, t$ time points, where each subject is observed under each time point several times (e.g., different vessel sections, left and right eye).
- One homogeneous group of subjects is observed repeatedly under $\ell = 1, \ldots, c$ conditions, where each subject is observed under each condition at $s = 1, \ldots, t$ time points.
- More than one homogeneous group of subjects (e.g., male and female, different treatment groups) are observed repeatedly at $s = 1, \ldots, t$ time points.

Typical questions that most practitioners deal with are:

- Do the treatments have the same effect?
- Is the time profile flat?
- Are the effects of the treatments similar over time? Are the treatment profiles parallel?

Alternatively, these practical questions can be translated into the statistical language as treatment effects, time effects, and interaction effects between treatment and time, respectively.

Measurements from the same experimental unit may be dependent, which leads to an extra level of complexity in longitudinal studies. Diggle, Liang, and Zeger (1994) provide a comprehensive overview of existing methods for longitudinal data analysis: generalized linear models (GLM) and their extensions, generalized linear mixed models (GLMM) and generalized estimating equations (GEE, Breslow and Clayton 1993; Zeger and Liang 1992; Liang and Zeger 1986). Most of these procedures are implemented in the commercial software SAS (SAS Institute Inc. 2003), e.g., SAS `PROC MIXED` for the analysis of linear mixed models. Publicly available software for longitudinal data analysis includes implementation of GLM/GLMM in R (R Development Core Team 2012) through the `glmmPQL()` function in **MASS** (Venables and Ripley 2002), the `lme()` function in **nlme** (Pinheiro, Bates, DebRoy, Sarkar, and R Development Core Team 2012), and the `lmer()` function in **lme4** (Bates, Maechler, and Bolker 2012); while GEE is available through **gee** (Carey, Lumley, and Ripley 2012) and **geepack** (Halekoh, Højsgaard, and Yan 2006, and references therein). However, all of these (semi)parametric procedures are based on specific model assumptions, e.g., existence of an expectation or homogeneous variances. In practice, such conditions can rarely be verified, and if observed measurements do not reflect the imposed conditions, e.g., in case of skewed distributions, outliers, or small sample sizes, parametric statistical procedures may result in unreliable or even false conclusions.

As an alternative, we can employ nonparametric rank-based methods that offer a flexible and robust framework for the analysis of a variety of longitudinal studies. In particular, in contrast to parametric procedures for factorial designs, the rank-based methodology is not restricted to data on a continuous scale and enables to analyze ordered categorical, dichotomous, and heavily skewed data in a systematic way (Konietschke, Bathke, Hothorn, and Brunner 2010).

Moreover, such nonparametric methods are robust to outliers and exhibit competitive performance for small sample sizes (Brunner, Domhof, and Langer 2002, Section 1.1, pp. 2). Note that, as discussed by Brunner *et al.* (2002), Robson (2002), Lehmann (2009), and Romano (2009), and references therein, if the assumptions of a parametric method are satisfied, the parametric procedure typically yields a higher power efficiency than its nonparametric counterpart. However, if these assumptions are not met or impossible to verify, the conclusions from the parametric method could be unreliable or even false. In such situations, more broadly applicable nonparametric approaches are preferred. Brunner and Puri (2001) and Brunner *et al.* (2002) provide a detailed overview of purely rank-based nonparametric methods for longitudinal data in factorial experiments, including descriptive point estimators of relative treatment effects (RTEs), confidence intervals, and test procedures. The hypotheses of "no treatment effect", "no time effect", and "no interaction between treatment and time" can be tested with nonparametric procedures for the analysis of data from factorial designs. Hereby, the hypotheses are not formulated in terms of expectations of treatment effects, but rather in terms of their distribution functions.

The working group from the Department of Medical Statistics, University of Göttingen, provides a SAS/IML macro library for nonparametric analyses of factorial longitudinal data. For each of different factorial designs, interactive matrix language (IML) code is implemented to test hypotheses formulated above and to compute confidence intervals. Given a high demand in publicly available software for robust longitudinal data analysis (Erceg-Hurn and Mirosevich 2008), we develop an R version of this SAS/IML macro library, i.e., a user friendly package **nparLD** that is freely available from the Comprehensive R Archive Network at `http://CRAN.R-project.org/package=nparLD` and offers nonparametric methodology for the most frequently arising factorial designs. The package **nparLD** is the first comprehensive R package that supports nonparametric methods in higher-way layouts. The name **nparLD** is interpreted as follows: "npar" stands for "nonparametric" while "LD" stands for "longitudinal data". All the implemented functions are accompanied by detailed help files and numerous examples. The goal of this paper is to introduce the developed nonparametric procedures in **nparLD** to a wide audience of statisticians and practitioners dealing with longitudinal studies.

The paper is organized as follows. In Section 2, we provide a brief overview of the nonparametric marginal model and commonly used factorial designs. In Section 3, we discuss the interpretation of results along with a review of relative treatment effects and their relationship to Wald-type statistics and analysis of variance (ANOVA)-type test statistics which are denoted by WTS and ATS, respectively, using case studies from dentistry, medicine, and biology. In Section 4, three different factorial longitudinal experiments are statistically evaluated with the new R package **nparLD**. Section 5 contains some conclusions and an outlook to future work.

2. Nonparametric factorial designs and hypotheses

We describe the idea of the nonparametric marginal model and its connection to different types of commonly arising factorial designs for longitudinal data. To classify common factorial designs, we introduce a notational system for each design depending on the number of factors. The factor which stratifies samples in independent groups, is called a *whole-plot factor*; while

	Data			Marginal distributions		
	Time (Factor T)			Time		
Subjects	$s=1$	$\cdots$	$s=t$	$s=1$	$\cdots$	$s=t$
$k=1$	X_{11}	$\cdots$	X_{1t}	F_1	$\cdots$	F_t
$\vdots$		$\vdots$			$\vdots$	
$k=n$	X_{n1}	$\cdots$	X_{nt}	F_1	$\cdots$	F_t

Table 1: The LD-F1 design and the corresponding marginal distributions.

the factor, stratifying repeated measurements, is called a *sub-plot factor*[1]. For example, when male and female subjects are observed at three time points, the factor sex is the whole-plot factor and the factor time is the sub-plot factor. Each design is denoted by Fx-LD-Fy, where x and y are the number of whole-plot and sub-plot factors, respectively, while "LD" stands for "longitudinal data". If only one group of homogeneous subjects is being considered, then such a design is denoted by LD-Fy instead. Examples of LD-F1, F1-LD-F1, and F2-LD-F1 designs are studied in Section 4. For more details and other common designs, we refer to Brunner *et al.* (2002, pp. 25 ff.).

2.1. LD-F1 design

We consider an experimental design consisting of $k = 1, \ldots, n$ experimental units (subjects). Each subject is observed repeatedly t times. The data from subject k are collected in the vector

$$\mathbf{X}_k = (X_{k1}, \ldots, X_{kt})^\top, \; k = 1, \ldots, n, \tag{1}$$

with marginal distributions

$$X_{1s}, \ldots, X_{ns} \sim F_s, \; s = 1, \ldots, t.$$

Here, F_s denotes the marginal cumulative distribution function from sample s. The total number of observations is $N = n \cdot t$. The data structure of such a trial is displayed in Table 1.

The hypothesis of "no time effect" is expressed in terms of the marginal distribution functions as

$$H_0^F(T) : F_1 = \ldots = F_t$$

and was introduced by Akritas and Arnold (1994).

As an illustration of the LD-F1 design, we may consider the psychiatric clinical trial by Bandelow *et al.* (1998) where the clinical global impression score (CGI) of 16 patients suffering from panic disorder was recorded during eight weeks under a treatment. Note that, since the response is measured on an ordered categorical scale, parametric and semiparametric mean-based approaches are inappropriate and may lead to unreliable conclusions.

In many factorial experiments, more than one homogeneous group of subjects is observed at multiple time points. Hence, in such cases, we get the Fx-LD-F1 design, where x is the number of whole-plot factors (e.g., treatments). The most common situation is the F1-LD-F1 design which is examined in Section 2.2.

[1]Here, whole-plot and sub-plot factors are referred to as between- and within-subjects factors, respectively.

		Data			Marginal distributions		
		Time (Factor T)			Time		
Factor A	Subjects	$s=1$	$\cdots$	$s=t$	$s=1$	$\cdots$	$s=t$
$i=1$	$k=1$	X_{111}	$\cdots$	X_{11t}	F_{11}	$\cdots$	F_{1t}
	$\vdots$		$\vdots$			$\vdots$	
	$k=n_1$	X_{1n_11}	$\cdots$	X_{1n_1t}	F_{11}	$\cdots$	F_{1t}
$\vdots$	$\vdots$		$\vdots$			$\vdots$	
$i=a$	$k=1$	X_{a11}	$\cdots$	X_{a1t}	F_{a1}	$\cdots$	F_{at}
	$\vdots$		$\vdots$			$\vdots$	
	$k=n_a$	X_{an_a1}	$\cdots$	X_{an_at}	F_{a1}	$\cdots$	F_{at}

Table 2: The F1-LD-F1 design and the corresponding marginal distributions.

2.2. F1-LD-F1 design

Suppose that a different groups of homogeneous subjects are observed repeatedly at t different time points and each group receives a randomly assigned treatment (treatment 1, treatment 2,..., treatment a). The statistical model underlying this design can be described by independent random vectors $\mathbf{X}_{ik} = (X_{ik1}, \ldots, X_{ikt})^\top, k = 1, \ldots, n_i$, with marginal distributions $X_{iks} \sim F_{is}, i = 1, \ldots, a; s = 1, \ldots, t$. The total number of observations is $N = n \cdot t$, where $n = \sum_{i=1}^{a} n_i$. The data structure of this F1-LD-F1 trial is displayed in Table 2.

The hypotheses of no main effect A, no main time effect T, and no interaction (AT) between A and T, are expressed in terms of the marginal distribution functions:

$$
\begin{aligned}
&H_0^F(A) : \overline{F}_{1\cdot} = \ldots = \overline{F}_{a\cdot} \\
&H_0^F(T) : \overline{F}_{\cdot 1} = \ldots = \overline{F}_{\cdot t} \\
&H_0^F(AT) : F_{is} = \overline{F}_{i\cdot} - \overline{F}_{\cdot s} + \overline{F}_{\cdot\cdot},\ i = 1, \ldots, a;\ s = 1, \ldots, t,
\end{aligned}
$$

where $\overline{F}_{i\cdot} = \frac{1}{t}\sum_{s=1}^{t} F_{is}$ denotes the mean distribution over time for treatment group i, $i = 1, \ldots, a$, $\overline{F}_{\cdot s} = \frac{1}{a}\sum_{i=1}^{a} F_{is}$ denotes the mean distribution over the treatment groups for time point s, $s = 1, \ldots, t$, and $\overline{F}_{\cdot\cdot} = \frac{1}{at}\sum_{i=1}^{a}\sum_{s=1}^{t} F_{is}$ denotes the overall mean distribution. Note that the hypotheses for the classical (parametric) linear longitudinal models are expressed in the same way with the expectations μ_{is}. For a discussion of the formulation of hypotheses by distribution functions, we refer to Akritas and Arnold (1994).

As an example, we may consider the study on efficacy of irrigation techniques in removing debris from irregularities in root canals with different apical sizes by Rödig, Sedghi, Konietschke, Lange, Ziebolz, and Hülsmann (2010). In this study, thirty extracted human pre-molars were randomly divided into three groups (each of size $n = 10$) followed by root canal preparation. In all three groups, three different irrigation procedures were performed and the amount of remaining debris was measured on an ordinal scale. This trial constitutes an F1-LD-F1 design. A more sophisticated setting may include an additional stratification of the whole-plot factor, which is illustrated in Section 2.3.

2.3. F2-LD-F1 design

The F2-LD-F1 design is one of the most widely used settings in factorial experiments in spite

			Data			Marginal distributions		
			Time (Factor T)			Time		
Factor A	Factor B	Subjects	$s=1$	$\cdots$	$s=t$	$s=1$	$\cdots$	$s=t$
$i=1$	$j=1$	$k=1$	X_{1111}	$\cdots$	X_{111t}	F_{111}	$\cdots$	F_{11t}
		$\vdots$		$\vdots$			$\vdots$	
		$k=n_{11}$	$X_{11n_{11}1}$	$\cdots$	$X_{11n_{11}t}$	F_{111}	$\cdots$	F_{11t}
	$\vdots$	$\vdots$		$\vdots$			$\vdots$	
	$j=b$	$k=1$	X_{1b11}	$\cdots$	X_{1b1t}	F_{1b1}	$\cdots$	F_{1bt}
		$\vdots$		$\vdots$			$\vdots$	
		$k=n_{1b}$	$X_{1bn_{1b}1}$	$\cdots$	$X_{1bn_{1b}t}$	F_{1b1}	$\cdots$	F_{1bt}
$\vdots$	$\vdots$	$\vdots$		$\vdots$			$\vdots$	
$i=a$	$j=1$	$k=1$	X_{a111}	$\cdots$	X_{a11t}	F_{a11}	$\cdots$	F_{a1t}
		$\vdots$		$\vdots$			$\vdots$	
		$k=n_{a1}$	$X_{a1n_{a1}1}$	$\cdots$	$X_{a1n_{a1}t}$	F_{a11}	$\cdots$	F_{a1t}
	$\vdots$	$\vdots$		$\vdots$			$\vdots$	
	$j=b$	$k=1$	X_{ab11}	$\cdots$	X_{ab1t}	F_{ab1}	$\cdots$	F_{abt}
		$\vdots$		$\vdots$			$\vdots$	
		$k=n_{ab}$	$X_{abn_{ab}1}$	$\cdots$	$X_{abn_{ab}t}$	F_{ab1}	$\cdots$	F_{abt}

Table 3: The F2-LD-F1 design and the corresponding marginal distributions.

of its complexity. Here, $a \cdot b$ different groups of homogeneous subjects are observed repeatedly at t different time points. Such a design can be employed, for example, when subjects are first stratified by their gender (female or male), then in each stratification, subjects are randomly assigned to different treatments (treatment 1, treatment 2,..., treatment a) and recorded at multiple time points. The statistical model of this trial can be described by independent random vectors $\mathbf{X}_{ijk} = (X_{ijk1}, \ldots, X_{ijkt})^\top, k = 1, \ldots, n_{ij}$, with marginal distributions $X_{ijks} \sim F_{ijs}, i = 1, \ldots, a; j = 1, \ldots, b; s = 1, \ldots, t$. The total number of observations is $N = n \cdot t$, where $n = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}$. The data structure of this F2-LD-F1 trial is displayed in Table 3, and the nonparametric hypotheses can be expressed in the same manner as in the F1-LD-F1 design.

To demonstrate an application of the F2-LD-F1 design, we may consider the shoulder tip pain study carried out with patients having undergone laparoscopic surgery in the abdomen, described by Lumley (1996). In this study, 41 patients were stratified by gender (Factor A), and in each stratification, the patients were randomly assigned to either the control or treatment group (Factor B). The pain scores of each patient were recorded at six time points (Factor T) to evaluate the effect of gender, treatment, and their interactions.

Instead of considering stratifications on the whole-plot factor, we may have stratifications on the sub-plot factor (time factor). Such designs include LD-F2, F1-LD-F2, and F2-LD-F2, which are discussed in detail by Brunner *et al.* (2002, Chapter 2).

In Section 3, we examine rank estimators of the relative treatment effects and test procedures for the hypotheses discussed above.

3. Nonparametric effects, estimators, and test procedures

The general model (1)[2] does not entail any parameters by which a difference between the distributions could be described. Therefore, the mean distribution function $H(x) = \frac{1}{t}\sum_{s=1}^{t} F_s(x)$ and the distribution functions $F_s(x)$ are used to define a treatment effect as a marginal summary measure between H and F_s:

$$p_s = \int H dF_s = P(Z < X_{1s}) + 0.5P(Z = X_{1s}), \;\; s = 1, \ldots, t, \tag{2}$$

where $Z \sim H$ denotes a randomly chosen observation from the whole data set independently from X_{1s}. These so-called *relative marginal effects*[3] p_s can be regarded as the probability that a randomly chosen observation X_{sk} at time point s tends to result in a larger value than Z. The interpretation of p_s is rather simple; i.e., X_{1j} tends to result in

- a smaller value than X_{2s}, if $p_j < p_s$,
- a larger value than X_{2s}, if $p_j > p_s$,
- neither a smaller nor larger value than X_{2s}, if $p_j = p_s$.

A graphical illustration of the tendency is given in Figure 1, where both the distribution functions $F_i(x), F_j(x)$, and the mean distribution function $H(x)$, are displayed.

Figure 1 suggests that, for an easier interpretation of the effect size measure p_s, it is sufficient to consider only the relation between p_j and p_s referring to $<, =$, and $>$, respectively. For a detailed discussion of p_s, we refer to Brunner *et al.* (2002, Section 3.1, pp. 35 ff).

The unknown quantities p_s can be estimated with overall ranks of the data by replacing all the observations $X_{11}, \ldots, X_{ns}$ with their overall ranks $R_{11}, \ldots, R_{ns}$. Further, let $\overline{R}_{\cdot s} =$

[2]To derive the results for factorial settings, sub-indices for the random vectors $\mathbf{X}_k$ in the model (1) are necessary.

[3]Here, we use the term "relative marginal effect" as a synonym for "relative treatment effect".

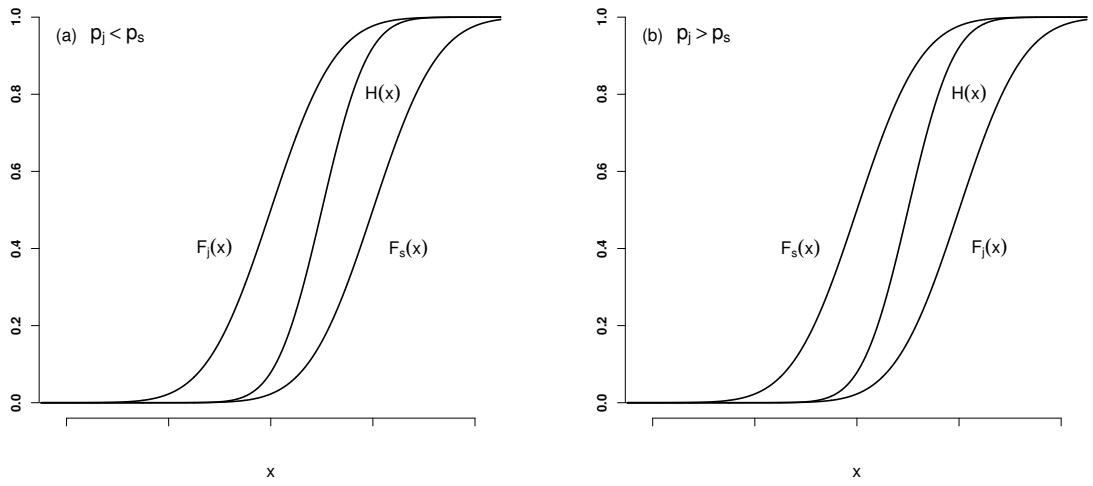

Figure 1: Stochastic tendency: (a) $p_j < p_s$, (b) $p_j > p_s$.

$\frac{1}{n}\sum_{k=1}^{n} R_{sk}$ denote the mean of the ranks in the marginal sample s. Then, an asymptotically unbiased and consistent estimator of p_s is given by

$$\widehat{p}_s = \frac{1}{N}\left(\overline{R}_{\cdot s} - \frac{1}{2}\right), \quad s = 1, \ldots, t,$$

where $N = n \cdot t$. To test the hypothesis $H_0^F : F_1 = \ldots = F_t$, let $\mathbf{C}$ denote a suitable contrast matrix, $\mathbf{F} = (F_1, \ldots, F_t)^\top$ the vector of distributions, $\mathbf{p} = (p_1, \ldots, p_t)^\top$ the vector of the relative marginal effects, and $\widehat{\mathbf{p}}$ the vector of corresponding estimators. Further, let $\widehat{\mathbf{V}}_n$ denote the empirical covariance matrix of the ranks, noting that $\widehat{\mathbf{V}}_n$ is a consistent estimator. Then, under the hypothesis $H_0^F : \mathbf{CF} = \mathbf{0}$, the Wald-type statistic (WTS)

$$Q_n(\mathbf{C}) = n\widehat{\mathbf{p}}^\top \mathbf{C}^\top [\mathbf{C}\widehat{\mathbf{V}}_n \mathbf{C}^\top]^+ \mathbf{C}\widehat{\mathbf{p}} \tag{3}$$

has, asymptotically, a χ^2_f distribution with $f = \text{rank}(\mathbf{C})$ degrees of freedom under some mild regularity conditions (Akritas, Arnold, and Brunner 1997). Here, $\mathbf{M}^+$ denotes the Moore-Penrose inverse of a matrix $\mathbf{M}$.

It is well known that the convergence of the distribution of Q_n to its limiting χ^2_f distribution is relatively slow (Akritas *et al.* 1997; Akritas and Brunner 1997; Brunner, Munzel, and Puri 1999; Brunner *et al.* 2002). In general, the asymptotic approximation deteriorates with an increase of a number of factor levels as well as for smaller sample sizes. Hence, we also present a small sample modification of this test statistic that maintains an accurate size of the test even for small sample sizes ($n \geq 7$).

Let $\mathbf{T} = \mathbf{C}^\top [\mathbf{C}\mathbf{C}^\top]^- \mathbf{C}$ denote the projection matrix obtained from $\mathbf{C}$ where $\mathbf{M}^-$ denotes a generalized inverse of $\mathbf{M}$. Then, the distribution of the ANOVA-type statistic (ATS)

$$A_n(\mathbf{C}) = \frac{n}{tr(\mathbf{T}\widehat{\mathbf{V}})} \widehat{\mathbf{p}}^\top \mathbf{T}\widehat{\mathbf{p}} \tag{4}$$

can be approximated by the $F_{(\widehat{f},\infty)}$ distribution, where $\widehat{f} = [tr(\mathbf{T}\widehat{\mathbf{V}})]^2 / tr(\mathbf{T}\widehat{\mathbf{V}}\mathbf{T}\widehat{\mathbf{V}})$ (Brunner, Dette, and Munk 1997; Brunner *et al.* 1999; Brunner and Puri 2001).

Note that, depending on additional stratification or experimental groups, an appropriate partition into sub-indices is necessary. For example, if the experimental units are divided into $i = 1, \ldots, a$ groups with $j = 1, \ldots, n_i$ units in the i-th group, then X_{ijs} denotes the random variable for the jth unit in the ith group at time point s.

For the main effects of the whole-plot factors and interactions involving only the whole-plot factors, the distribution of ATS can be further approximated by applying a finite denominator degrees of freedom $\widehat{f}_0$, i.e., by the $F_{(\widehat{f},\widehat{f}_0)}$ distribution, taking advantage of the diagonal covariance matrix resulting from independence of subjects [4]. (For further details on $\widehat{f}_0$, see Brunner *et al.* (1997) and Brunner *et al.* (2002, Section 8.3.3, pp. 134 ff).) Real-life applications of the so-called modified ATS can be found in Sections 4.2 and 4.3.

Note that, the adjusted degrees of freedom used for the approximation of the distribution of ATS may appear to be quite different from the conventional degrees of freedom employed in the traditional repeated measures ANOVA. However, such an adjustment for degrees of

[4] Bathke, Schabenberger, Tobias, and Madden (2009) mention that ATS becomes too conservative for testing an effect with a finite denominator degrees of freedom when a sub-plot factor is involved. Therefore, the use of $F_{(\widehat{f},\infty)}$ instead of $F_{(\widehat{f},\widehat{f}_0)}$ is recommended for ATS involving sub-plot factors.

freedom can be viewed as a generalization of the conventional degrees of freedom in the heteroscedastic case. For instance, for the repeated measures data in the LD-F1 design, assuming sphericity, e.g., compound symmetry structure, of the covariance matrix, approximate degrees of freedom for the distribution of ATS, which is equivalent to treatment sum of squares divided by residual sum of squares, are equal to $(\widehat{f}, \widehat{f}_0) = (t-1, (n-1)\cdot(t-1))$ (Bathke *et al.* 2009). However, in general, ranked observations are heteroscedastic even if the original observations are homoscedastic (Akritas 1990), and thus it is reasonable to assume an arbitrary (unstructured) covariance matrix (Brunner *et al.* 2002, Section 2.2, pp. 33)[5]. Therefore, for such covariance matrix structure, an appropriate adjustment for degrees of freedom is necessary to draw a valid inference using ATS.

4. Examples

In this section, we provide examples that illustrate how different factorial designs can be analyzed using the package **nparLD**. Along with the individual functions for specific designs (`ld.f1()`, `ld.f2()`, `f1.ld.f1()`, `f2.ld.f1()` and `f1.ld.f2()`), the package provides a wrapper function `nparLD()` that automatically identifies the most suitable design through the formula provided by a user. The wrapper function `nparLD()` creates a class object called `nparLD` from which users may obtain short and extended summaries as well as a plot of the results using `print()`, `summary()`, and `plot()` for the `nparLD` object, respectively. In particular, the `print()` function displays basic results about the model formula and results from the WTS and ATS; the `summary()` function shows the relative treatment effects in detail additionally to the output provided by `print()`. Finally, `plot()` creates plots of the relative treatment effects and their corresponding confidence intervals at different time points.

4.1. Study from dentistry

Our first case study is related to a growth curve problem where the LD-F1 design may be employed. Potthoff and Roy (1964) assess distances (in millimeters) between the center of the pituitary and the pterygomaxillary fissure of 16 boys and 11 girls on four different occasions, i.e., at the ages 8, 10, 12, and 14, and conclude that two separate growth curves are required for boys and girls. In this example, we focus on the homogeneous group of the 16 boys. (The data for boys are available in `dental`. A complete dataset for both boys and girls is available in `Orthodont` in the package **nlme**.) In particular, we are interested in testing the hypothesis $H_0^F(T) : F_8 = F_{10} = F_{12} = F_{14}$ of no time effect, where F_s denotes the marginal distribution of the distances at age s. We start our analysis by examining the box plot shedding light on the distribution of the data for each age group, and by observing the plot of the relative treatment effect for each age group along with the pointwise 95% confidence intervals. The plots are generated using the following code:

```
R> library("nparLD")
R> data("dental")
R> boxplot(resp ~ time, data = dental, lwd = 2, xlab = "time",
+    font.lab = 2, cex.lab = 2, main = "Box Plots")
```

[5]When data are assumed to have an equal correlation structure, Schörgendorfer, Madden, and Bathke (2011) suggest to utilize a heterogeneous compound symmetry (CSH) structure for ranked data, in order to reduce the number of estimated parameters and to obtain less conservative results using ATS.

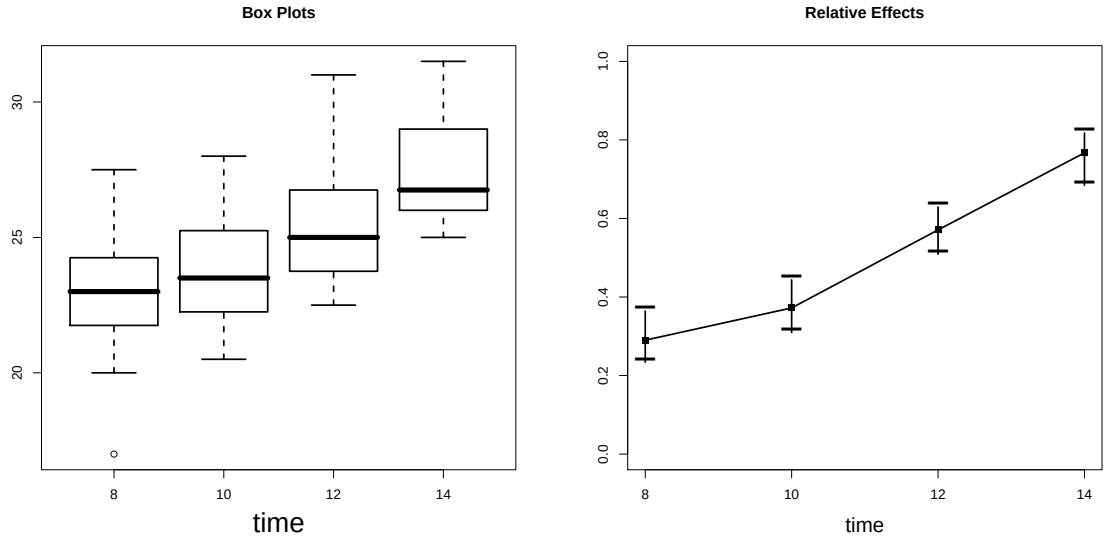

Figure 2: Box plots and 95% confidence intervals for p_s in the dental study.

```
R> ex.f1np <- nparLD(resp ~ time, data = dental, subject = "subject",
+    description = FALSE)
R> plot(ex.f1np)

 LD F1 Model
 ------------------------
 Check that the order of the time level is correct.
 Time level:   8 10 12 14
 If the order is not correct, specify the correct order in time.order.
```

Remark. Note that, as a precautionary check, the function `nparLD()` automatically provides information about the selected model as well as the order of levels in the sub-plot factor. The user may choose to hide this information by setting `order.warning = FALSE`. Moreover, more detailed information about the data and abbreviations used in the output become available by setting `description = TRUE`.

The box plots at the left panel of Figure 2 show the minimum, first quartile, median, third quartile, and the maximum distance measured for each time point separately. They indicate that the measured distances have a somewhat skewed distribution (especially as the age goes up). The increase in median gives rise to a time effect. The 95% confidence intervals at the right panel of Figure 2 present the lower bound, point estimate, and the upper bound for each time point separately using `plot(ex.f1np)`, where `ex.f1np` is an `nparLD` class object. The point estimates increase, meaning the older the boys, the larger the observed distances between pituitary and the pterygomaxillary fissure. The exact values for the bounds and point estimates are obtainable by using the code `plot(ex.f1np)$Conf.Int`. The results can be explored further by using `summary(ex.f1np)`.

Remark. Note that, a further insight about dependence of the measurements per subject can be readily obtained by using the `groupedData()` function from the R package **nlme** (Pinheiro *et al.* 2012) that produces subject-specific line plots in factorial experiments.

```
R> summary(ex.f1np)

Model:
LD F1 Model

Call:
resp ~ time

Relative Treatment Effect (RTE):
       RankMeans Nobs       RTE
time8   19.06250   16 0.2900391
time10  24.31250   16 0.3720703
time12  37.03125   16 0.5708008
time14  49.59375   16 0.7670898

Wald-Type Statistc (WTS):
     Statistic df      p-value
time  94.47718  3 2.391503e-20

ANOVA-Type Statistc (ATS):
     Statistic       df      p-value
time  31.48774 2.700785 1.437729e-18
```

For each age group s, the rank mean of the overall ranks (RankMeans), the number of observations (Nobs) and the point estimate $\widehat{p}_s$ of the relative treatment effect (RTE) are displayed. The obtained result of 0.29 for the age group 8 (time8) can be interpreted, for example, as follows: a randomly chosen observation from the whole dataset results in a smaller value than a randomly chosen observation from the age group 8 with an estimated probability of 29%. Further, since $\widehat{p}_8 < \widehat{p}_{10} < \widehat{p}_{12} < \widehat{p}_{14}$, the observations from the age group 8 tend to result in smaller values than those from the age group 10 which, in return, also tends to result in smaller values than the measurements from the age groups 12 and 14, respectively. Thus, an increase in the effect seems to indicate the increase in the measured distances.

To test the hypothesis $H_0^F(T)$ of no time effect, WTS in (3) and ATS in (4) can be applied, which are also displayed in the output of `summary(ex.f1np)`. The column `df` for ATS is the numerator degrees of freedom of the F distribution as the denominator degrees of freedom is set to infinity. Both WTS and ATS yield highly statistically significant p values of 2.392×10^{-20} and 1.438×10^{-18}, respectively, indicating that the null hypothesis of no time effect is to be rejected. To investigate the question about which of the four distribution functions differ, we can apply multiple comparisons with the Bonferroni adjustment as described below:

```
R> m810 <- which(((dental$time == 8) + (dental$time == 10)) == 1)
R> m812 <- which(((dental$time == 8) + (dental$time == 12)) == 1)
R> m814 <- which(((dental$time == 8) + (dental$time == 14)) == 1)
```

Comparison	Hypothesis	p value	Adjusted p value
Time 8 vs. Time 10	$H_0^F : F_8 = F_{10}$	0.2204	0.6612
Time 8 vs. Time 12	$H_0^F : F_8 = F_{12}$	< 0.0001	< 0.0001
Time 8 vs. Time 14	$H_0^F : F_8 = F_{14}$	< 0.0001	< 0.0001

Table 4: Multiple comparisons against the control in the dental study with Bonferroni adjustment.

```
R> ex.f1np810 <- nparLD(resp ~ time, data = dental[m810,],
+    subject = "subject", description = FALSE)
R> ex.f1np812 <- nparLD(resp ~ time, data = dental[m812,],
+    subject = "subject", description = FALSE)
R> ex.f1np814 <- nparLD(resp ~ time, data = dental[m814,],
+    subject = "subject", description = FALSE)
R> summary(ex.f1np810)
R> summary(ex.f1np812)
R> summary(ex.f1np814)
```

The results are presented in Table 4, where, for brevity, only the p values obtained from ATS are reported.

In Table 4, the Bonferroni-adjusted p value of 0.6612, obtained for testing the age group 8 against the age group 10 (Time 8 vs. Time 10), is calculated by multiplying the original p value of 0.2204 by 3. Similar calculations are also performed for the other pairwise comparisons. From the results, we can conclude that the distance between the center of the pituitary and the pterygomaxillary fissure significantly increases over time by observing the p values of < 0.0001 from both WTS and ATS. In addition, we notice significant differences between the distributions of the measured distances for the age groups 8 and 12, and age groups 8 and 14, respectively. To compare the obtained results and conclusions with parametric methods, we further reanalyze the data with the `lme()` function in the R package **nlme** (Pinheiro *et al.* 2012) as described below:

```
R> library("nlme")
R> ex.f1lme <- lme(resp ~ time, data = dental, random = ~ 1 | subject)
R> summary(ex.f1lme)
```

We obtain an overall significant time effect (p value < 0.0001). Regarding the multiple comparisons against age group 8, using the code

```
R> ex.f1lme810 <- lme(resp ~ time, data = dental[m810,],
+    random = ~ 1 | subject)
R> ex.f1lme812 <- lme(resp ~ time, data = dental[m810,],
+    random = ~ 1 | subject)
R> ex.f1lme814 <- lme(resp ~ time, data = dental[m810,],
+    random = ~ 1 | subject)
R> summary(ex.f1lme810)
R> summary(ex.f1lme812)
R> summary(ex.f1lme814)
```

and multiplying the original p value by 3, we obtain the adjusted p value of 0.4395 for the comparison "Time 8 vs. Time 10", as well as the p values of 0.0009 and < 0.0001 for "Time 8 vs. Time 12" and "Time 8 vs. Time 14", respectively. Thus, both parametric and nonparametric procedures result in similar conclusions in this example, which is not surprising since the data exhibit only a minor degree of skewness as indicated by the box plots (see the left panel of Figure 2).

4.2. Rat growth study

Our next example deals with the study of body weights of 27 rats (Box 1950; Wolfinger 1996). Each rat was randomly assigned to one of three treatments (control, thyroxin, or thiouracil) with sample sizes 10, 7, and 10, respectively. Thyroxin is a thyroid hormone typically applied in hypothyroidism, and thiouracil is a drug that suppresses generation of thyroxin. The first group was kept as a control while the second and third group had thyroxin and thiouracil added to their drinking water, respectively. The weight (in grams) of each rat was recorded at baseline and subsequent four weeks. Thus, this experiment has the F1-LD-F1 design structure with treatment being the whole-plot factor. As is pointed out by Wolfinger (1996), the body weights of the 27 rats show a "fanning effect", indicating an increase in variability over time. Although a standard technique of the logarithmic transformation makes the data more homoscedastic, there is a concern that such a nonlinear transformation distorts the time and treatment effects. Therefore, we analyze the data using our nonparametric methods. Similar to the dental study, we first examine the box plots of the data and the plot of the relative treatment effect estimates and their corresponding confidence intervals.

```
R> library("nparLD")
R> data("rat")
R> boxplot(resp ~ group * time, data = rat, names = FALSE,
+    col = c("grey", 2, 3), lwd = 2)
R> axis(1, at = 2, labels = "Time 0", font = 2, cex = 2)
R> axis(1, at = 5, labels = "Time 1", font = 2,cex = 2)
R> axis(1, at = 8, labels = "Time 2", font = 2,cex = 2)
R> axis(1, at = 11, labels = "Time 3", font = 2,cex = 2)
R> axis(1, at = 14, labels = "Time 4", font = 2,cex = 2)
R> legend(2, 190, c("Control", "Thiour", "Thyrox"), lwd = c(3, 3, 3),
+    col = c("grey", 2, 3), cex = 2)
R> ex.f1f1np <- nparLD(resp ~ time * group, data = rat,
+    subject = "subject", description = FALSE)
R> plot(ex.f1f1np)
```

```
F1 LD F1 Model
----------------------
Check that the order of the time and group levels are correct.
Time level:   0 1 2 3 4
Group level:   control thyrox thiour
If the order is not correct, specify the correct order in time.order or
group.order.
```

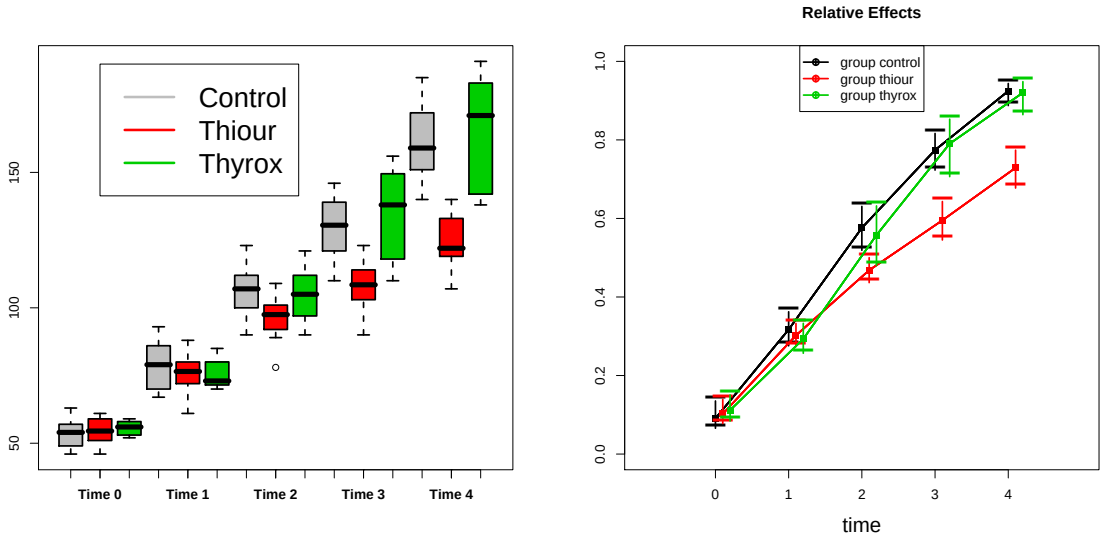

Figure 3: Box plots and 95% confidence intervals for p_{is} in the rat growth study. Thiouracil and thyroxin are denoted by thiour and thyrox, respectively.

Figure 3 shows box plots (the left panel) and 95% confidence intervals (the right panel) for the relative treatment effects in the rat growth study for the three treatments and each time point separately. The box plots indicate that the data follow a skewed distribution. The 95% confidence intervals further imply that the weights increase over time in all treatment groups.

The main question of this experiment is whether the time profiles of the three experiments are parallel, i.e., if there exists a statistical interaction between treatment and time. Absence of such an interaction would be indicated by parallel time profiles. Regarding the box plots and the 95% confidence intervals in Figure 3, the time profile from the thyroxin group does not seem to be parallel to both the control and the thiouracil group time profiles. A secondary question of interest would be to know whether the treatments affect the weight gains. The `nparLD()` and `summary()` functions can be applied to test the hypothesis of no interaction between treatment and time formulated in terms of the distribution functions, which help us answer questions raised above.

```
R> summary(ex.f1f1np)
```

For the F1-LD-F1, F1-LD-F2, and F2-LD-F1 designs, in addition to WTS and ATS, the modified ATS using the Box (1954) approximation for the whole-plot factors and their interaction (Brunner *et al.* 1997) are available. The modified ATS has a finite denominator degrees of freedom (denoted by $\widehat{f}_0$ as discussed in Section 3) as opposed to ATS which has the denominator degrees of freedom equal to infinity, in order to improve approximation of the distribution under the hypothesis of "no treatment effect" and "no interaction between the whole-plot factors". The output for each test is presented below:

```
Model:
F1 LD F1 Model
```

```
Call:
resp ~ time * group

Relative Treatment Effect (RTE):
                  RankMeans Nobs        RTE
groupcontrol       72.92000   50 0.53644444
groupthyrox        72.67143   35 0.53460317
groupthiour        59.81000   50 0.43933333
time0              14.20714   27 0.10153439
time1              41.55714   27 0.30412698
time2              72.57857   27 0.53391534
time3              97.68810   27 0.71991182
time4             116.30476   27 0.85781305
groupcontrol:time0  12.75000   10 0.09074074
groupcontrol:time1  43.30000   10 0.31703704
groupcontrol:time2  78.25000   10 0.57592593
groupcontrol:time3 105.00000   10 0.77407407
groupcontrol:time4 125.30000   10 0.92444444
groupthyrox:time0   15.57143    7 0.11164021
groupthyrox:time1   40.07143    7 0.29312169
groupthyrox:time2   75.78571    7 0.55767196
groupthyrox:time3  107.21429    7 0.79047619
groupthyrox:time4  124.71429    7 0.92010582
groupthiour:time0   14.30000   10 0.10222222
groupthiour:time1   41.30000   10 0.30222222
groupthiour:time2   63.70000   10 0.46814815
groupthiour:time3   80.85000   10 0.59518519
groupthiour:time4   98.90000   10 0.72888889

Wald-Type Statistc (WTS):
            Statistic df      p-value
group        12.52657  2 1.904977e-03
time       3619.03739  4 0.000000e+00
group:time   70.34311  8 4.199050e-12

ANOVA-Type Statistc (ATS):
             Statistic       df      p-value
group         5.286582 1.922792 5.654723e-03
time       1008.512138 1.990411 0.000000e+00
group:time   11.093940 3.516933 3.616929e-08

Modified ANOVA-Type Statistic for the Whole-Plot Factors:
      Statistic      df1      df2    p-value
group  5.286582 1.922792 19.23468 0.01563658
```

The `summary()` function automatically creates output tables for the hypotheses of "no treatment effect", "no time effect", and "no interaction" using both WTS in (3) and ATS in (4).

In this study, the hypothesis of no interaction, i.e., parallel time profiles, is rejected at the 1% level using both WTS and ATS with the p values of 4.199×10^{-12} and 3.617×10^{-8}, respectively. To investigate the question of whether the hypothesis of no interaction is rejected at each time point, multiple comparisons can be performed by using the `nparLD()` function repeatedly with the baseline and week 1 values, followed by comparison of the baseline, week 1, and week 2 values, etc. Similarly as in the dental study, the Bonferroni adjustment can be applied to control the Type I error.

4.3. Respiratory disorder study

Our last example concerns about a clinical trial for patients with a respiratory disorder (Koch, Carr, Amara, Stokes, and Uryniak 1990). A total of 111 patients from two centers were randomly assigned to two treatments (active or placebo). The status of each patient was recorded on an ordinal scale (0 = terrible, 1 = poor, 2 = fair, 3 = good, 4 = excellent) at baseline and subsequent four visits. An advantage of the rank-based nonparametric methods is that they can handle ordinal data in the same manner without requiring any further transformation. We analyze the effects of treatment, center, visit, and their interactions similar to the study conducted by Koch *et al.* (1990, pp. 458) using the ordinal scale from 0 to 4. The main question of this trial is whether or not the active treatment group (A) reveals a significantly better clinical record than the placebo group (P). Since there are two whole-plot factors (treatment and center), an appropriate factorial design to consider is F2-LD-F1. We start our analysis by observing the relative treatment effect and its corresponding confidence interval at each time point for the two centers, using the following code:

```
R> library("nparLD")
R> data("respiration")
R> par(mfrow = c(1, 2))
R> center <- respiration[,"center"]
R> boxplot(resp ~ treatment * time, data = respiration[which(center == 1),],
+    names = FALSE, ylim = c(-1, 5), col = c("grey", 2), lwd = 2,
+    main = "Center 1")
R> axis(1, at = 2, labels = "Time 1", font = 2, cex = 2)
R> axis(1, at = 5, labels = "Time 2", font = 2, cex = 2)
R> axis(1, at = 8, labels = "Time 3", font = 2, cex = 2)
R> axis(1, at = 11, labels = "Time 4", font = 2, cex = 2)
R> axis(1, at = 14, labels = "Time 5", font = 2, cex = 2)
R> legend(2, 5, c("Treat A", "Treat P"), lwd = c(2, 2), col = c("grey", 2),
+    cex = 1.2)
R> boxplot(resp ~ treatment * time, data = respiration[which(center == 2),],
+    names = FALSE, ylim = c(-1, 5), col = c("grey", 2), lwd = 2,
+    main = "Center 2")
R> axis(1, at = 2, labels = "Time 1", font = 2, cex = 2)
R> axis(1, at = 5, labels = "Time 2", font = 2, cex = 2)
R> axis(1, at = 8, labels = "Time 3", font = 2, cex = 2)
R> axis(1, at = 11, labels = "Time 4", font = 2, cex = 2)
R> axis(1, at = 14, labels = "Time 5", font = 2, cex = 2)
R> legend(2, 5, c("Treat A", "Treat P"), lwd = c(2, 2), col = c("grey", 2),
```

```
+    cex = 1.2)
R> ex.f2f1np <- nparLD(resp ~ time * center * treatment, data = respiration,
+    subject = "patient", description = FALSE)
R> plot(ex.f2f1np)

 F2 LD F1 Model
 ----------------------
 Check that the order of the time, group1, and group2 levels are correct.
 Time level:   1 2 3 4 5
 Group1 level:   1 2
 Group2 level:   A P
 If the order is not correct, specify the correct order in time.order,
 group1.order, or group2.order.
```

As the box plots suggest (see the upper panel of Figure 4), scores under the active treatment are larger than under placebo, and that there exists a difference in scores between the two centers. The lower panel of Figure 4 shows the 95% confidence intervals for the relative treatment effects p_{ijs}. We observe that, in both centers, the effects for the placebo group seem to remain constant over time, while the effects for the active treatment group increase. In Center 1, the clinical record seems to decrease at the last visit, which cannot be observed in Center 2. The time effect, treatment effect, center effect, and their interactions can be analyzed using the `print()` function.

```
R> print(ex.f2f1np)

Model:
F2 LD F1 Model

Call:
resp ~ time * center * treatment

Wald-Type Statistc (WTS):
                      Statistic df     p-value
center                10.2569587  1 0.001361700
treatment              9.3451482  1 0.002235766
time                  17.4568433  4 0.001575205
center:treatment       1.2365618  1 0.266134717
center:time            8.7200395  4 0.068491057
treatment:time        17.5434583  4 0.001515158
center:treatment:time  0.2898785  4 0.990458142

ANOVA-Type Statistc (ATS):
                       Statistic       df      p-value
center                10.25695866 1.000000 0.0013616998
treatment              9.34514819 1.000000 0.0022357657
time                   4.43527016 3.320559 0.0028528788
center:treatment       1.23656176 1.000000 0.2661347165
```

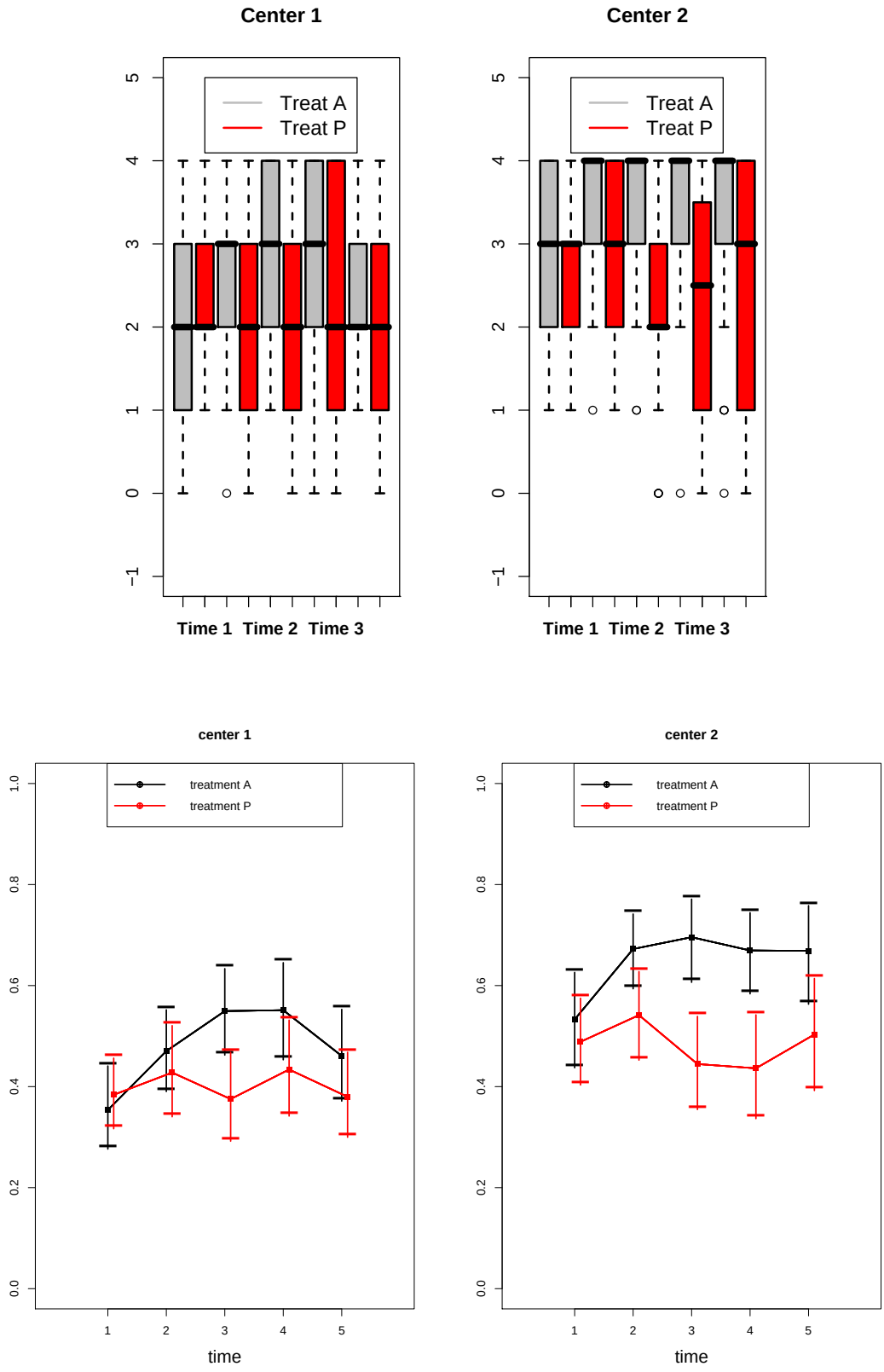

Figure 4: Box plot and 95% confidence intervals for p_{ijs} in the respiratory disorder study.

```
center:time            1.60699585 3.320559 0.1802120504
treatment:time         5.46185031 3.320559 0.0005867392
center:treatment:time  0.05915234 3.320559 0.9866660535

Modified ANOVA-Type Statistic for the Whole-Plot Factors:
                 Statistic df1      df2     p-value
center           10.256959   1 104.9255 0.001803091
treatment         9.345148   1 104.9255 0.002836284
center:treatment  1.236562   1 104.9255 0.268676117
```

The `print()` function automatically prints out tables of results for all main effects and interactions in the F2-LD-F1 design using WTS, ATS, and a modified ATS for the whole-plot factors. Both WTS and ATS indicate a significant interaction between treatment and time at the 5% level, with the p value of 0.0006 using ATS. This confirms the interpretation of the confidence intervals regarding the time profiles in Figure 4. Moreover, both a significant treatment (p value of 0.0028) and center effect (p value of 0.0018) are observed from the modified ATS, where the significant center effect should be further investigated.

Now, let us compare the obtained results with the conclusions provided by the parametric methods. In particular, we apply `lme()` from **nlme** to our respiratory data below:

```
R> library("nlme")
R> ex.f2f1lme <- lme(resp ~ time * treatment * center, data = respiration,
+    random = ~ 1 | patient)
R> summary(ex.f2f1lme)

Linear mixed-effects model fit by REML
 Data: respiration
       AIC      BIC    logLik
  1585.142 1628.187 -782.5712

Random effects:
 Formula: ~1 | patient
        (Intercept)  Residual
StdDev:   0.8673626 0.8073981

Fixed effects: resp ~ time * treatment * center
                            Value Std.Error  DF    t-value p-value
(Intercept)             1.3148148 0.5216425 440  2.5205287  0.0121
time                    0.1740741 0.1098730 440  1.5843212  0.1138
treatmentP              0.3729931 0.7261302 107  0.5136724  0.6085
center                  0.8111111 0.3299157 107  2.4585408  0.0156
time:treatmentP        -0.1399608 0.1529440 440 -0.9151113  0.3606
time:center            -0.0407407 0.0694898 440 -0.5862840  0.5580
treatmentP:center      -0.2782293 0.4604257 107 -0.6042872  0.5469
time:treatmentP:center -0.0209588 0.0969789 440 -0.2161167  0.8290
```

Note that, the parametric method of `lme()` yields completely different conclusions compared to the nonparametric procedure of `nparLD()`. In particular, the results from `lme()` imply that

time is an insignificant variable, i.e., the obtained p value is 0.1138, while `nparLD()` provides a highly statistically significant p value of 0.0029, concluding that the scores do not evolve in time. Similarly, treatment is declared as an insignificant factor by `lme()`, with the p value of 0.6085; in contrast, `nparLD()` concludes that there exists a significant treatment effect, with the resulting p value of 0.0022. Such contradictory results are not surprising since the respiratory data are observed on an ordered categorical scale and, hence, parametric methods are inapplicable. Thus, following the output of `nparLD()`, we are inclined to conclude that scores are dynamic in treatment and time, and that interaction between treatment and time is significant. These results are also confirmed by the graphical diagnostics provided by the box plots of the respiratory data (see the upper panel of Figure 4).

5. Conclusion and future work

The R package **nparLD** implements a broad range of rank-based nonparametric methods for analyzing longitudinal data in factorial experiments. A notable novel feature of **nparLD** is that it accommodates various factorial designs, including higher-way layouts. The users can easily evaluate the treatment and time effects as well as their interactions via the robust ANOVA-type statistic (ATS), which accurately controls the Type I error rate even for small sample sizes, and the classical Wald-type statistic (WTS). We plan to update the package **nparLD** on a regular basis with new nonparametric statistical procedures available for longitudinal data. In particular, we aim to implement the multiple contrast testing procedures discussed by Konietschke *et al.* (2010). In addition, we plan to undertake a major update of the code and release **nparLD** in the S4 style.

Acknowledgments

The research of Y. Gel was supported by a grant from the National Science and Engineering Research Council (NSERC) of Canada, and the research of F. Konietschke was supported by a grant from the German Academic Exchange Service (DAAD). This work was supported in part by the German Research Foundation grant DFG-BR 655/16-1.

References

Akritas MG (1990). "The Rank Transform Method in Some Two-Factor Designs." *Journal of the American Statistical Association*, **85**, 73–78.

Akritas MG, Arnold SF (1994). "Fully Nonparametric Hypotheses for Factorial Designs I: Multivariate Repeated Measures Designs." *Journal of the American Statistical Association*, **89**, 336–343.

Akritas MG, Arnold SF, Brunner E (1997). "Nonparametric Hypotheses and Rank Statistics for Unbalanced Factorial Designs." *Journal of the American Statistical Association*, **92**, 258–265.

Akritas MG, Brunner E (1997). "A Unified Approach to Rank Tests for Mixed Models." *Journal of Statistical Planning and Inference*, **61**, 249–277.

Bandelow B, Brunner E, Broocks A, Beinroth D, Hajak G, Pralle L, Rüther E (1998). "The Use of the Panic and Agoraphobia Scale in a Clinical Trial." *Psychiatry Research*, **77**, 43–49.

Bates D, Maechler M, Bolker B (2012). ***lme4****: Linear Mixed-Effects Models Using S4 Classes*. R package version 0.999999-0, URL http://CRAN.R-project.org/package=lme4.

Bathke AC, Schabenberger O, Tobias RD, Madden LV (2009). "Greenhouse-Geisser Adjustment and the ANOVA-Type Statistic: Cousins or Twins?" *The American Statistician*, **63**, 239–246.

Box GEP (1950). "Problems in the Analysis of Growth and Wear Curves." *Biometrics*, **6**, 362–389.

Breslow NE, Clayton DG (1993). "Approximate Inference in Generalized Linear Mixed Models." *Journal of the American Statistical Association*, **88**, 9–25.

Brunner E, Dette H, Munk A (1997). "Box-Type Approximations in Nonparametric Factorial Designs." *Journal of the American Statistical Association*, **92**, 1494–1502.

Brunner E, Domhof S, Langer F (2002). *Nonparametric Analysis of Longitudinal Data in Factorial Experiments*. John Wiley & Sons, New York.

Brunner E, Munzel U, Puri ML (1999). "Rank Score Tests in Factorial Designs with Repeated Measures." *Journal of Multivariate Analysis*, **70**, 286–317.

Brunner E, Puri ML (2001). "Nonparametric Methods in Factorial Designs." *Statistical Papers*, **42**, 1–52.

Carey VJ, Lumley T, Ripley BD (2012). ***gee****: Generalized Estimation Equation Solver*. R package version 4.13-18, URL http://CRAN.R-project.org/package=gee.

Diggle PJ, Liang KY, Zeger SL (1994). *Analysis of Longitudinal Data*. Oxford University Press, Oxford.

Erceg-Hurn DM, Mirosevich VM (2008). "Modern Robust Statistical Methods. An Easy Way to Maximize the Accuracy and Power of Your Research." *American Psychologist*, **63**, 591–601.

Halekoh U, Højsgaard S, Yan J (2006). "The R Package **geepack** for Generalized Estimating Equations." *Journal of Statistical Software*, **15**(2), 1–11. URL http://www.jstatsoft.org/v15/i02/.

Koch GG, Carr GJ, Amara IA, Stokes ME, Uryniak TJ (1990). "Categorical Data Analysis."

Konietschke F, Bathke AC, Hothorn LA, Brunner E (2010). "Testing and Estimation of Purely Nonparametric Effects in Repeated Measures Designs." *Computational Statistics & Data Analysis*, **54**, 1895–1905.

Lehmann E (2009). "Parametric versus Nonparametrics: Two Alternative Methodologies." *Journal of Nonparametric Statistics*, **21**, 397–405.

Liang KY, Zeger SL (1986). "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, **73**, 13–22.

Lumley T (1996). "Generalized Estimating Equations for Ordinal Data: A Note on Working Correlation Structures." *Biometrics*, **52**, 354–361.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Development Core Team (2012). ***nlme**: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-104, URL http://CRAN.R-project.org/package=nlme.

Potthoff RF, Roy SN (1964). "Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems." *Biometrika*, **51**, 313–326.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Robson C (2002). *Real World Research: A Resource for Social Scientists and Practitioner-researchers*. John Wiley & Sons, New York.

Rödig T, Sedghi M, Konietschke F, Lange K, Ziebolz D, Hülsmann M (2010). "Efficacy of Syringe Irrigation, RinsEndo®and Passive Ultrasonic Irrigation in Removing Debris from Irregularities in Root Canals with Different Apical Sizes." *International Endodontic Journal*, **43**, 581–589.

Romano J (2009). "Discussion of 'Parametric versus Nonparametrics: Two Alternative Methodologies'." *Journal of Nonparametric Statistics*, **21**, 419–424.

SAS Institute Inc (2003). *The SAS System, Version 9.1*. SAS Institute Inc., Cary, NC. URL http://www.sas.com/.

Schörgendorfer A, Madden LV, Bathke AC (2011). "Choosing Appropriate Covariance Matrices in a Nonparametric Analysis of Factorials in Block Designs." *Journal of Applied Statistics*, **38:4**, 833–850.

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. 4th edition. Springer-Verlag, New York.

Wolfinger RD (1996). "Heterogeneous Variance: Covariance Structures for Repeated Measures." *Journal of Agricultural, Biological, and Environmental Statistics*, **1**, 205–230.

Zeger SL, Liang KY (1992). "An Overview of Methods for the Analysis of Longitudinal Data." *Statistics in Medicine*, **11**, 1825–1839.

Affiliation:

Kimihiro Noguchi
Department of Statistics
University of California, Davis

Davis, California 95616, United States of America
E-mail: kinoguchi@ucdavis.edu

Yulia R. Gel
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario N2L 3G1, Canada
E-mail: ygl@math.uwaterloo.ca

Edgar Brunner, Frank Konietschke
Department of Medical Statistics
University of Göttingen
37073 Göttingen, Germany
E-mail: brunner@ams.med.uni-goettingen.de, fkoniet@gwdg.de

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 50, Issue 12 *Submitted:* 2011-02-15
September 2012 *Accepted:* 2012-08-22