



Article scientifique

Article

2010

Accepted version

Open Access

This is an author manuscript post-peer-reviewing (accepted version) of the original publication. The layout of the published version may differ .

---

## An exact permutation method for testing any effect in balanced and unbalanced fixed effect ANOVA

---

Kherad Pajouh, Sara; Renaud, Olivier

### How to cite

KHERAD PAJOUH, Sara, RENAUD, Olivier. An exact permutation method for testing any effect in balanced and unbalanced fixed effect ANOVA. In: Computational statistics & data analysis, 2010, vol. 54, p. 1881–1893. doi: 10.1016/j.csda.2010.02.015

This publication URL: <https://archive-ouverte.unige.ch/unige:12140>

Publication DOI: [10.1016/j.csda.2010.02.015](https://doi.org/10.1016/j.csda.2010.02.015)

# An Exact Permutation Method for Testing Any Effect in Balanced and Unbalanced Fixed Effect ANOVA

Sara Kherad Pajouh<sup>a,1</sup> and Olivier Renaud<sup>a</sup>

<sup>a</sup> Methodology & Data Analysis Group, Dept. of Psychology, Boulevard du Pont d'Arve 40,  
University of Geneva, Switzerland

postprint – Journal of Computational Statistics and Data Analysis 54 (2010)  
1881–1893

## Abstract

The ANOVA method and permutation tests, two heritages of Fisher, have been extensively studied. Several permutation strategies have been proposed by others to obtain a distribution-free test for factors in a fixed effect ANOVA (i.e., single error term ANOVA). The resulting tests are either approximate or exact. However, there exists no universal exact permutation test which can be applied to an arbitrary design to test a desired factor. An exact permutation strategy applicable to fixed effect analysis of variance is presented. The proposed method can be used to test any factor, even in the presence of higher-order interactions. In addition, the method has the advantage of being applicable in unbalanced designs (all-cell-filled), which is a very common situation in practice, and it is the first method with this capability. Simulation studies show that the proposed method has an actual level which stays remarkably close to the nominal level, and its power is always competitive. This is the case even with very small datasets, strongly unbalanced designs *and* non-Gaussian errors. No other competitor show such an enviable behavior.

## 1 Introduction

ANOVA is probably the single most used method in many fields of science, ranging from biology to psychology. Historically, a balanced design was preferable since the formulae in this case are much easier and can be handled by hand, even with a large number of observations. By balanced, we mean that each cell has exactly the same number of observations. With the use of statistical packages, this is not a concern anymore, and in many fields a balanced design is almost impossible to obtain, either because of (missing at random) drop-outs or by design. It is therefore useful to provide methods that also work for unbalanced designs. We will assume in what follows that there is at least one observation per cell (all-cell-filled). Analysing empty-cell designs is a complex topic that goes beyond the scope of this paper, and we refer to Searle (1987).

To test different effects or factors in an ANOVA, i.e., main effects and interaction effects, some authors advocate a sequential approach; first allocate the part of the explained variance to the main effects (one after the other), then to the two-way interaction(s) (one after the other) and then to increasingly and higher order interactions if present. In several statistical packages, this approach is called *sequential* or *Type I sum of squares*. Other authors advocate a related but not equivalent technique that we will call *step-down procedure*: Test first the higher-order interaction and stop the procedure if the interaction is significant; otherwise, proceed and test the effects at one lower level, using the simplified model that does not contain the higher-order interactions. The last approach, called *marginal* or *Type III sum of squares*, consists, for any factor to be tested (main effects, interaction effects), in comparing the full model versus the model that contains all the factors except the factor to be tested.

In balanced designs, the Type I and Type III sum of squares approaches are equivalent, but the results and their interpretation might differ in unbalanced cases. Searle (1987) shows that the null hypotheses tested differ depending on the type of sum of squares used. Each of these null hypotheses can be expressed as a linear combination of the cell means. Searle (1987) claims that, in unbalanced cases, the cell-mean hypothesis that very often meets the researcher's goals is the one linked to Type III sum of squares (see also the help files of StatSoft, 2007, for a nice summary). Again this is valid only in the all-cell-filled case. The method we introduce is correct for Type III sum of squares and for hierarchical models. Since for balanced designs all types of sum of squares are equivalent, the method applies to all types of sum of squares in this case. For unbalanced designs, some additional work would, however be needed to use Type I sum of squares.

It is very likely that the assumption that the error terms are distributed according to a Gaussian distribution does not hold in many applied datasets. In this case, the required conditions for a parametric ANOVA test are not satisfied and a test that does not depend on parametric hypothesis is desirable, at least as a check for the parametric test. One famous class of so-called distribution-free tests is permutation or randomization tests. Permutation tests date back to Fisher (1935), but given the power of computers, they have resurfaced in recent decades. Permutation tests have lots of applications not only in ANOVA but in different fields (Brombin and Salmaso, 2009, e.g.). For more details on the history of permutation tests, see David (2008).

Exchangeability under the null hypothesis is the only requirement for having an exact permutation test. Much stronger requirements, including the normality assumption of error terms, are necessary in (parametric) ANOVA. Also, let us mention that even when the conditions of parametric tests hold, a permutation test is almost as powerful as parametric tests (Welch and Gutierrez, 1988, e.g.). The definition of exchangeability (in the case of continuous variables) is that the joint density  $f$  satisfies  $f(z_1, z_2, \dots, z_n) = f(z_{\pi(1)}, z_{\pi(2)}, \dots, z_{\pi(n)})$  for any permutation  $\pi$  of the indices. This means that, under the null hypothesis, any permutation of the data is an equally probable outcome (Good, 2004). Then, the observed statistic (based on actual data) has to be compared to the empirical distribution of all the statistics obtained by permutations of the data. The proportion of statistics higher than the observed statistic gives the  $p$ -value (Pesarin, 2001).

The permutation test is well studied for the comparison of two samples (the Wilcoxon rank test is actually a special case of permutation tests). However, using a permutation test for more general ANOVA designs has attracted much less attention in the statistics literature, and there is no unique accepted approach for applying this class of the tests. First note that permutation tests can be classified as exact or approximate. A test is exact if, under the null hypothesis, the probability of type I error is equal to the nominal level of the test (Good, 1994; Anderson and Ter Braak, 2003).

In experimental designs with more than one factor, different strategies have been proposed. They can be divided into two groups: permutation of observations (or raw data) and permutation of different types of residuals. In this article we will focus on the permutation of residuals and mention the other approaches only for completeness. We refer to a comprehensive review of the literature (Anderson and Ter Braak, 2003) for more details. For permutation of raw data, there are three important approaches. The first is the unrestricted permutation of raw data (Manly, 1997). This is not an exact test, except in the simple case of one-way ANOVA (otherwise the effect of the other factor(s) makes the data not exchangeable). The second one is the restricted permutation of raw data. This was introduced by Good (1994, chapter 4) and Edgington (1995, chapter 6). It is an exact test when testing the main effects in the absence of interaction terms. However, for larger designs, the restrictions required leave no possible permutation and no test is available. The third one is the synchronized permutation test, which also relies on restrictions and was introduced by Salmaso (2003) for  $2^k$  design and Basso et al. (2007) for balanced two-way factorial designs. In Gonzalez and Manly (1998), different permutation strategies have been studied and compared. The simulation results show that no method is uniformly better than another over all kinds of designs. According to their simulations, permutation of raw data has relatively more power than the other approaches in some cases. Anderson and Ter Braak (2003) advocate the permutation of raw data. They present guidelines to determine the exchangeable units to be permuted to obtain an exact permutation test.

Concerning the permutation of residuals, some authors advise using permutation tests based on residuals under the full model (Ter Braak, 1992, e.g.). As its name suggests, one should first compute the residuals based on the full model, which removes the effects of all factors from the original observations. Then, one constructs a new sample or pseudo-data by adding the fitted value to an unrestricted permuted version of the residuals. Based on these new samples, one uses a statistic whose null hypothesis is that the true parameter is equal to its empirical value in the original sample (Ter Braak, 1992). Note that, with this approach, the same “new samples” are used for all the tests. The only difference is that the statistic used depends on which factor is tested. This test is approximate and closely related to one of the bootstrap approaches used for regression (Davison and Hinkley, 1997, e.g.).

Another approximate method is the permutation of residuals under the reduced model (Still and White, 1981; Freedman and Lane, 1983). In this method, the residuals are computed from a model that includes all parameters except the parameter(s) of interest. This means that only the effect of the parameters that are of interest (that are tested) are possibly present in these residuals. However, under the null hypothesis, there should be no systematic effect in these residuals, and therefore an unrestricted permutation of these residuals is used to test the parameters of interest. Anderson and Ter Braak (2003) showed that this method has relatively more power in comparison with the other methods.

The technique we provide in this paper is permutation of residuals under a modified model, which is an exact permutation method based on the residuals under the reduced model. The method will be fully explained in Section 2.2. It is important to mention that we are considering only single error term models. The idea is to remove that part of the design matrix that is not to be tested, much like the residual maximum likelihood (REML) approach in mixed effect models. It is based on an article by Jung et al. (2006). However our formulation generalizes their approach in three important directions. (1) We have a general formula that is applied to any ANOVA design, whereas they provided formulae only for three particular cases. (2) It can be applied to all factors in the ANOVA, for example to test a main effect in the presence of interaction, while their formulae apply only for the highest-order factor. (3) Our formula can be applied to unbalanced designs, whereas they rely on balanced cases. In addition, note that our formulation is actually not restricted to ANOVA and can be applied in regression settings as well. Huh and Jhun (2001) use also this type of approach in regression, but only for a univariate hypothesis (they do also mention a multivariate test but it is actually based on residuals under the full model). In the case of ANOVA and hierarchical regression, a method that can deal with multivariate hypotheses is necessary, as is provided in the present article.

An important advantage of this exact method compared to exact methods using restriction of raw data is that this method is also applicable to designs with small sample sizes, because the proposed method will rely on a much larger number of possible permutations. Note that this method uses a separate permutation method to test each factor of interest.

The structure of the paper is the following. Section 2 presents the parametric tests in ANOVA. Then we introduce an approximate method of permutation of residuals under the reduced model and show that since the exchangeability assumption does not hold, it is not an exact test. In the next part, the residuals under the modified model are introduced, and it is shown that they satisfy the exchangeability condition, implying that they lead to an exact permutation test. Section 3 shows a geometric illustration of the method. The construction of one important ingredient of the method is also provided. The details for one specific ANOVA design are made explicit in Section 4. Finally, Section 5 is devoted to a simulation study to evaluate the level and the power of the proposed method and competitors.

## 2 The method

The fully general model for regression and analysis of variance with single error term (ANOVA) can be written as

$$y = X\beta + \varepsilon, \quad (1)$$

where  $y$  is the vector of the explained variable,  $X$  is the design matrix,  $\beta$  the vector of parameters and  $\varepsilon$  the vector of unknown errors with mean 0 and variance  $\sigma^2$ . More details about the conditions for the error term will be discussed in the following sections. For ANOVA, we will suppose that all cells have at least one observation, a so-called all-cell-filled design (Searle, 1987, see).

In regression,  $X$  is of full rank, but this is not necessarily the case for ANOVA. Although they have some advantages and advocates in other settings (Kirk, 1994; Searle, 1987), the approaches using an *overparameterized* design matrix or *cell-mean model* design matrix will not be used here. The following theory is much easier using *sigma-restricted* (SR) parameterization for the design matrix, which contains all the necessary information on the model. In this case the design matrix is of full rank, which implies that normal equations can be inverted without generalized inverse and the degrees of freedom correspond to the number of columns (see below). Moreover, the subspace spanned by the columns of interest is directly the projection subspace (whereas the subspace spanned by an interaction term in the overparameterized parameterization contains the main effects as well).

An easy way to obtain the SR design matrix is to code the level on a set of (not necessarily orthogonal) contrasts like Helmert or “treatment”. The number of columns corresponding to each factor is equal to the number of degrees of freedom of that factor. For example, for the “treatment” contrasts, each element gets the values 1, 0 and -1. Consider for instance a main effect  $A$ , with four levels. In the design matrix, there are three columns corresponding to this factor. The first column takes the value 1 for the observations which are in the first level of  $A$ , -1 if they are in the last level of  $A$ , and 0 if they are in the other two levels. The same idea applies to the two other columns, but switching the first with either the second or the third level (Cardinal and Aitken, 2006, see).

In the following we will suppose that  $X$  is of full rank, filled either with covariates or based on the *sigma-restricted* (SR) parameterization. For any test of interest, we can divide the design matrix  $X$  and the vector of parameters  $\beta$  into the component not of interest and the component of interest:

$$X = [X_1 | X_2], \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix},$$

and then write Eq. (1) as  $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ . The dimension of  $X$  is  $N \times p$ .  $X_1$  is  $N \times q$  and corresponds to the columns of  $X$  that are not of interest for the particular test, and  $X_2$  is  $N \times (p - q)$  and corresponds to the tested part of  $X$ . Also  $y$  and  $\varepsilon$  are vectors of size  $N \times 1$ . Similarly,  $\beta$  is a vector that includes  $p$  parameters and  $\beta_1$  and  $\beta_2$  are of dimensions  $q$  and  $(p - q)$ , respectively. The corresponding hypotheses are

$$H_0 : \beta_2 = 0 \quad \text{vs.} \quad H_1 : \beta_2 \neq 0. \quad (2)$$

If the null hypothesis is not rejected, then the explanatory variable(s) in  $X_2$  might not be necessary and the simpler model in this case would be  $y = X_1\beta_1 + \varepsilon$ .

### 2.1 Parametric approach

If the errors are supposed Gaussian, the following theorem gives the statistic for the well-known  $F$ -test in ANOVA, which is equivalent to the  $t$ -test or Wald test in regression and the  $F$ -test in hierarchical regression.

**Theorem 1** *The  $F$ -test statistic for the hypothesis on Eq. (2) for testing  $\beta_2$  is given by*

$$F_{ss} = \frac{(\text{RSS}(\hat{\beta}_1) - \text{RSS}(\hat{\beta})) / (p - q)}{\text{RSS}(\hat{\beta}) / (N - p)}, \quad (3)$$

where  $\text{RSS}(\cdot)$  is the residual sum of squares for the model with the corresponding parameters. If the errors are Gaussian,  $F_{ss}$  is distributed as  $(p - q, N - p)$ .

The proof of the theorem can be found in many books on regression or ANOVA (Davison and Hinkley, 1997, e.g.). To introduce our proposal, we first need to write this statistic in another form, as shown in the next theorem.

**Theorem 2**  *$F_{ss}$  is equal to  $F_R$ , defined as*

$$F_R = \frac{y'(X_{rr}(X'_{rr}X_{rr})^{-1}X'_{rr})y / (p - q)}{y'(I - X(X'X)^{-1}X')y / (N - p)}, \quad (4)$$

where

$$X_{rr} = (I - H_1)X_2, \quad \text{with } H_1 = X_1(X'_1X_1)^{-1}X'_1. \quad (5)$$

Note that in a fully balanced case (with SR parameterization),  $X_{rr} = X_2$ .

The proof of this theorem can be found in the Appendix. In the following, we show how to transform the data so that they can be permuted to obtain a genuine permutation test, while still using the same statistic.

## 2.2 Permutation of residuals under the reduced model

As already stated, the condition to obtain a permutation test is to handle exchangeable objects. The elements in  $y$  are certainly not exchangeable. Indeed, under the null hypothesis (2),  $\mathbb{E}(y_i) = \{X_1\beta_1\}_i$ , which differs depending on  $i$ .

Using the residuals under the reduced model solves this first obstacle, as shown next. In a very similar way to the residual maximum likelihood (REML) approach in mixed effect models, we remove the part of the design matrix which is not of interest for the test. This can be done by premultiplying both sides of Eq. (1) by  $I - H_1$ , and we obtain the reduced model

$$y_{rr} = X_{rr}\beta_2 + \varepsilon_{rr}, \quad (6)$$

where  $y_{rr} = (I - H_1)y$ ,  $X_{rr} = (I - H_1)X_2$  and  $\varepsilon_{rr} = (I - H_1)\varepsilon$ . We will call  $y_{rr}$  the residuals under the reduced model or briefly the reduced residuals (hence the subscript  $rr$ ). Note that, in the balanced case, since  $X'_1X_2 = 0$ , we have that  $X_{rr} = X_2$ .

We will show in Theorem 5 that the usual  $F$ -statistic (3) can be written based on the reduced residuals  $y_{rr}$  as follows:

$$F_{rr} = \frac{y'_{rr}(X_{rr}(X'_{rr}X_{rr})^{-1}X'_{rr})y_{rr} / (p - q)}{y'_{rr}(I - X(X'X)^{-1}X')y_{rr} / (N - p)}. \quad (7)$$

Under the null hypothesis (2), it is easy to see that, even with i.i.d. assumptions on the error term,  $y_{rr}$  is distributed as

$$y_{rr} = \varepsilon_{rr} \sim (0, \sigma^2(I - H_1)). \quad (8)$$

All elements of  $y_{rr}$  having the same expectation under the null, one might think that a permutation scheme would be to permute the elements in the vector of residuals directly, say  $y_{rr}^*$ , and plug them in the formula of  $F_{rr}$  to obtain

$$F_{rr}^* = \frac{y_{rr}^{*'}(X_{rr}(X'_{rr}X_{rr})^{-1}X'_{rr})y_{rr}^* / (p - q)}{y_{rr}^{*'}(I - X(X'X)^{-1}X')y_{rr}^* / (N - p)}. \quad (9)$$

The  $p$ -value is defined as  $(\#F_{rr}^* \text{ such that } F_{rr}^* \geq F_{rr})/(\#F_{rr}^*)$ , for all or at least a large number of permutations. However, exchangeability does not concern only expectations, but the entire joint distribution, including variances and covariances and higher-order moments. From Eq. (8), we see that  $y_{rr}^*$ , a randomly permuted vector of  $y_{rr}$ , is distributed differently from  $y_{rr}$ . The exchangeability is not met and therefore the corresponding test would not be an exact permutation test. The next section shows the conditions needed to obtain exchangeable observations.

### 2.3 Permutation of residuals under the modified model

We explain how to obtain uncorrelated (transformed) observations to achieve an exact permutation test. Since the matrix  $X_1$  has full rank  $q$ , there exists a matrix  $V$  of dimension  $N \times (N - q)$  whose columns form an orthonormal basis for the subspace orthogonal to  $(X_1)$ . Orthonormality of the columns of  $V$  implies  $V'V = I_{N-q}$ . On the other hand, the subspaces spanned by  $X_1$  and  $V$  are complementary by construction, and hence, the sum of the corresponding projections equal to the identity matrix, i.e.  $I = X_1(X_1'X_1)^{-1}X_1' + V(V'V)^{-1}V'$ . Therefore, the matrix  $V$  satisfies the following equations simultaneously:

$$VV' = I_N - H_1 \quad \text{and} \quad V'V = I_{N-q}. \quad (10)$$

The construction of  $V$  is explained in Section 3.2. Premultiplying by  $V'$  both sides of (6), we get the modified model as follows:

$$y_{mr} = X_{mr}\beta_2 + \varepsilon_{mr} \quad (11)$$

where  $y_{mr} = V'y_{rr}$  are the residuals under the modified model, or briefly the modified residuals,  $X_{mr} = V'X_{rr}$  and  $\varepsilon_{mr} = V'\varepsilon_{rr}$ . Note that unlike  $y$  and  $y_{rr}$  which are of dimension  $N$ , the dimension of  $y_{mr}$  is  $N - q$ . The variance of  $\varepsilon_{mr}$  is  $(\varepsilon_{mr}) = (V'\varepsilon_{rr}) = V'(\sigma^2(I - H_1))V = \sigma^2V'VV'V = \sigma^2I_{N-q}$ . Hence, under the null hypothesis (2),

$$y_{mr} = \varepsilon_{mr} \sim (0, \sigma^2I_{N-q}). \quad (12)$$

The following two theorems give sufficient conditions for the distribution of the error on the original equation to have (weak/strong) exchangeable error in the modified model.

**Theorem 3** *If the elements of the original model error  $\varepsilon$  are exchangeable, then the modified model errors,  $\varepsilon_{mr}$ , are weakly exchangeable; that is, they are exchangeable up to the second moment. In particular, if the elements of  $\varepsilon$  are i.i.d., then the elements of  $\varepsilon_{mr}$  are weakly exchangeable.*

**Theorem 4** *If the (joint) distribution of the original model error  $\varepsilon$  is spherical, then the modified model errors,  $\varepsilon_{mr}$ , are strongly exchangeable. More formally, if  $\varepsilon$ , has a spherical density, i.e.,  $f_\varepsilon(\cdot)$ , the (joint) probability density function (p.d.f.) of  $\varepsilon$  satisfies*

$$f_\varepsilon(w) = f_\varepsilon(Tw),$$

*for any unitary matrix  $T$  and vector  $w$ , then the modified model errors  $\varepsilon_{mr}$  are exchangeable; that is, the joint p.d.f. of  $\varepsilon_{mr}$  remains fixed under any permutation of its elements, i.e.,*

$$f_{\varepsilon_{mr}}(u) = f_{\varepsilon_{mr}}(Pu),$$

*for any permutation matrix  $P$ , and vector  $u$*

The proofs of Theorem 3 and Theorem 4 are given in the Appendix.

One can obtain a new statistic based on the modified residuals. That is the  $F$ -statistic:

$$F_{mr} = \frac{y'_{mr}(X_{mr}(X'_{mr}X_{mr})^{-1}X'_{mr})y_{mr}/(p - q)}{y'_{mr}(I_{N-q} - X_{mr}(X'_{mr}X_{mr})^{-1}X'_{mr})y_{mr}/(N - p)}, \quad (13)$$

where  $y_{mr}$  is defined in (11). The following theorem gives the justification of the proposed method.

**Theorem 5** *For the observed data, the three statistics defined in Eqs. (3), (7) and (13) are equivalent:*

$$F_{mr} = F_{rr} = F_{ss}. \quad (14)$$

The proof can be found in the Appendix. Now, we have all the ingredients to make an exact permutation test: obtain  $y_{mr}^*$  which is a permuted version of  $y_{mr}$ , and replace it in  $F_{mr}$  with the suitable  $F$ -statistic for permuted modified residuals:

$$F_{mr}^* = \frac{y_{mr}^{*'}(X_{mr}(X_{mr}'X_{mr})^{-1}X_{mr}')y_{mr}^*/(p-q)}{y_{mr}^{*'}(I_{N-q} - X_{mr}(X_{mr}'X_{mr})^{-1}X_{mr}')y_{mr}^*/(N-p)}.$$

The  $p$ -value is defined as  $(\#F_{mr}^* \text{ such that } F_{mr}^* \geq F_{mr})/(\#F_{mr}^*)$ , for all or at least a large number of permutations. The following theorem summarizes the main result of this paper.

**Main result.** *In linear models  $y = X\beta + \varepsilon$ , as in Eq. (1), with  $X$  of full rank (regression or ANOVA), many testing hypotheses can be written as in (2). This includes testing an interaction in balanced and unbalanced designs and testing a main effect in the presence of higher-order interactions (equivalent to the Type III sum of squares). With the conditions for the error term given in Theorems 3 and 4, an exact permutation test can be computed using the permutation of the residuals under the modified model ( $y_{mr}$ ) and the  $F_{mr}$ -statistic given in Eq. (13).*

### 3 Some additional points about the new method

#### 3.1 Graphical view of the proposed method

The graphical view depicted in Fig. 1 can help in understanding Theorem 5 and the link between the different statistics. In the graph, the lines with the same color and length are parallel;  $X_1$  and  $X_2$  are one-dimensional;  $y$ ,  $X_1$  and  $X_2$  are displayed as vectors in  $\mathbb{R}^N$ , here  $\mathbb{R}^3$ . A small arrow represents an orthogonal projection and (a) segment(s) represent(s) the subspace into which the projection is done.  $y_{rr}$  is the projection into the space orthogonal to  $X_1$ .

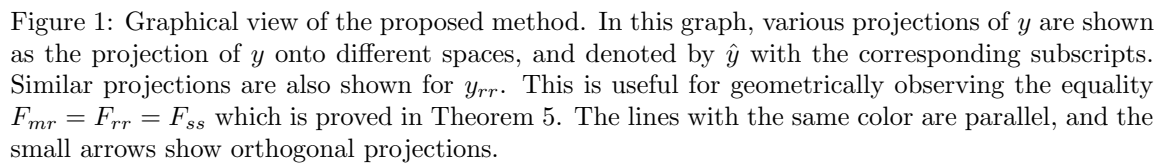
Concerning Eq. (3),  $RSS(\beta)$ ,  $RSS(\beta_1)$  are the squared distances of the vectors of residuals  $y - \hat{y}_{X_1}$  and  $y - \hat{y}_{X_1, X_2}$ . In other words  $RSS(\beta_1)$  and  $RSS(\beta)$  are the squared distances between  $y$  and the line generated by  $X_1$ , respectively the subspace generated by  $X_1$  and  $X_2$ . By Pythagoras' theorem, the difference between the two is  $SS(X_2)$ , also represented, which is the numerator of (3). The denominator of (4) corresponds to the same vector as the denominator of (3). Its numerator corresponds to the squared norm of the projection of  $y$  onto  $X_{rr}$ , named  $SS_{rr}(X_2)$ , and the graphical representation shows that it is of same length as  $SS(X_2)$ . Finally, given that  $y_{rr}$  is the projection into the orthogonal space of  $X_1$ , Eq. (7) relies on  $RSS_{rr}(\beta)$  and  $SS_{rr}(X_2)$ , which have the same lengths as  $RSS(\beta)$  and  $SS(X_2)$ , respectively. Concerning Eq. (13), since  $y_{rr}$  belongs to a subspace orthogonal to  $X_1$ , and  $V$  just makes a change of basis in this subspace, the distances will remain the same.

It is worth mentioning that the position of  $y_{mr}$  is the same as that of  $y_{rr}$ , but in a reduced dimension space. Hence a similar argument shows that  $F_{mr}$  is also equivalent to  $F_{ss}$ .

#### 3.2 Construction of the matrix $V$

In this section we explain the construction of the matrix  $V$ . This construction relies on the eigen-decomposition of  $(I - H_1)$ . Since  $(I - H_1)$  is a symmetric and idempotent matrix, it possesses only two eigenvalues: 0 and 1. Let  $(I - H_1) = UDU'$  be the eigen-decomposition of  $(I - H_1)$ , where  $D$  is a diagonal matrix containing the eigenvalues of  $(I - H_1)$ , and  $U$  is a unitary matrix, whose columns are the eigenvectors of  $(I - H_1)$ . Since  $(I - H_1)$  has rank  $N - q$ , there are exactly  $N - q$  ones and  $q$  zeros in the diagonal of  $D$ . It is easy to show that both conditions in (10) are satisfied by choosing  $V$  as the columns of  $U$  corresponding to the non-zero diagonal elements of  $D$ . It turns out that the





It is worth mentioning that there is an  $(N - q)$ -dimensional subspace which is mapped to itself under the projection induced by  $I - H_1$ . Any orthonormal basis of such subspace can be chosen as the eigenvectors of  $I - H_1$  corresponding to the non-zero eigenvalues. However, different choices of such basis vectors result in different performance of the test in terms of level and power. Therefore, it is worth considering some specifications for choices which give better performance. Assume that  $p - q = 1$ , i.e. that  $X_2$  and thus  $X_{mr}$  are one-dimensional. In this case, the  $F_{mr}$  statistic is equivalent to (a monotonic transformation of)  $CF_{mr} = (y'_{mr} \cdot X_{mr})^2$ .

- These two examples show that, if the matrix  $V$  is such that the statistic can take many different values, it can be considered as a suitable  $V$ .

- postprint – Journal of Computational Statistics and Data Analysis 54 (2010) 1881–1893

$$y_{mr} = V'y_{rr}.$$

Consider a model under the alternative hypothesis. In order to have a large power, we need the  $p$ -values to take in most of the cases a small value, i.e.,  $CF_{mr}^* < CF_{mr}$  holds with high probability. Ignoring the noise part of the observations, the statistics can be written as

$$|y'_{mr} X_{mr}| \simeq |\beta'_2 X'_{mr} X_{mr}| = |\beta_2| |X_{mr}|^2,$$

and

$$|y'^*_{mr} X_{mr}| \simeq |\beta'_2 X'^*_{mr} X_{mr}| = |\beta_2| |X^*_{mr}| |X_{mr}| |\cos \theta^*| = |\beta_2| |X_{mr}|^2 |\cos \theta^*|,$$

where  $\theta^*$  is the angle between the vectors  $X^*_{mr}$  and  $X_{mr}$ , and where we have used the fact that  $|X_{mr}| = |X^*_{mr}|$  in the last equality. Therefore, a good choice of  $V$  should give us  $|\cos \theta^*| \ll 1$ , or  $\theta^*$  close to  $\pm\pi/2$  for many permutations. For example, if  $V$  is such that all the components of  $X_{mr}$  share the same sign (positive or negative), this condition is not satisfied.

A good, reasonable choice for  $V$  is one that results in a  $X_{mr}$  that has negative and positive components which are all different. It is advised to check it before running the test.

## 4 Example

Consider a fixed effect balanced two-way ANOVA design with the following model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad i = 1 \dots a, \quad j = 1 \dots b, \quad k = 1 \dots n, \quad (15)$$

where  $\alpha_i$  and  $\beta_j$  are the main effects of (fixed) factors  $A$  and  $B$ ,  $\mu$  is the overall mean,  $(\alpha\beta)_{ij}$  is the interaction effect in the  $ij$ th cell and  $\varepsilon_{ijk}$  is the unknown error associated with observation  $y_{ijk}$ . Imagine we want to test a main effect. Some statisticians recommend to first check that there is no (significant) interaction effect, and then to test the main effect in the model with no interaction. However, applied researchers often want to test all effects at once, and Searle (1987) showed that this approach based on marginal tests corresponds to the following genuine hypotheses:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a \quad \text{vs.} \quad H_1 : \text{not } H_0. \quad (16)$$

The well-known statistic to test  $H_0$  is

$$F_A = \frac{\sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 / (a-1)}{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2 / (ab(n-1))},$$

where a dot represent an average over the index. Under Gaussian assumptions, the distribution of  $F_A$  under the null is  $(a-1, ab(n-1))$ .

Here, we use reduced residuals in order to test the main effect  $A$ . The advantage of using reduced residuals is that the effects of the non-interesting factors, i.e., the overall mean and the factors  $B$  and  $AB$ , are removed from  $y_{ijk}$ . They are therefore defined as

$$\begin{aligned} y_{(rr)ijk} &= y_{ijk} - \bar{y}_{...} - (\bar{y}_{.j.} - \bar{y}_{...}) - (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) \\ &= y_{ijk} - \bar{y}_{ij.} + \bar{y}_{i..} - \bar{y}_{...} \end{aligned} \quad (17)$$

By direct computation, we can see that the above  $F$ -statistic can be rewritten as

$$F_{(rr)A} = \frac{\sum_{i=1}^a (\bar{y}_{(rr)i..} - \bar{y}_{(rr)...})^2 / (a-1)}{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{(rr)ijk} - \bar{y}_{(rr)ij.})^2 / (ab(n-1))}.$$

The equivalent matrix formulation is obtained by first defining the design matrix  $X$  as

$$\left[ \underbrace{\begin{bmatrix} \vdots \\ X_0, X_b, X_{ab} \\ \vdots \end{bmatrix}}_{X_1} \quad \underbrace{\begin{bmatrix} \vdots \\ X_a \\ \vdots \end{bmatrix}}_{X_2} \right], \quad (18)$$

using a sigma-restricted parameterization of the factors  $A$ ,  $B$ , and  $AB$ . Since the design is balanced, the columns of  $X_0$ ,  $X_b$ ,  $X_{ab}$ , and  $X_a$  are orthogonal. So, by explicit computation of Eq. (3), it is easy to show that  $F_{ss}$  is equivalent to  $F_A$ . Similarly, if we define  $y_{rr} = (I - H_1)y = (I - X_1(X_1'X_1)^{-1}X_1')y$ , this leads to the same equation as (17). Moreover, Eq. (17) clearly shows that two  $y_{(rr)ijk}$  are correlated differently depending on whether they share a  $i$  and/or a  $j$ . Therefore the exchangeability condition does not hold and even in this simple balanced two-way ANOVA we have to resort to the computation of the residuals under the modified model  $y_{mr}$ , through the use of the  $V$  matrix as explained in Section 2.3.

It is important to note that the use of the sigma-restricted parameterization allows us to test any factor, in particular to test a main effect in the presence of interactions. Jung et al. (2006)'s formulation did not allow this. A second advantage of our matrix formulation is that it generalizes to the unbalanced case, which again was not present in the cited article.

Finally, Theorem 5 states that  $F_{mr} = F_{rr} = F_{ss}$  are equal to  $F_J$  (see in the proof, Eq. (33)) for original observations. However, not all statistic are suited for permutation. Jung et al. (2006) propose using  $F_J$ . A simulation study (not shown here) provided evidence that the level of this permutation test can be far from the nominal level. One reason to explain this fact is that  $y_{rr}$  is in a subspace spanned by  $I - H_1$  but  $y_{rr}^*$  is not restricted to lie in this subspace, and thus, for the permuted  $y_{rr}^*$ ,  $F_J$  does not project onto the correct space, as does  $F_{rr}$ .

## 5 Simulation study

In this part, three simulation studies are carried out to explore the performance of the new exact method concerning level and power. In the first one, we consider testing the interaction in a two-way balanced ANOVA with two observations per cell and two levels for  $\alpha$  and  $\beta$ .

To assess the level and power, we need somehow to standardize the effect of interest. Here we investigate the interaction, which must satisfy the sigma-restriction condition  $\sum_{i=1}^a(\alpha\beta)_{ij} = \sum_{j=1}^b(\alpha\beta)_{ij} = 0$ . For comparison of the power, we consider a variation measure for the gradual increase for the parameter of interest of  $(\alpha\beta)$  as  $\theta = \left(\sum_{i,j=1}^{a,b}((\alpha\beta)_{ij})^2/(ab)\right)^{1/2}$ . In the simulations, the interaction that will be used is  $(\alpha\beta)_{ij} \cdot t/\theta$ , where  $t$  is a constant that changes the parameter of variations according to the level and power of interest. This means that for estimation of the significance level,  $t$  is equal to 0, which makes the effects of  $\alpha\beta$  equal zero, and for the estimation of power it makes changes in the original  $\alpha\beta$  according to the new variations of  $t > 0$ . After generating the values in the last part, we add the error terms which is generated from different distributions, namely  $N(0, 1)$ ,  $\exp(1) - 1$ ,  $t(4)$  and  $U(-\sqrt{3}, \sqrt{3})$ .

The new exact method (denoted  $Y_{mr}$ ) will be compared with three other methods, which are the Normal-theory  $F$ -test (denoted  $F$ ) and two approximate permutation tests methods: unrestricted permutation of residuals under the reduced model or permutation of  $y_{rr}$  (see Eq. (9), denoted  $Y_{rr}$ ) and unrestricted permutation of raw data or permutation of  $y_{ijk}$ , using the statistic (3) (denoted  $Y$ ). The permutation tests are based on 1000 permutations and comparison of the level and power of the methods is based on 5000 Monte Carlo replications. The results are given in Table 1 for different values of  $t$ , including  $t = 0$ , representing the level.

The second study concerns a balanced three-way ANOVA. In this case the model considered is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijk}$$

$$i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c$$

In the simulations, we choose  $a = 2$ ,  $b = 3$ ,  $c = 4$  and  $n = 2$ , and the same four different noises are added. The results are given in Table 2 for the level of the test for  $\beta\gamma$  and Table 3 for the level of the test for  $\gamma$ . The results show that all methods are close to the nominal  $\alpha = 0.05$ . This is probably due to the fact that the total number of observations is larger than in the first study.

The last study is carried out for an unbalanced two-way ANOVA model, as presented in Table 4. The data are generated as in the first study, except that the number of observations for the four

Table 1: Simulation results for a balanced two-way ANOVA with  $a = 2$ ,  $b = 2$ , and  $n = 2$  to test the  $\alpha\beta$  interaction for different values of  $t$ . The column with  $t = 0$  corresponds to the level, and the other ones to power. In bold, values that are significantly different from the type I error, which is set to 0.05.

	stat.	$t = 0$	$t = 0.5$	$t = 1$	$t = 1.5$	$t = 2$	$t = 2.5$
$\mathcal{N}(0, 1)$	$Y_{mr}$	0.0493	0.1843	0.5460	0.8413	0.9533	0.9918
	$Y_{rr}$	<b>0.0318</b>	0.1373	0.4590	0.7857	0.9460	0.9910
	$Y$	<b>0.0615</b>	0.2308	0.6260	0.8942	0.9815	0.9988
	F-test	0.0510	0.1908	0.5770	0.8782	0.9828	0.9998
$U(-\sqrt{3}, \sqrt{3})$	$Y_{mr}$	0.0500	0.0800	0.2867	0.6623	0.9060	0.9969
	$Y_{rr}$	<b>0.0390</b>	0.0647	0.2483	0.6130	0.8977	0.9917
	$Y$	<b>0.0737</b>	0.1083	0.3743	0.7828	0.9733	0.9999
	F-test	0.0537	0.0870	0.3160	0.7337	0.9737	0.9902
$\exp(1) - 1$	$Y_{mr}$	0.0507	0.2450	0.6147	0.8270	0.9019	0.9928
	$Y_{rr}$	<b>0.0330</b>	0.2217	0.5717	0.5717	0.9270	0.9910
	$Y$	<b>0.0647</b>	0.3347	0.6897	0.8870	0.9573	0.9912
	F-test	0.0520	0.0520	0.6007	0.8193	0.8983	0.9902
$t(4)$	$Y_{mr}$	0.0450	0.1543	0.7060	0.8600	0.8833	0.9440
	$Y_{rr}$	<b>0.0254</b>	0.1170	0.6080	0.7830	0.8460	0.9050
	$Y$	<b>0.0618</b>	0.1827	0.7220	0.8540	0.9115	0.9460
	F-test	<b>0.0429</b>	0.1372	0.6400	0.8250	0.8628	0.9390

Table 2: Simulation results for the level in a balanced three-way ANOVA with  $a = 2$ ,  $b = 3$ ,  $c = 4$  and  $n = 2$  to test the  $\beta\gamma$  interaction. The type I error is set to 0.05.

Three-way	$Y_{mr}$	$Y_{rr}$	$Y$	F-test
$\mathcal{N}(0, 1)$	0.0503	0.0493	0.0497	0.0509
$U(-\sqrt{3}, \sqrt{3})$	0.0516	0.0551	0.0529	0.0547
$\exp(1) - 1$	0.0490	0.0462	0.0470	0.0488
$t(4)$	0.0546	0.0542	0.0524	0.0500

Table 3: Simulation results for a balanced three-way ANOVA with  $a = 2$ ,  $b = 3$ ,  $c = 4$  and  $n = 2$  to test the main effect of  $\gamma$ . The type I error is set to 0.05.

Three-way	$Y_{mr}$	$Y_{rr}$	$Y$	F-test
$\mathcal{N}(0, 1)$	0.0503	0.0523	0.0547	0.0510
$U(-\sqrt{3}, \sqrt{3})$	0.0516	0.0542	0.0459	0.0523
$\exp(1) - 1$	0.0526	0.0556	0.0543	0.0514
$t(4)$	0.0542	0.0552	0.0568	0.0534

cells is  $[2, 4; 2, 4]$  (left part of Table 4) or  $[1, 3; 1, 3]$  (right part of Table 4). The level is given in the column with  $t = 0$  and one value of power in the column with  $t = 1$ .

In both cases, the new method is the only one that keeps a level close to the nominal size. The parametric  $F$ -test can have a level above 10% and the other permutation methods can be even higher. In the case when several methods have the correct level, the new method is also quite competitive in terms of power, which shows its good efficiency.

It is important to mention that in using the permutation of residuals under the modified model, we had to use a separate permutation method to test each factor of interest. In other words, we cannot test different factors simultaneously. For any given design, the matrix  $X$  is the same, but depending on the parameter of interest to test,  $X_1$  and  $X_2$  (Eq. (2)) differ. Computationally, for each parameter of interest,  $(I - H)$ ,  $V$  and  $y_{mr}$  have to be computed (once). Then,  $y_{mr}^*$  and  $F_{mr}^*$  are computed for each permutation. As a comparison, for permutation of residuals under the reduced model, the same computation has to be carried out, except the step involving  $V$ . For the permutation of raw data, only the permutation step is present. However, it seems that the permutation is the most computationally demanding step, so there are no large differences in computation time in practice between the three approaches.

## 6 Conclusion

A general formula to test any factor in balanced and unbalanced designs with non-empty cell is presented. It can be applied to any fixed effect ANOVA design or single error term ANOVA. In this paper, we modify an approximate test and provide an exact permutation strategy based on permutation of residuals under the reduced model. The main idea is to remove the correlation between residuals, using the decomposition of an idempotent matrix. This makes the residuals exchangeable. It is shown that if the error terms are i.i.d. or exchangeable, the modified residuals are exchangeable up to the second moment, and under the spherical condition of the error terms, the modified residuals are strongly exchangeable.

Simulation studies compare the significance level and power of the new method to the parametric method. We also extended our numerical comparison to two other approximate methods, namely, permutation of raw data and permutation of residuals under the reduced model, using several different error distributions. Results of these simulations show that the proposed method has a level which stays remarkably close to the nominal level and its power is always competitive. This is the case even with very small data sets, strongly unbalanced design *and* non-Gaussian errors. No other tested method shares this enviable property.

Table 4: Simulation results for the level in an unbalanced two-way ANOVA with  $a = 2$ ,  $b = 2$ , to test the  $\alpha\beta$  interaction for different values of  $t$ . In bold, values that are significantly different from the Type I error, which is set to 0.05.

Error dist.	stat.	[2,4;2,4]		[1,3;1,3]	
		$t = 0$	$t = 1$	$t = 0$	$t = 1$
$\mathcal{N}(0, 1)$	$Y_{mr}$	0.0510	0.784	0.0482	0.442
	$Y_{rr}$	<b>0.0612</b>	0.833	<b>0.0720</b>	0.560
	$Y$	<b>0.0624</b>	0.812	<b>0.0740</b>	0.544
	F-test	0.0504	0.802	0.0490	0.460
$U(-\sqrt{3}, \sqrt{3})$	$Y_{mr}$	0.0500	0.830	0.0460	0.387
	$Y_{rr}$	0.0510	0.867	<b>0.0645</b>	0.512
	$Y$	<b>0.0402</b>	0.852	<b>0.0650</b>	0.495
	F-test	0.0518	0.850	0.0472	0.404
$\exp(1) - 1$	$Y_{mr}$	0.0528	0.282	0.0564	0.542
	$Y_{rr}$	<b>0.1246</b>	0.654	<b>0.1112</b>	0.639
	$Y$	<b>0.1720</b>	0.728	<b>0.1122</b>	0.654
	F-test	<b>0.1020</b>	0.572	<b>0.0824</b>	0.560
$t(4)$	$Y_{mr}$	0.0482	0.602	0.0502	0.324
	$Y_{rr}$	0.0452	0.613	<b>0.0814</b>	0.438
	$Y$	0.0454	0.610	<b>0.0830</b>	0.417
	F-test	0.0483	0.603	0.0542	0.339

## References

- Anderson, M. and C. Ter Braak (2003). Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation* 73(2), 85–113.
- Basso, D., M. Chiarandini, and L. Salmaso (2007). Synchronized permutation tests in replicated  $I \times J$  designs. *Journal of Statistical Planning and Inference* 137(8), 2564–2578.
- Brombin, C. and L. Salmaso (2009). Multi-aspect permutation tests in shape analysis with small sample size. *Computational Statistics and Data Analysis* 53(12), 3921–3931.
- Cardinal, R. and M. Aitken (2006). *ANOVA for the Behavioural Sciences Researcher*. Lawrence Erlbaum Associates.
- David, H. (2008). The beginnings of randomization tests. *The American Statistician* 62(1), 70–72.
- Davison, A. and D. Hinkley (1997). *Bootstrap methods and their applications*. Cambridge, Cambridge University Press.
- Edgington, E. (1995). *Randomization Tests*. CRC Press.
- Fisher, R. (1935). The design of experiments. *Oliver & Boyd Edinburgh, Scotland*.
- Freedman, D. and D. Lane (1983). A nonstochastic interpretation of reported significance levels. *Journal of Business and Economic Statistics* 1(4), 292–98.
- Gonzalez, L. and B. Manly (1998). Analysis of variance by randomization with small data sets. *Environmetrics* 9(1), 53–65.
- Good, P. (1994). *Author Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer.
- Good, P. (2004). *Permutation, Parametric, and Bootstrap Tests of Hypotheses (Springer Series in Statistics)*. Springer-Verlag New York, Inc. Secaucus, NJ, USA.

- Huh, M. and M. Jhun (2001). Random permutation testing in multiple linear regression. *Communications in Statistics: Theory and Methods* 30(10), 2023–2032.
- Jung, B. C., M. Jhun, and S. H. Song (2006). A new random permutation test in ANOVA models. *Statistical Papers* 48, 47–62.
- Kirk, R. (1994). *Experimental design: Procedures for the behavioral sciences*. Wadsworth.
- Manly, B. (1997). *Randomization, bootstrap and Monte Carlo methods in biology*. Chapman and Hall, Landan.
- Pesarin, F. (2001). *Multivariate permutation tests: with applications in biostatistics*. Wiley Chichester.
- Salmaso, L. (2003). Synchronized permutation tests in  $2^k$  factorial designs. *Communications in Statistics-Theory and Methods* 32(7), 1419–1437.
- Searle (1987). *Linear Models for Unbalanced Data*. John Wiley and Sons, New York.
- StatSoft, I. (2007). Statistica (data analysis software system). version 8.0. www.statsoft.com.
- Still, A. and A. White (1981). The approximate randomization test as an alternative to the F test in analysis of variance. *British Journal of Mathematical and Statistical Psychology* 34(2), 243–252.
- Ter Braak, C. (1992). Permutation versus bootstrap significance tests in multiple regression and ANOVA. *Bootstrapping and Related Techniques*, 79–86.
- Welch, W. and L. Gutierrez (1988). Robust permutation tests for matched-pairs designs. *J. American Statistical Assoc* 83(402), 450–455.

## Appendix

### Proof of Theorem 2

Using Eq. (1), the residuals can be written as

$$r = y - \hat{y} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y = (I - H)y \quad (19)$$

where  $H = X(X'X)^{-1}X'$  is an  $N \times N$  matrix, which is known as the Hat matrix. Note that  $H$  is the projection matrix to the subspace spanned by the columns of  $X$ . It is easy to show that  $H$  is an idempotent matrix for any design matrix  $X$ ; that is,  $H^2 = H$ . Similarly,  $I - H$  is also idempotent. Using the idempotent property of  $H$ , the residual sum of squares under the full model can be written as

$$\text{RSS}(\hat{\beta}) = r'r = y'(I - H)^2y = y'(I - H)y = y'(I - X(X'X)^{-1}X')y, \quad (20)$$

which shows the equality of the denominators of  $F_{ss}$  and  $F_R$ . Similarly, we can show that  $\text{RSS}(\hat{\beta}_1) = y'(I - H_1)y$ , with  $H_1 = X_1(X_1'X_1)^{-1}X_1'$ .

It remains to show that  $y'(I - H_1)y = y'(I - H)y = y'(H_{rr})y$ , with  $H_{rr} = X_{rr}(X_{rr}'X_{rr})^{-1}X_{rr}'$ . This is implied by  $H = H_1 + H_{rr}$ , which is proved in the following two lemmas.

Let  $X_1$  and  $X_2$  be two real matrices each with  $N$  rows. Define  $X_{rr} = (I - H_1)X_2$ , where  $H_1 = X_1(X_1'X_1)^{-1}X_1'$ . Then the subspaces spanned by the union of the columns of  $X_1$  and  $X_2$  is the same as that of  $X_1$  and  $X_{rr}$ , that is;

$$([X_1|X_2]) = ([X_1|X_{rr}]). \quad (21)$$

*Proof of Lemma 6:* In order to show the equality of the two subspaces, we will show that each of them is a subspace of the other one.

- i)  $([X_1|X_2]) \subseteq ([X_1|X_{rr}])$ . We just have to show that the columns of  $X_2$  lie in  $([X_1|X_{rr}])$ ; i.e., there exist matrices  $A$  and  $B$ , such that  $X_2 = X_1A + X_{rr}B$ . The definition of  $X_{rr}$  implies that

$$X_{rr} = (I - H_1)X_2 = (I - X_1(X_1'X_1)^{-1}X_1')X_2,$$

or, equivalently,  $X_2 = X_1((X_1'X_1)^{-1}X_1'X_2) + X_{rr}$ , which shows the claim.

- ii)  $([X_1|X_{rr}]) \subseteq ([X_1|X_2])$ . Rewriting  $X_{rr}$  as

$$X_{rr} = X_1(-(X_1'X_1)^{-1}X_1'X_2) + X_2 \quad (22)$$

shows that  $X_{rr}$  is a linear combination of  $X_1$  and  $X_2$ . This completes the proof of this lemma. ■

$$H = H_1 + H_{rr}.$$

*Proof of Lemma 6:* We showed in Lemma 6 that  $(X) = (X_{1rr})$ , where  $X = [X_1|X_2]$  and  $X_{1rr} = [X_1|X_{rr}]$ . Therefore,  $H$  and  $H_{1rr} = X_{1rr}(X_{1rr}'X_{1rr})^{-1}X_{1rr}'$ , the projection matrices onto this subspace are equal. It remains to show  $H_{1rr} = H_1 + H_{rr}$ .

Using the definition of  $X_{rr}$ , we have

$$X_1'X_{rr} = X_1'(I - H_1)X_2 = X_1'(I - X_1(X_1'X_1)^{-1}X_1')X_2 = 0. \quad (23)$$

The columns of the matrices  $X_1$  and  $X_{rr}$  are therefore orthogonal. Hence  $(X_1)$  and  $(X_{rr})$ , the subspaces spanned by the columns of these two matrices, are orthogonal. Thus,

$$\begin{aligned} X_{1rr}'X_{1rr} &= \begin{bmatrix} X_1' \\ X_{rr}' \end{bmatrix} [X_1 \ X_{rr}] \\ &= \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_{rr}'X_{rr} \end{bmatrix} \end{aligned}$$

is a block diagonal matrix, and its inverse is simply obtained by

$$(X_{1rr}'X_{1rr})^{-1} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_{rr}'X_{rr})^{-1} \end{bmatrix}. \quad (24)$$

Therefore, we have

$$\begin{aligned} H_{1rr} &= X_{1rr}(X_{1rr}'X_{1rr})^{-1}X_{1rr}' \\ &= [X_1 \ X_{rr}] \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_{rr}'X_{rr})^{-1} \end{bmatrix} \begin{bmatrix} X_1' \\ X_{rr}' \end{bmatrix} \\ &= X_1(X_1'X_1)^{-1}X_1' + X_{rr}(X_{rr}'X_{rr})^{-1}X_{rr}' \\ &= H_1 + H_{rr}. \end{aligned} \quad (25)$$

This completes the proof. ■

### Proof of Theorem 3

Note that the mean of the original error is zero, and so is the mean of the modified model error, since it does not change under a linear transformation. Moreover,

$$\varepsilon_{mr} = V'\varepsilon_{rr} = V'(I - H_1)\varepsilon = V'VV'\varepsilon = V'\varepsilon. \quad (26)$$

The exchangeability of the raw data implies that all the non-diagonal elements of its covariance matrix are the same (and similarly all its diagonal elements are the same); i.e., the covariance



matrix can be written as  $(\varepsilon) = (\sigma^2 - \rho)I_N + \rho'$ , where  $\rho'$  is the all-one column of length  $N$ . Hence the covariance matrix of the modified model error can be written as

$$\begin{aligned} (\varepsilon_{mr}) &= V'((\varepsilon))V \\ &= V'((\sigma^2 - \rho)I_N)V + \rho V''V \\ &= (\sigma^2 - \rho)I_{N-q} + \rho V''V. \end{aligned}$$

In the presence of an intercept,  $\rho'$  is a vector in  $(X_1)$ , and it does not change under a projection onto this space. Hence,  $VV' = (I - H_1)1 = 0$ . This equality, together with the fact that  $V$  has full column rank, implies that  $V'1 = 0$ . Therefore, the covariance of  $\varepsilon_{mr}$  (and  $y_{mr}$ ) is a compound symmetric matrix, and the elements of  $\varepsilon_{mr}$  are exchangeable. It is clear that since the class of independent and identical distributions is a special subclass of the exchangeable distributions, the elements of  $\varepsilon_{mr}$  (and  $y_{mr}$ ) are also exchangeable under the i.i.d. assumption. ■

We state and prove the following lemma which plays a key role in the proof of Theorem 4.

Let  $V$  be an  $N \times (N - q)$  matrix which satisfies  $VV' = I_N - H_1$  and  $V'V = I_{N-q}$ . For any  $(N - q) \times (N - q)$  permutation matrix  $P$ , there exists a unitary matrix  $T_P$  such that  $P'V'T_P = V'$ . *Proof of Lemma 6:* Recall that  $I - H_1$  has  $N - q$  eigenvalues equal to 1, and  $q$  eigenvalues equal to 0. Let  $M$  be an  $N \times q$  matrix containing the normalized eigenvectors of  $I - H_1$  corresponding to the zero eigenvalues. We claim that the square matrix  $T_P = MM' + VPV'$  fulfills the requirements of the lemma.

Note that the columns of  $V$  are the eigenvectors of  $I - H_1$  corresponding to non-zero eigenvalues. Since  $I - H_1$  is an idempotent matrix, its eigenvectors are orthonormal; i.e.,  $V'M = M'V = 0$ . The first property of  $T_P$  can be easily seen through

$$P'V'T_P = P'V'(MM' + VPV') = P'(V'M)M' + P'(V'V)P'V' = V', \quad (27)$$

where the last equality follows from  $P'P = I_{N-q}$ , which holds for any permutation matrix.

In order to show that  $T_P$  is unitary, we can write

$$\begin{aligned} T_P'T_P &= (MM' + VPV')(MM' + VPV') \\ &= MM' + I_N - H_1 = I_N, \end{aligned} \quad (28)$$

where, in the last equality, we used the fact that  $M'M = I_q$ . Note that the eigenvectors of  $I - H_1$  corresponding to the zero eigenvalues are exactly the eigenvectors of  $H_1$  corresponding to the eigenvalues equal to 1. Thus,  $MM' = H_1$ , which completes the proof. ■

## Proof of Theorem 4

Note that by definition of Eq. (26), one can write

$$f_{\varepsilon_{mr}}(u) = \int_{x: V'x=u} f_{\varepsilon}(x)dx.$$

Let  $P$  be an arbitrary permutation matrix. We have

$$\begin{aligned} f_{\varepsilon_{mr}}(Pu) &= \int_{x:V'x=Pu} f_{\varepsilon}(x)dx \\ &= \int_{x:P'V'x=u} f_{\varepsilon}(x)dx \end{aligned} \quad (29)$$

$$= \int_{w:P'V'T_Pw=u} f_{\varepsilon}(T_Pw)dw \quad (30)$$

$$= \int_{w:V'w=u} f_{\varepsilon}(T_Pw)dw \quad (31)$$

$$\begin{aligned} &= \int_{w:V'w=u} f_{\varepsilon}(w)dw \\ &= f_{\varepsilon_{mr}}(u). \end{aligned} \quad (32)$$

In the above chain of equalities, we use the fact that  $P'P = I$  in (29); we use a change of variable  $x = T_Pw$  in (30) where the matrix  $T_P$  is the unitary matrix introduced in Lemma 6 and its property is used in (31). Finally, in (32) we use the fact that  $\varepsilon$  has a spherical density. It is worth mentioning that the determinant of the Jacobian matrix arising in the change of variable in (30) equals 1, since  $T_P$  is a unitary matrix. ■

## Proof of Theorem 5

Define

$$F_J = \frac{y'_{rr}(X_{rr}(X'_{rr}X_{rr})^{-1}X'_{rr})y_{rr}/(p-q)}{y'_{rr}(I - X_{rr}(X'_{rr}X_{rr})^{-1}X'_{rr})y_{rr}/(N-p)}. \quad (33)$$

In order to prove the theorem, we will show that 1)  $F_{rr} = F_R$ , 2)  $F_J = F_{rr}$  and finally 3)  $F_{mr} = F_J$ . These, together with Theorem 2 ( $F_{ss} = F_R$ ), show the claim.

1) We will show that the numerators and denominators of  $F_{rr}$  and  $F_R$  are the same. Note that Eq. (23) implies that

$$\begin{aligned} H_1H_{rr} &= (X_1(X'_1X_1)^{-1}X'_1)(X_{rr}(X'_{rr}X_{rr})^{-1}X'_{rr}) \\ &= X_1(X'_1X_1)^{-1}(X'_1X_{rr})(X'_{rr}X_{rr})^{-1}X'_{rr} = 0. \end{aligned}$$

Therefore,

$$(I - H_1)H_{rr}(I - H_1) = H_{rr}. \quad (34)$$

Using this equality, one can easily see that the two numerators are equal:

$$y'_{rr}H_{rr}y_{rr} = y'(I - H_1)H_{rr}(I - H_1)y = y'H_{rr}y.$$

Similarly the denominators of  $F_{rr}$  and  $F_R$  are equal because

$$\begin{aligned} y'_{rr}(I - H)y_{rr} &= y'_{rr}(I - H_1 - H_{rr})y_{rr} \\ &= y'(I - H_1)((I - H_1) - H_{rr})(I - H_1)y \\ &= y'(I - H_1 - (I - H_1)H_{rr}(I - H_1))y \\ &= y'(I - H_1 - H_{rr})y \\ &= y'(I - H)y. \end{aligned}$$

Note that we have used the fact that  $H = H_1 + H_{rr}$  in the first line of the above chain, as shown in Lemma 6.

2) The numerators of  $F_J$  and  $F_{rr}$  are the same by their definition. Concerning the proof of equality of  $F_J$  and  $F_{rr}$ , it remains to show that their denominators are equal. We have

$$\begin{aligned} y'_{rr}(I - H_{rr})y_{rr} &= y'(I - H_1)(I - H_{rr})(I - H_1)y \\ &= y'((I - H_1) - (I - H_1)H_{rr}(I - H_1))y \\ &= y'(I - H_1 - H_{rr})y \\ &= y'(I - H)y. \end{aligned}$$

3) The proof of  $F_J = F_{mr}$  is a direct consequence of the three following chains of equalities:

$$\begin{aligned} y'_{mr}X_{mr} &= y_{rr}VV'X_{rr} = y_{rr}(I - H_1)X_{rr} = y'_{rr}X_{rr}, \\ X'_{mr}X_{mr} &= X'_{rr}VV'X_{rr} = X'_{rr}(I - H_1)X_{rr} = X'_{rr}X_{rr}, \end{aligned}$$

and

$$y'_{mr}y_{mr} = y'_{rr}VV'y_{rr} = y'_{rr}(I - H_1)y_{rr} = y'_{rr}y_{rr},$$

in which where we have used  $VV' = I - H_1$ . This completes the proof. ■

## Acknowledgment

The authors thank the editor and referees for their helpful comments, especially one referee for motivating us to add Theorem 3 and 4. The work is supported by the Swiss National Foundation under the grant number 105211-112465/1 and 105214-116269/1.