Semiparametric Inference for the Two-way Layout Under Order Restrictions

ORI DAVIDOV
Department of Statistics, University of Haifa

KONSTANTINOS FOKIANOS
Department of Mathematics and Statistics, University of Cyprus

GEORGE ILIOPOULOS
Department of Statistics and Insurance Science, University of Piraeus

ABSTRACT. There are many situations in which a researcher would like to analyse data from a two-way layout. Often, the assumptions of linearity and normality may not hold. To address such situations, we introduce a semiparametric model. The model extends the well-known density ratio model from the one-way to the two-way layout and provides a useful framework for semiparametric analysis of variance type problems under order restrictions. In particular, the likelihood ratio order is emphasized. The model enables highly efficient inference without resorting to fully parametric assumptions or the use of transformations. Estimation and testing procedures under order restrictions are developed and investigated in detail. It is shown that the model is robust to misspecification, and several simulations suggest that it performs well in practice. The methodology is illustrated using two data examples; in the first, the response variable is discrete, whereas in the second, it is continuous.

Key words: density ratio model, likelihood ratio order, maximum likelihood, model misspecification, two-way layout

1. Introduction

This article provides an alternative approach to the classical analysis of the two-way layout for ANOVA-type problems. Our formulation emphasizes order restricted inference. The proposed methodology simultaneously relaxes the standard assumptions of linearity and normality and does not require any data transformations. The approach is quite general and can be easily extended to multiway layouts. For expository purposes, we restrict ourselves to the two-way design.

To motivate our study, consider an agricultural experiment, which explores the effect of nitrogen and phosphorus additives on the yield of maize. In this example, which we analyse in more detail in Section 6, researchers are interested in showing that yield of maize increases, marginally, with the amount of both nitrogen and phosphorus. They may also be interested in exploring if there is some interaction effect. However, if the outcome does not follow a normal distribution and/or the errors are not homoscedastic, then inferences drawn from a standard ANOVA-type analysis may not be correct. Moreover, researchers may not be only interested in ordering the means, that is, the expected yield at a given combination of the experimental factors, but more generally in ordering their entire distributions.

We start with a description of the proposed model. Suppose that \((i, j) \in \mathcal{M} = \mathcal{R} \times \mathcal{C}\), where \(i \in \mathcal{R} = \{1, \ldots, R\}\) indexes the row factor and \(j \in \mathcal{C} = \{1, \ldots, C\}\) the column factor. Let \(Y\) be the outcome of interest, and let \(g_{ij}, G_{ij}\) denote its density (or mass) function and
distribution function (DF), respectively, at the \((i, j)\)th level. The model that we initially consider has the form

\[ g_{ij}(y) = f(y) \exp \{ \alpha_{ij} + \lambda_{ij} \phi(y) \}. \]  

(1)

Here, \( f \) is a completely unspecified baseline density function, \( \alpha = (\alpha_{ij}) \) are the logarithms of the normalizing constants and \( \lambda = (\lambda_{ij}) \) are the row by column effects. The function \( \phi \) is assumed to be known and monotone increasing. The misspecification of \( \phi \) is discussed in Section 3. The assumption of monotonicity turns out to be important for inducing stochastic order relations among the \( R \times C \) groups. We also assume that \( \lambda_{11} = 0 \) and consequently \( \alpha = 0 \), so \( g_{11}(y) = f(y) \) for all \( y \in \mathbb{R} \). Model (1) generalizes the standard two-way normal ANOVA model where it is assumed that observations from the \((i, j)\)th treatment group follow a \( \mathcal{N}(\mu_{ij}, \sigma^2) \) distribution. Indeed, if \( g_{ij} \) is the density of a \( \mathcal{N}(\mu_{ij}, \sigma^2) \) random variable, then straightforward calculations show that \( g_{ij}(y)/g_{11}(y) = \exp(\alpha_{ij} + \lambda_{ij} \phi(y)) \), where

\[ \alpha_{ij} = \left( \mu_{11}^2 - \mu_{ij}^2 \right)/(2\sigma^2), \quad \lambda_{ij} = (\mu_{ij} - \mu_{11})/\sigma^2 \]  

and \( \phi(y) = y \). With this identification, \( \lambda_{ij} \) is a (scaled) contrast between the means \( \mu_{ij} \) and \( \mu_{11} \).

Model (1) extends the well-known density ratio model (Qin & Zhang, 1997; Qin, 1998; Gilbert et al., 1999; Gilbert, 2000; Fokianos et al., 2001; Fokianos, 2004) from the one-way to the two-way layout. The density ratio model provides an attractive modelling framework because it combines the strengths of both parametric and non-parametric approaches. Treatment effects are easy to interpret and understand, yet the shape of the distribution, determined by the baseline density, need not be specified in advance. In fact, many standard models belonging to the exponential family are instances of (1). An important feature of model (1) is that it can be employed for testing for the likelihood ratio ordering without resorting to fully parametric assumptions.

The proposed model is applicable for the analysis of both observational studies and designed experiments. Data of this type often arise in the context of ANOVA models. It is impossible to provide a comprehensive review of ANOVA methods and applications. Scheffé (1959), Speed (1987) and Gelman (2005) were excellent introductions to the parametric approach. Non-parametric methods, based on ranks, have been proposed by Akritas & Arnold (1994), Akritas et al. (1997) and Brunner et al. (1997). Order restricted inference in the context of normal theory ANOVA models is briefly discussed by Robertson et al. (1988) and Silvapulle & Sen (2005). To the best of our knowledge, there does not exist a method, which addresses the problem at hand in a semiparametric context. Davidov et al. (2010) developed ordered restricted inference for the one-way layout in the specific case of the power bias family. Here, their method is extended to the two-way layout using a general function \( \phi \). The extension from the one-way to the two-way layout (and beyond) raises a variety of new statistical and methodological issues and broadens the range of applications and scientific questions, which the methodology can address. Also note that (1) is a biased sampling model with a weight function depending upon unknown parameters. For non-parametric and semiparametric inference in the context of biased sampling, see Vardi (1982, 1985), Gill et al. (1988), Gilbert et al. (1999), Gilbert (2000), Bickel et al. (1998) and Davidov & Iliopoulos (2009).

Several research hypotheses may be of interest in the context of model (1). The null hypothesis most often considered is

\[ H_g : G_{ij} = F \quad \text{for all } (i, j) \in \mathcal{M}, \]  

(2)

which in the terminology of ANOVA means that both the row and column factors have no effect. Other interesting nulls are
When (3) holds, then all distributions within the same row are equal, that is, there is no column effect. Similarly, (4) indicates that there is no row effect. Hypotheses of the type (3) and/or type (4) are of special interest when the row/column factor is used for blocking.

The most general alternative hypothesis is

\[ A_G : G_{ij} \neq F \quad \text{for some } (i, j) \in M, \]  

which specifies that there exists at least one combination of treatments that is different from all the rest. In this paper, we are particularly interested in alternatives that incorporate stochastic order restrictions. The unordered case is briefly presented in Appendix A. Three interesting ordered alternatives are

\[ A_c : G_{ij} \preceq_{lt} G_{ij'} \quad \text{for all } j \leq j' \text{ and } i \in R, \]  
\[ A_r : G_{ij} \preceq_{lt} G_{i'j} \quad \text{for all } i \leq i' \text{ and } j \in C, \]  
\[ A_m : G_{ij} \preceq_{lt} G_{i'j'} \quad \text{for all } (i, j) \leq (i', j') \in M, \]  

where the inequality \((i, j) \leq (i', j')\) holds if \(i \leq i'\) and \(j \leq j'\). The symbol \(\preceq_{lt}\) denotes ordering with respect to the likelihood ratio order, and in all alternatives, at least one strict inequality is required. The relationship specified by (8) is known as the matrix ordering (Robertson et al., 1988; Teoh et al., 2008).

Recall that the DFs \(G_{ij}\) and \(G_{i'j'}\) are said to be ordered by the likelihood ratio order if the ratio \(g_{i'j'}(y)/g_{ij}(y)\) is increasing in \(y\) on the union of their supports. Further, note that if \(G_{ij} \preceq_{lt} G_{i'j'}\), then \(G_{ij}\) and \(G_{i'j'}\) are also ordered by the usual stochastic order, denoted \(G_{ij} \preceq_{st} G_{i'j'}\), and consequently \(\mu_{ij} = \mathbb{E}G_{ij}(X) \leq \mathbb{E}G_{i'j'}(X) = \mu_{i'j'}\); that is, their means are also ordered (Shaked & Shanthikumar, 2007). We emphasize that the likelihood ratio ordering is much stronger than the ordering of the means. In fact, if \(G_{ij} \sim_{lt} G_{i'j'}\), there exists a value \(x\) with \(g_{ij}(x) = g_{i'j'}(x)\), such that for every measurable set \(A \subseteq (x, \infty)\), we have \(\mathbb{P}G_{i'j'}(X \in A) \geq \mathbb{P}G_{ij}(X \in A)\), whereas for \(A \subseteq (-\infty, x)\), the inequality is reversed. This means that large values are more likely under \(G_{i'j'}\) than under \(G_{ij}\) and vice-versa. There are a variety of applications where such an ordering is meaningful for the scientific question at hand.

Observe now that the ratio

\[ g_{i'j'}(y)/g_{ij}(y) = \exp\{\left(\alpha_{i'j'} - \alpha_{ij}\right) + \left(\lambda_{i'j'} - \lambda_{ij}\right) \phi(y)\} \]

is increasing in \(y\) if and only if \(\lambda_{ij} \leq \lambda_{i'j'}\). In other words, \(G_{ij} \preceq_{lt} G_{i'j'}\) provided that \(\lambda_{ij} \leq \lambda_{i'j'}\). This observation implies that hypotheses (2)–(8) can be written in terms of elementary inequality relations among the parameters \(\lambda_{ij}\)’s. For example, testing (2) versus (8) reduces to

\[ H_G : \lambda_{ij} = 0 \quad \text{for all } (i, j) \in M \quad \text{versus} \quad A_m : \lambda_{ij} \leq \lambda_{i'j'} \quad \text{for all } (i, j) \leq (i', j') \in M. \]  

Other hypotheses can be similarly reformulated. For instance, \(A_c\) implies that the \(\lambda_{ij}\)’s are ordered along the rows \((\lambda_{ij} \leq \lambda_{ij'} \quad \text{for all } j \leq j' \text{ and } i \in R)\), whereas \(A_r\) implies that the \(\lambda_{ij}\)’s are ordered along the columns \((\lambda_{ij} \leq \lambda_{i'j} \quad \text{for all } i \leq i' \text{ and } j \in C)\). We conclude that model (1) is useful for testing for order in the two-way layout without resorting to fully parametric assumptions.
The paper is organized in the following way. In Section 2, we describe semiparametric inference procedures for model (1). In particular, constrained estimation and testing procedures for the likelihood ratio ordering among the $R \times C$ groups are developed. In Section 3, we examine the effect of misspecifying the function $\phi$ and show that this is not a serious issue in the context of testing. In fact, our analysis shows that the proposed methodology is consistent even if the density ratio model does not hold, but the distributions are ordered. In Section 4, we reparametrize model (1) and explain how it can be used in order to test for factor interaction. In Sections 5 and 6, respectively, we present simulation results and the analysis of two real data examples. It is made clear that our approach is quite general, flexible and efficient when compared with parametric methods. We conclude with a brief discussion. A supplementary file contains three appendices: Appendix A in which some results for the unordered case are presented, Appendix B, which contains the proofs of our theorems, and Appendix C, wherein some additional numerical results as well as tables and figures can be found.

2. Maximum empirical likelihood estimation

The first step in any ordered analysis is to derive the unconstrained estimators and their large sample properties. This is the material of the first subsection, whereas in the sequel, we discuss the constrained estimators as well as testing procedures for the hypotheses presented in Section 1.

2.1. Estimation without restrictions

Recall model (1). Let $S_{ij}, (i, j) \in \mathcal{M}$, be independent random samples of sizes $n_{ij}$ from $G_{ij}$. In addition, let $Y_k, k = 1, \ldots, n = \sum_{i,j} n_{ij}$, denote the observations in the combined sample $S$, say, and set $Z_k = \phi(Y_k), k = 1, \ldots, n$. Then, the empirical likelihood (cf. Owen, 2001) of $A, A$ and $F$ is

$$L(A, A, F) = \prod_{(i,j) \in \mathcal{M}} \prod_{Y_k \in S_{ij}} g_{ij}(Y_k) = \left\{ \prod_{k=1}^{n} p_k \right\} \left\{ \prod_{i=1}^{R} \prod_{j=1}^{C} \prod_{Y_{k} \in S_{ij}} \exp(\alpha_{ij} + \lambda_{ij} Z_k) \right\},$$

(10)

where $p_k = dF(Y_k) = F(Y_k) - F(Y_k-)$ is the observed jump of $F$ at $Y_k$. As mentioned in Section 1, we assume that $\alpha_{11} = \lambda_{11} = 0$. This guarantees identifiability of model (1). A unique maximizer of (10) exists provided that at least two observations are different and that the samples $S_{ij}$ can not be partitioned to non-overlapping groups (cf. Davidov & Iliopoulos, 2009). By standard arguments (Fokianos et al., 2001), it can be shown (for more details, see Appendix A) that the log profile likelihood for $(A, A)$ is

$$l(A, A) = \sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij} \alpha_{ij} + \sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij} \lambda_{ij} \bar{Z}_{ij} - \sum_{k=1}^{n} \log \left\{ \sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij} \exp(\alpha_{ij} + \lambda_{ij} Z_k) \right\},$$

(11)

where $\bar{Z}_{ij} = n_{ij}^{-1} \sum_{Z_k \in S_{ij}} Z_k$. The maximum likelihood estimator (MLE) of $(A, A)$, denoted by $(\hat{A}, A)$, is the unique solution to the score equations obtained by equating the partial derivatives with respect to $\alpha_{ij}$ and $\lambda_{ij}$ for all $(i, j) \neq (1, 1)$ to zero. It is well known that the solution can be calculated using standard software because the resulting score equations
are essentially the score equations of a multinomial logistic regression model (Fokianos et al., 2001). Then, the DFs $G_{ij}$ can be estimated consistently by $\hat{G}_{ij}(x) = \sum_{k=1}^{R} \hat{g}_{ij}(Y_k)\mathbb{I}(Y_k \leq x)$, where

$$\hat{g}_{ij}(Y_k) = \frac{\exp \left( \hat{\alpha}_{ij} + \hat{\lambda}_{ij} Z_k \right)}{\sum_{i=1}^{C} \sum_{j=1}^{R} \exp \left( \hat{\alpha}_{ij} + \hat{\lambda}_{ij} Z_k \right)}, \quad k = 1, \ldots, n,$$

(12)

$I(\cdot)$ denotes the indicator function and $\hat{\alpha}_{11} = \hat{\lambda}_{11} = 0$. Note that $p_k$ is estimated by $\hat{p}_k = \hat{g}_{11}(Y_k)$.

Let $\alpha = \text{vec}(A^T)$ and $\lambda = \text{vec}(A^T)$ excluding $\alpha_{11}$ and $\lambda_{11}$, respectively, and let $\hat{\alpha}$ and $\hat{\lambda}$ be their estimators. By standard asymptotic theory for $M$-estimators (theorem 5.41 in Van der Vaart, 2000), we have

$$\sqrt{n} \left( \frac{\hat{\alpha} - \alpha}{\hat{\lambda} - \lambda} \right) \Rightarrow N_{2(RC-1)}(0, \Psi) \quad \text{as } n \to \infty.$$  

(13)

Here, $\Rightarrow$ denotes weak convergence, $0$ is a vector of zeros, $\Psi = S^{-1}V S^{-1}$ with $S = -n^{-1} \mathbb{E}(\nabla^2 l)$ and $V = n^{-1} \mathbb{E}\{\nabla l(\nabla l)^T\}$, and it is assumed that $n_{ij}/n \to \rho_{ij} \in (0, 1)$ as $n \to \infty$. The matrix $\Psi$ is a function of the unknown parameters, and its explicit form can be derived as in Fokianos et al. (2001) and Davidov et al. (2010). It can be consistently estimated from the data by plugging in the corresponding MLEs. Therefore, (13) implies that, under appropriate regularity conditions, large sample inferential procedures for $(A, A)$ are readily developed. Note that when $A = 0$, then necessarily $A = 0$ and that the limit distribution in (13) is singular (e.g. Fokianos et al., 2001). In this case,

$$n^{1/2} \hat{\lambda} \Rightarrow N_{RC-1}(0, \Sigma_0).$$

(14)

The matrix $\Sigma_0$ is non-singular and defined as follows. For each $k, k' \in \{1, \ldots, RC\}$, find the unique $i, i' \in \{1, \ldots, R\}$ and $j, j' \in \{1, \ldots, C\}$ such that $k = (i - 1)C + j$ and $k' = (i' - 1)C + j'$. Let $\Omega = \{\omega_{kk'}\}$ be the $(RC) \times (RC)$ matrix with $\omega_{kk'} = \delta_{kk'} \delta_{kk'} \rho_{ij} (\delta_{kk'} - \rho_{ij})$, where $\delta_{st}$ denotes Kronecker’s delta. Then,

$$\Sigma_0 = \left[ \text{Var}_F \{\phi(Y)\} \right]^{-1} \Omega'.$$  

(15)

where $\Omega'$ arises from $\Omega$ by deleting its first row and column.

2.2. Estimation under order restrictions

Let $\mathcal{L}_c$, $\mathcal{L}_r$ and $\mathcal{L}_m$ denote the parameter space of $A$ under the hypotheses $A_c$, $A_r$ and $A_m$ in (6)–(8), respectively. In what follows, we will focus on estimation under $A_m$; the situation for the other cases is similar.

Consider estimation of $(A, A, F)$ subject to the usual constraints that $\sum_{k=1}^{R} p_k \exp(\alpha_{ij} + \lambda_{ij} Z_k) = 1$ for $(i, j) \in M$ and the additional constraint that $A \in \mathcal{L}_m$. It is important to note that the profiling procedure depends exclusively on $A$. Therefore, the log profile likelihood of model (1) under the restriction $A \in \mathcal{L}_m$ is given by (11).

Let

$$(A_m^*, A_m^*) = \text{arg max}\{l(A, A) : A \in \mathcal{L}_m\}$$

(16)

be the restricted MLEs. Denote also by $\lambda_m^*$ the vector version of $A_m^*$. It can be shown that the restrictions imposed on $A$ induce restrictions on $A$. In fact, it can be shown that if $0 < \lambda_{ij} \leq \lambda_{i'j'}$, then $\alpha_{ij}/\lambda_{ij} \leq \alpha_{i'j'}/\lambda_{i'j'}$ and $\alpha_{ij} = \alpha_{i'j'}$ whenever $\lambda_{ij} = \lambda_{i'j'}$. However, $A$
is completely determined by \( A \) and \( F \). Therefore, the additional restrictions are automatically satisfied, and thus, they need not be considered in the estimation procedure.

Observe that (16) is a convex optimization problem in standard form. Several algorithms exist for solving this problem (see, for example, Boyd & Vandenberghe, 2004). The standard approach in the statistical literature to such problems is to project the unconstrained estimator onto the constrained space (cf. Piegorsch, 1990; Silvapulle, 1994; Silvapulle & Sen, 2005). That is, to consider

\[
\lambda^{**}_m = \Pi_{\Sigma} \left( \hat{\lambda}_m \mid \mathcal{L}_m \right) \equiv \arg \min_{\lambda \in \mathcal{L}_m} \left\{ (\lambda - \hat{\lambda}_m)^T \Sigma^{-1} (\lambda - \hat{\lambda}_m) \right\},
\]

where \( \Sigma \) is the estimated (unconstrained) covariance matrix of \( \hat{\lambda} \) and \( \Pi_{\Sigma} \) is the projection operator onto the constrained parameter space. Obviously, \( \lambda^{**}_m \) is a solution to a quadratic problem. However, we have developed a much simpler, exact, estimation method motivated by the following observation. Let \( \hat{g}_{ij} \) be the probability mass function with mean \( \bar{Z}_{ij} \) defined in (12). Now

\[
\hat{g}_{ij}(y) = \exp \left\{ \left( \hat{\alpha}_{i'j'} - \hat{\alpha}_{ij} \right) + \left( \hat{\lambda}_{i'j'} - \hat{\lambda}_{ij} \right) \phi(y) \right\}, \quad y \in \mathcal{S},
\]

so \( \hat{g}_{ij} \prec_{\mathcal{S}} \hat{g}_{i'j'} \) if and only if \( \hat{\alpha}_{ij} < \hat{\alpha}_{i'j'} \) and \( \hat{\lambda}_{ij} = \hat{\lambda}_{i'j'} \). But because \( \hat{g}_{ij} \) and \( \hat{g}_{i'j'} \) are necessarily ordered by the likelihood ratio order, the direction of the ordering is determined by the means \( \bar{Z}_{ij} \) and \( \bar{Z}_{i'j'} \). Thus, the aforementioned conditions are equivalent to \( \bar{Z}_{ij} < \bar{Z}_{i'j'} \) and \( \bar{Z}_{ij} = \bar{Z}_{i'j'} \), respectively. This observation suggests a computationally simple algorithm for finding \((A^*, A^*)\), which requires solving one quadratic programme and then fitting a multinomial logistic model. Both computational tasks are readily implemented in the \( \mathbb{R} \) environment. The algorithm is formally presented in Appendix B where it is further shown that it yields the restricted MLE and that both \((A^*, A^*)\) and \((A^{**}, A^{**})\) are consistent for \((A, A)\).

2.3. Testing for ordering

We start by reformulating the hypotheses (2)–(8). To do so, it is convenient for any \( K \geq 2 \) to define the \( K \times K \) matrix \( T_{0,K} = (t_{ij}) = (\delta_{ij} - \delta_{i,j-1}) \), where \( \delta_{it} \) denotes Kronecker’s delta. Further, let \( T_{1,K} \) be the \( K \times (K - 1) \) matrix arising from \( T_{0,K} \) by deleting its first column. Next, observe that testing the hypotheses (2) or (3) versus (6) can be expressed equivalently as

\[
H_g: \ A T_{0,C} = 0 \quad \text{versus} \quad A_c: \ A T_{1,C} \geq 0,
\]
\[
H_c: \ A T_{1,C} = 0 \quad \text{versus} \quad A_c: \ A T_{1,C} \geq 0,
\]

respectively. Similarly, for testing (2) or (4) versus (7), we have

\[
H_g: \ T_{0,R}^T A = 0 \quad \text{versus} \quad A_r: \ T_{1,R}^T A \geq 0,
\]
\[
H_r: \ T_{1,R}^T A = 0 \quad \text{versus} \quad A_r: \ T_{1,R}^T A \geq 0,
\]

respectively, whereas testing (2) versus (8) is equivalent to testing

\[
H_g: \ T_{1,R}^T A = 0 \quad \text{and} \quad A T_{1,C} = 0 \quad \text{versus} \quad A_m: \ T_{1,R}^T A \geq 0 \quad \text{and} \quad A T_{1,C} \geq 0.
\]

By linearity, all of the aforementioned hypotheses can be expressed in terms of the vector \( \lambda \). Let \( I_K \) be the identity matrix of order \( K \) and denote by \( \otimes \) Kronecker’s product. Then \( T_{1,R}^T A = ( \geq 0 ) \) is equivalent to \( R_1 \lambda = ( \geq 0 ) \), where \( R_1 \) is the \( C(R - 1) \times (RC - 1) \) matrix
obtained after deleting the first column of the matrix $I_C \otimes T_{1,R}$. Similarly, denote by $C_0$, $C_1$, $R_0$ and $M$ the matrices corresponding to the hypotheses about $AT_{0,C}$, $AT_{1,C}$, $T_{0,R}^T A$, and (jointly) $T_{1,R}^T A$ and $AT_{1,C}$, respectively. Finally, notice that the number of rows in these matrices are $RC - 1$, $(C - 1)R$, $RC - 1$ and $2RC - R - C$, respectively, whereas the number of columns always equals $RC - 1$.

Let $\Lambda_{\mu,v}$ denote the likelihood ratio statistic (LRT) for testing $H_\mu$ against $A_v$, for $\mu \in \{g, r, c\}$ and $v \in \{r, c, m\}$, that is,

$$\Lambda_{\mu,v} = 2 \left\{ \sup_{A_v} l(A, A) - \sup_{H_\mu} l(A, A) \right\}.$$  \hspace{1cm} (18)

The large sampling distribution of (18) is given in the following theorem.

**Theorem 1.** Under the corresponding null hypotheses,

$$\lim_{n \to \infty} \mathbb{P}(\Lambda_{gc} > x) = \sum_{k=0}^{R(C-1)} w_k \left(R(C-1), C_1 \Sigma_0 C_1^T \right) \mathbb{P} \left( \chi^2_{R-1+k} > x \right),$$

$$\lim_{n \to \infty} \mathbb{P}(\Lambda_{cc} > x) = \sum_{k=0}^{R(C-1)} w_k \left(R(C-1), C_1 \Sigma_0 C_1^T \right) \mathbb{P} \left( \chi^2_{C} > x \right),$$

$$\lim_{n \to \infty} \mathbb{P}(\Lambda_{gr} > x) = \sum_{k=0}^{C(R-1)} w_k \left(C(R-1), R_1 \Sigma_0 R_1^T \right) \mathbb{P} \left( \chi^2_{C-1+k} > x \right),$$

$$\lim_{n \to \infty} \mathbb{P}(\Lambda_{rr} > x) = \sum_{k=0}^{C(R-1)} w_k \left(C(R-1), R_1 \Sigma_0 R_1^T \right) \mathbb{P} \left( \chi^2_{R} > x \right),$$

$$\lim_{n \to \infty} \mathbb{P}(\Lambda_{gm} > x) = \sum_{k=0}^{RC-1} w_k \left(RC - 1, M \Sigma_0 M^T \right) \mathbb{P} \left( \chi^2_{k} > x \right),$$

where $\Sigma_0$, $\Sigma'_0$ and $\Sigma''_0$ are the covariance matrices of the asymptotic normal distribution of $\hat{\lambda}$ under the various null hypotheses, given that model (1) holds.

Theorem 1 states that the LRTs follow chi-bar-square distributions, which are mixtures of chi-square distributions with different degrees of freedom. The quantities $w_k$ are known as the chi-bar-square weights or level probabilities. The weights are non-negative; they sum to unity, and they are functions of the asymptotic covariance matrix obtained under the null and the constraints specified in the alternative hypothesis. For completeness, we mention that the weights can be computed in closed form, when $\Sigma_0$ (alternatively $\Sigma'_0$ or $\Sigma''_0$) is known, only if the dimension of the mixture is less than 5. Note further that although $\Sigma_0$ depends only on the sample sizes, the matrices $\Sigma'_0$ and $\Sigma''_0$ depend also on the unknown value of $A$. Nevertheless, a large sample estimate of the weights can be obtained by simulation, as discussed in detail by Silvapulle & Sen (2005).

However, because we would like to consider relatively small samples and avoid large sample approximations, we have implemented our test by simulating the permutation distribution of the LRT under the null and comparing the observed value of our statistic with the tail of that distribution. For example, under hypothesis (2), all distributions are equal, and one can permute between all of them. Under $H_c$, all distributions within the same row are equal but may differ across rows. Thus, the permutation distribution of the LRT is simulated by permuting the observations within rows.

© 2013 Board of the Foundation of the Scandinavian Journal of Statistics.
3. Model misspecification

It is important to recognize that the functional form of $\phi$, used to model the log density ratio, may be known only in limited situations. For example, if the distributions $G_{ij}$ belong to the exponential family, then a good guess as to the form of $\phi$ may be made. Because $\phi(x) = x$ for normal densities, it is reasonable to use the identity function in other location problems where means are to be compared. Another example is the function $\phi(x) = \log x$, which is the correct specification under the so-called power biased sampling model (cf. Davidov et al., 2010). In most situations, however, the functional form of $\phi$ is unknown, and such modelling assumptions may be unwarranted. The issue of model misspecification for the density ratio model has not been well addressed in the literature.

Unfortunately, it turns out that estimating $\phi$ non-parametrically seems to be a difficult problem. Figure 1 displays plots of $\log \left( \hat{g} / \hat{f} \right)$ when $f$ and $g$ are the densities of $\mathcal{N}(0, 1)$ and $\mathcal{N}(2, 1)$, and $\hat{f}$ and $\hat{g}$ are either their optimal Gaussian kernel estimators (smooth line) or their non-parametric MLEs under the likelihood ratio ordering described by Carolan & Tebbs (2005) (the resulting step function is represented by dots). The log-ratio is plotted for the middle half of the combined sample, that is, excluding the tails where the estimators are less precise. Given that here $\log \left( \frac{g(x)}{f(x)} \right) = 2x - 2$, it is clear that such plots can be quite misleading even for moderate to large sample sizes.

An alternative parametric approach is to assume that $\phi$ belongs to a low-dimensional parametric family of functions such as the Box–Cox family, that is, $\phi(x) = (x^\xi - 1)/\xi$ for $\xi \neq 0$ and $\log x$ for $\xi = 0$, and consider $\xi$ as a parameter to be estimated. This approach can be combined with goodness of fit testing in the following way. Choose a number of $\phi$ functions belonging to the Box–Cox family, such as $\log x$, $-1/x$, $x^{1/2}$, $x$ and $x^2$. Fit the model with each one separately and choose the best fitting model, that is, the model minimizing some prechosen criterion such as sums of Kolmogorov–Smirnov distance statistics; see Zhang (2002) for more details. See also Fokianos & Kaimi (2006) who chose the value $\xi$ on the basis of a different criterion, namely, maximization of the empirical likelihood. Clearly, these approaches are not satisfactory.

It follows that, in general, correctly specifying $\phi$ is a difficult problem, which has not received proper attention in the literature. Nevertheless, we have found that model misspecification is not a cause of great concern in the context of order restricted inference as is shown in the succeeding text.

Fig. 1. Plots of $\log \left( \hat{g} / \hat{f} \right)$ based on random samples of sizes $n_1$ and $n_2$ from $f \sim \mathcal{N}(0, 1)$ and $g \sim \mathcal{N}(2, 1)$, respectively. The densities are estimated either using Gaussian kernels (smooth line) or by maximization of the empirical likelihood under the likelihood ratio ordering (dots), and $\log \left( \hat{g} / \hat{f} \right)$ is plotted for the middle half of combined sample. The dashed line corresponds to the true ratio of densities $g(x)/f(x) = 2x - 2$. 

© 2013 Board of the Foundation of the Scandinavian Journal of Statistics.
Theorem 2. The test based on fitting the model (1) using \( \psi \) (a misspecified function) instead of \( \phi \) (the true function) is (i) asymptotically of level \( \alpha \) under the null; and (ii) consistent under the alternative.

Theorem 2 shows that the test based on a misspecified model maintains the true level under the null and is consistent under the alternative (i.e. as \( n \to \infty \), it rejects the null with probability one). In fact, our proofs show that the test is consistent even if the density ratio model does not hold, provided that the distributions are indeed ordered as specified by the alternative. In general, though, misspecification leads to some loss in power. This is intuitively clear and empirically verified by a simulation study presented in Section 5. It is important to realize that although tests are not much affected, estimation of some features of the underlying distributions may be sensitive to misspecification. Generally speaking, in a semiparametric model, which depends on a finite-dimensional and an infinite-dimensional parameter, misspecification of the parametric part can greatly affect inference on the non-parametric part. In particular, the estimator of the DF may not be consistent, and therefore, empirical functionals may not converge to population functionals. See also Fokianos & Kaimi (2006) who studied the effect of \( \phi \) misspecification in the context of unordered inference.

It is worth mentioning that this robustness of the test to misspecification in the context of order restricted inference does not carry over to the unordered case. In fact, as discussed in Appendix A, there are cases where the asymptotic power of the test under \( A_g \) equals its asymptotic size (Theorem A.1).

4. Modelling factor effects

Note that one may express \( \lambda_{ij} \) as

\[
\lambda_{ij} = \beta_i + \gamma_j + \theta_{ij},
\]

(19)

where \( \beta_i \) and \( \gamma_j \) are row and column main effects and \( \theta_{ij} \) is an interaction term. Additionally, for identifiability, set \( \beta_1 = \gamma_1 = 0 \) as well as \( \theta_{i1} = \theta_{1j} = 0 \) for all \( (i, j) \in M \). This reparametrization implies that \( \beta_i = \lambda_{i1}, \gamma_j = \lambda_{1j} \) and \( \theta_{ij} = \lambda_{ij} - \beta_i - \gamma_j \), for \( i \in R, j \in C \).

The reparametrization (19) is useful as in standard two-way ANOVA models. Consider testing

\[
H_i : \theta_{ij} = 0 \quad \text{for all } i, j \quad \text{against} \quad A_i : \theta_{ij} \neq 0 \quad \text{for at least one pair } (i, j).
\]

(20)

that is, the hypothesis of no interaction. If \( H_i \) is not rejected, we may set \( \theta_{ij} = 0 \) and proceed with the ‘additive’ model

\[
g_{ij}(x) = g_{11}(x) \exp\{\alpha_{ij} + (\beta_i + \gamma_j)\phi(x)\}.
\]

(21)

When the sample sizes \( n_{ij} \) are moderate to large, (20) can be tested by comparing \( \hat{\theta}_{ij} \)’s with their standard errors, which can be estimated by bootstrapping non-parametrically within each group. In the case of small sample sizes, one can use the following approach: fit first both the additive and the full model and calculate the log LRT \( \Lambda_i = 2\{\sup(A, A) - \sup H_i l(A, A)\} \).

Estimate \( \hat{g}_{ij} \)’s under the additive model, that is, under \( H_i \) and then bootstrap from the estimated distributions (which all have support \( S \)) in order to approximate the null bootstrap distribution of \( \Lambda_i \). Finally, reject \( H_i \) if the observed \( \Lambda_i \) is large compared with the bootstrapped ones.

Under model (21), hypotheses (2)–(8) are expressed in terms of \( \beta_i \)’s and \( \gamma_j \)’s. In particular, \( A_c \) is equivalent to \( \gamma_j \leq \gamma_j' \) for all \( j \); \( A_r \) is equivalent to \( \beta_i \leq \beta_i' \) for all \( i \), whereas \( A_m \) holds if and only if both \( A_r \) and \( A_c \) hold. See Table 1 for formal expressions.

© 2013 Board of the Foundation of the Scandinavian Journal of Statistics.
Table 1. *Hypotheses (2)–(8) expressed in terms of the parameters of model (21)*

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Expression in terms of $\mathbf{\beta}$ and/or $\mathbf{\gamma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_g$</td>
<td>$\mathbf{\beta} = \mathbf{0}$ and $\mathbf{\gamma} = \mathbf{0}$</td>
</tr>
<tr>
<td>$H_c$</td>
<td>$\mathbf{\gamma} = \mathbf{0}$</td>
</tr>
<tr>
<td>$A_r$</td>
<td>$C_1(\mathbf{\beta}^T, \mathbf{\gamma}^T)^T \geq \mathbf{0}$, where $\tilde{C}<em>1 = \left( \mathbf{0}</em>{C-1,R-1} \mathbf{T}^T_{0,C-1} \right)$</td>
</tr>
<tr>
<td>$A_m$</td>
<td>$\tilde{M}(\mathbf{\beta}^T, \mathbf{\gamma}^T)^T \geq \mathbf{0}$, where $\tilde{M} = \left( \mathbf{T}^T_{0,R-1} \mathbf{0}_{R-1,C-1} \right)$</td>
</tr>
</tbody>
</table>

Here, $\mathbf{0}_{K_1,K_2}$ is the $K_1 \times K_2$ matrix of zeros, and $\mathbf{T}_{0,K}$ is as in Section 2.3.

Let $\Lambda_{\mu,v}$ be the LRT for testing $H_\mu$ against $A_v$, for $\mu \in \{g, c\}$ and $v \in \{r, c, m\}$. As described in Section 2.3., the finite sample permutation distribution of the LRTs can be easily simulated. The corresponding large sample distributions under the various null hypotheses are given by the following theorem:

**Theorem 3.** *Under the corresponding null hypotheses,*

\[
\lim_{n \to \infty} \mathbb{P}(\Lambda_{gc} > x) = \sum_{k=0}^{C-1} w_k \left( C - 1, \tilde{C}_1 \mathbf{\Sigma}_0 \tilde{C}_1^T \right) \mathbb{P} \left( \chi^2_{R-1+k} > x \right),
\]

\[
\lim_{n \to \infty} \mathbb{P}(\Lambda_{cc} > x) = \sum_{k=0}^{C-1} w_k \left( C - 1, \tilde{C}_1 \mathbf{\Sigma}_0^{(c)} \tilde{C}_1^T \right) \mathbb{P} \left( \chi^2_k > x \right),
\]

\[
\lim_{n \to \infty} \mathbb{P}(\Lambda_{gr} > x) = \sum_{k=0}^{R-1} w_k \left( R - 1, \tilde{R}_1 \mathbf{\Sigma}_0 \tilde{R}_1^T \right) \mathbb{P} \left( \chi^2_{C-1+k} > x \right),
\]

\[
\lim_{n \to \infty} \mathbb{P}(\Lambda_{rr} > x) = \sum_{k=0}^{R-1} w_k \left( R - 1, \tilde{R}_1 \mathbf{\Sigma}'_0 \tilde{R}_1^T \right) \mathbb{P} \left( \chi^2_k > x \right),
\]

\[
\lim_{n \to \infty} \mathbb{P}(\Lambda_{gm} > x) = \sum_{k=0}^{R+C-2} w_k \left( R + C - 2, \tilde{M} \mathbf{\Sigma}_0 \tilde{M}^T \right) \mathbb{P} \left( \chi^2_k > x \right),
\]

where $\tilde{C}_1$, $\tilde{R}_1$ and $\tilde{M}$ are defined in Table 1 and $\mathbf{\Sigma}_0$, $\mathbf{\Sigma}_0^{(c)}$ and $\mathbf{\Sigma}_0''$ are the covariance matrices of the asymptotic distribution of $\left( \hat{\mathbf{\beta}}_2, \ldots, \hat{\mathbf{\beta}}_R, \hat{\mathbf{\gamma}}_2, \ldots, \hat{\mathbf{\gamma}}_C \right)^T$ under the various nulls, given that model (21) holds.

Comparing the asymptotic chi-bar-square distributions in theorems 3 and 1, we see a reduction in the degrees of freedom. This gain (in degrees of freedom) occurs because model (21) is more parsimonious than model (1).

### 5. Simulations

The performance of the proposed methodology was investigated by simulation. In this section, we focus on the evaluation of the type I error and power of the proposed test for order restricted inference in the two-way layout. The superiority of the constrained estimators relative to the
unconstrained ones in terms of their mean squared error is demonstrated in Table S1 in the Supporting information.

In order to evaluate the proposed testing procedure, we start by generating $n_{ij}$ observations for $i = 1, \ldots, R$ and $j = 1, \ldots, C$ from the lognormal distribution with parameters 0 and 1. We consider the configurations $(R, C) = (2, 2), (2, 3), (3, 3)$, and for each one, we generated both balanced and unbalanced samples so that the total number of observations is $10RC$. The unbalanced designs for the aforementioned $R\times C$ configurations were $[(15, 10), (10, 5)], [(15, 10, 5), (8, 12, 10)]$ and $[(15, 8, 7), (12, 5, 8), (8, 17, 10)]$, respectively. Note that the lognormal assumptions means that model (1) holds with $\phi(x) = \log x$. We then test the null hypothesis (2) versus the alternative (8). We implemented the test $\Lambda_{gm}$ using 500 permutations. The type I error was calculated by 1000 simulation runs. We compare our test with two competitors, which we denote by $F_{gm}$ and $F_{tr}^{gm}$. The $F_{gm}$-test is the standard ANOVA-based test for order conducted on the original scale (Silvapulle and Sen, 2005), whereas the $F_{tr}^{gm}$ is the same test after applying the logarithmic transformation to each data point. (The superscript ‘tr’ indicates that the data are transformed.) Because the data are lognormal, the $F_{tr}^{gm}$-test is in fact the procedure of choice, that is, the likelihood ratio test for normally distributed observations. We also considered the $F_{gm}$-test based on $\phi(x) = x$.

Table 2 shows the nominal versus the actual significance levels. The actual levels of all tests are in general close to the nominal ones regardless the design is balanced or unbalanced. This is true also for the case of misspecified function $\phi(x) = x$. These results are expected because these tests are all based on resampling methods.

To further investigate the behaviour of these test statistics, we conduct a power study. We simulate data from the lognormal distribution for such parameter configurations so that the matrix order holds true. In particular, we draw data from the lognormal distribution

$$g_{ij}(y) = \frac{1}{\sqrt{2\pi}y} \exp\left\{-\frac{1}{2}(\log y - \mu_{ij})^2\right\}, \quad y > 0,$$

Table 2. Estimated size and power for testing hypothesis (2) versus (8) based on balanced (b) and unbalanced (u) small samples

<table>
<thead>
<tr>
<th>Design</th>
<th>$F_{gm}$-test</th>
<th>$F_{tr}^{gm}$-test</th>
<th>$\Lambda_{gm}$-test $\phi(x) = x$</th>
<th>$\Lambda_{gm}$-test $\phi(x) = \log x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>$C$</td>
<td>.05</td>
<td>.10</td>
<td>.05</td>
</tr>
<tr>
<td>2</td>
<td>2 (b)</td>
<td>.036</td>
<td>.094</td>
<td>.049</td>
</tr>
<tr>
<td>2</td>
<td>2 (u)</td>
<td>.053</td>
<td>.105</td>
<td>.055</td>
</tr>
<tr>
<td>2</td>
<td>3 (b)</td>
<td>.044</td>
<td>.082</td>
<td>.046</td>
</tr>
<tr>
<td>2</td>
<td>3 (u)</td>
<td>.050</td>
<td>.090</td>
<td>.062</td>
</tr>
<tr>
<td>3</td>
<td>2 (b)</td>
<td>.046</td>
<td>.110</td>
<td>.062</td>
</tr>
<tr>
<td>3</td>
<td>3 (u)</td>
<td>.060</td>
<td>.098</td>
<td>.050</td>
</tr>
<tr>
<td></td>
<td>Estimated size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 (b)</td>
<td>.526</td>
<td>.670</td>
<td>.627</td>
</tr>
<tr>
<td>2</td>
<td>2 (u)</td>
<td>.455</td>
<td>.608</td>
<td>.575</td>
</tr>
<tr>
<td>2</td>
<td>3 (b)</td>
<td>.654</td>
<td>.772</td>
<td>.786</td>
</tr>
<tr>
<td>2</td>
<td>3 (u)</td>
<td>.590</td>
<td>.754</td>
<td>.794</td>
</tr>
<tr>
<td>3</td>
<td>2 (b)</td>
<td>.802</td>
<td>.880</td>
<td>.944</td>
</tr>
<tr>
<td>3</td>
<td>3 (u)</td>
<td>.854</td>
<td>.928</td>
<td>.972</td>
</tr>
</tbody>
</table>

Under (2), the data are generated from the lognormal distribution with parameters 0 and 1 while under (8) from the lognormal distributions specified by (22). In each case, the number of simulations is 1000 and of permutations 500.
with
\[ \mu_{ij} = \alpha_i + \beta_j, \quad \alpha_i = i/R, \ i = 0, 1, \ldots, R - 1, \quad \beta_j = j/C, \ j = 0, 1, \ldots, C - 1. \]  

(22)

The simulation results are summarized also in Table 2. Note that the power of the proposed likelihood ratio test \( \Lambda_{gm} \) with the correct function \( \phi(x) = \log x \) is practically equivalent to the power of its parametric counterpart \( F_{gm}^{tr} \) for testing (2) versus (8). On the other hand, our test based on \( \phi(x) = x \) has less power, yet it is always more powerful than the \( F_{gm} \)-test. For example, when \( R = C = 2 \) and the nominal level equals 5 per cent, the power of \( \Lambda_{gm} \)-test based on \( \phi(x) = \log x \) was .639, whereas that of \( F_{gm}^{tr} \)-test was .627. On the other hand, the power of \( \Lambda_{gm} \)-test based on \( \phi(x) = x \) was .585, whereas the power of \( F_{gm} \)-test was only .526.

Furthermore, we evaluated the performance of the proposed test based on \( \Lambda_{gm} \) when the function \( \phi \) is misspecified. For this purpose, we first fitted model (1) by choosing \( \phi(x) = \sqrt{x} \) and \( x \) instead of the correct function \( \phi(x) = \log x \). Results of these simulations are displayed in Table 3. The type I error of the tests is close to the nominal level.

We also observe that the power is largest when the model is correctly specified. When \( \phi \) is misspecified, the power is somewhat reduced. Observe also that the power when using \( \sqrt{x} \) is larger than when using \( x \). This happens because \( \sqrt{x} \) is ‘closer’ to the correct function \( \log x \) than \( x \). It is also important to note that the power of \( \Lambda_{gm} \)-test is comparable with that of the original ANOVA-based test \( F_{gm}^{tr} \) even under misspecification of \( \phi \).

In order to further investigate the effect of \( \phi \) misspecification to the power of our test, we considered several functions within the Box–Cox family with the corresponding parameter \( \zeta \) ranging from \(-2\) to \(2\). The results are displayed in Table 4 and confirm our previous statement that the power decreases as \( \phi \) gets farther from the correct choice. Because the underlying lognormal distribution corresponds to \( \zeta = 0 \), the maximum power is attained at this point. Observe though that the loss of power is not more than 25 per cent when \( \zeta = \pm 2 \).

A similar study when the underlying distribution is exponential can be found in Table S2.

We conclude that the proposed methodology works well. It is highly efficient relative to the fully parametric analysis, and it performs better than the standard method when the underlying parametric assumptions do not hold. Moreover, the methodology works well even if the function \( \phi \) is misspecified. This is especially true when the misspecified model is ‘close’ to the true model. This is a great advantage in applications.

### 6. Real data examples

In this section, we illustrate the proposed methodology using two real data examples. In the first example, the response variable is discrete, and the group sample sizes are moderate, whereas in the second, the response is continuous, and the sample sizes are small. The examples are used for illustrative purposes because, in practice, the ordering among the distributions should be specified (and justified) prior to examination of the data.

<table>
<thead>
<tr>
<th>( R )</th>
<th>( C )</th>
<th>( \phi(x) = \log x )</th>
<th>( \phi(x) = \sqrt{x} )</th>
<th>( \phi(x) = x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>.059</td>
<td>.120</td>
<td>.053</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>.051</td>
<td>.100</td>
<td>.061</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>.040</td>
<td>.083</td>
<td>.052</td>
</tr>
</tbody>
</table>

The data are generated from the lognormal distributions (and so the correct choice would be \( \phi(x) = \log x \)) specified in Section 5. The number of simulations is 1000, and the number of observations per cell is 15.
Table 4. Power of 5 per cent-level and 10 per cent-level $\Lambda_{gm}$-test as functions of the Box–Cox transformation parameter $\xi$ under two $2 \times 2$ configurations for the lognormal distribution described in Section 5 on the basis of balanced and unbalanced samples

<table>
<thead>
<tr>
<th>$\xi$</th>
<th>$\alpha$-level</th>
<th>Balanced $2 \times 2$ design ($n_{11} = n_{12} = n_{21} = n_{22} = 15$)</th>
<th>Power of $\Lambda_{gm}$-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.05</td>
<td>.59 .63 .70 .74 .77 .78 .79 .77 .75 .71 .67 .63 .58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.10</td>
<td>.71 .75 .83 .85 .87 .88 .89 .88 .86 .84 .80 .75 .71</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\xi$</th>
<th>$\alpha$-level</th>
<th>Unbalanced $2 \times 2$ design ($n_{11} = 15$, $n_{12} = n_{21} = 10$, $n_{22} = 5$)</th>
<th>Power of $\Lambda_{gm}$-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.05</td>
<td>.41 .44 .47 .47 .47 .50 .50 .47 .46 .44 .41 .37 .35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.10</td>
<td>.57 .60 .64 .65 .66 .67 .68 .65 .60 .58 .57 .53 .49</td>
<td></td>
</tr>
</tbody>
</table>

The number of simulated samples is 1000, and the number of random permutations is 500. The correct choice is $\phi(x) = \log x$, which corresponds to $\xi = 0$. Among the tests based on these transformations, the largest power corresponds to this choice, and the power decreases as $\xi$ moves away from zero.

6.1. Days absent from school

Aitkin (1978) reported on the number of days a student was absent from school, his/her ethnic background (foreigner or native), gender (female or male), age and learner status. The dataset can be found in MASS library of R. We focus exclusively on the relationship between the response (days absent from school) with ethnic background and gender. Figure S2 in the Supporting information shows boxplots of the response by the levels of the two factors. Let $Y_{ijk}, k = 1, \ldots, n_{ij}$, denote the values of the response variable at the level $i$ of ethnicity and $j$ of gender for $i, j = 1, 2$. We use the coding Foreigner = 1 and Native = 2 for the factor ethnicity and Female = 1 and Male = 2 for the factor gender. The group sample sizes are $(n_{11}, n_{12}, n_{21}, n_{22}) = (42, 38, 35, 31)$.

We start by fitting a standard two-way ANOVA model with interaction to the data we see that there is high heteroscedasticity (Levene’s test $p$-value $\approx .001$) and the residuals are very far from normality. A suitable transformation to normality for these data is $Y_{ijk} \mapsto \log(Y_{ijk} + 2)$ leading to the two-way ANOVA model

$$\log(Y_{ijk} + 2) = \mu + \beta_i + \gamma_j + \theta_{ij} + \epsilon_{ijk}$$

(23)

where $\beta_i$ is the main effect of ethnicity, $\gamma_j$ is the main effect of gender and $\theta_{ij}$ denotes the interaction of the two factors, $i, j = 1, 2$. We set $\beta_1 = \gamma_1 = \theta_{11} = \theta_{12} = \theta_{21} = 0$ for identifiability. Fitting the ANOVA model (23), we find that there is no significant interaction between the two factors ($p = .755$) nor a significant gender effect ($p = .351$). On the other hand, there is a highly significant ethnicity effect ($p < 10^{-4}$). The estimate of $\gamma_2$ has a positive value, and this indicates that natives tend to be absent from school more often than foreigners. More precisely, the mean of the transformed number of days absent from school for natives is larger than the corresponding mean for foreigners. For model (23), the normal distribution fits well to the residuals, although there is still some indication for heteroscedasticity (Levene’s test $p$-value $\approx .047$).

We applied the proposed methodology to this data example. We fit model (1) to the data using the identity function $\phi(x) = x$. We found that $\hat{\lambda}_{12} = .0324(.0222), \hat{\lambda}_{21} = .0559(.0187)$ and $\hat{\lambda}_{22} = .0579(.0187)$. The values inside the parentheses correspond to the standard errors.
of the estimates and have been estimated by 1000 bootstrap iterations within each group. The p-value for the permutation test of $H^G$ versus $A^m$ was found to be very small ($< .001$). Hence, we may conclude that there is an ordering of the number of days a student is absent from school by his/her ethnicity and gender. This was also the case for the test of $H^G$ versus $A^r$. On the other hand, the p-value for the test of $H^G$ versus $A^c$ was found to be non-significant ($\approx .122$), which means that there is no statistically significant gender effect. On the basis of these two facts, we conclude that the response variable is ordered with respect to the student’s ethnicity in the sense of likelihood ratio ordering. In particular, this implies that the number of days natives are absent from school is stochastically larger than the corresponding number days for foreigners, and this in turn means that the mean number of days for natives is larger than the mean number of days for foreigners. Hence, our conclusions agree with those from a standard ANOVA model, but they are stronger because they refer to the distribution of the response variable rather than its mean and do not require any assumptions about the underlying distribution nor data transformation.

Let us now model the factor effects as described in Section 4. Note that the estimate of the interaction effect is $\hat{\theta}_{22} = \hat{\lambda}_{22} - \hat{\lambda}_{12} - \hat{\lambda}_{21} = -.0304(.0263)$, which indicates that there is no significant interaction between the two factors. Moreover, it seems that $\hat{\lambda}_{12}$ is not significant, which also suggests that there is only a row factor (i.e. ethnicity) effect. Having found that there is no interaction, we proceed by fitting the additive model (21). The p-value for the test of $H^G$ versus $A^m$ was found to be very small ($p < 10^{-3}$), so we again conclude that there is indeed ordered factor effects. Next, we test $H^r$ versus $A^r$, that is, no row (ethnicity) effect versus ordered row effect. Here, we found the same as in the previous test, that is, there is an ordered row effect. On the other hand, the p-value for the test of $H^c$ versus $A^c$ was found approximately equal to .130. Hence, we conclude once more that there is no significant column (i.e. gender) effect, and the row ordered effect model is adequate.

The aforementioned row effect model estimates the ratio of the densities $g(x|\text{Native})/g(x|\text{Foreigner})$ by the quantity $\exp(-.621 + .039x)$. As noted in Section 1, because the two distributions are ordered by the likelihood ratio order, there exists a value $x^*$ such that the probability that a native is absent for $x$ days is larger than the probability that a foreigner is absent for $x$ days for all $x > x^*$. Under the aforementioned model, this value is estimated to be $\hat{x}^* = (-.621)/(.039) \approx 16.1$ days ($\hat{\sigma} = 1.36$ on the basis of 1000 bootstrap iterations; 95 per cent bootstrap ci $= [13.1, 18.6]$). Once again, note that this is a much stronger conclusion than simply ordering the means of the two distributions.

Because the true functional form of $\phi$ is unknown, we repeated the aforementioned analysis using several different functions within the Box–Cox family of transformations. Because of the existence of some zero data values, we first add one to all observations. By maximizing the likelihood, we found that the optimal value of the Box–Cox exponent was approximately $-1/2$. This is equivalent to fitting model (1) with $\phi(x) = -1/\sqrt{x + 1}$. Motivated by (23), we also applied some Box–Cox transformations to $\log(x + 2)$. The conclusions are essentially the same as those obtained by using the identity function; see Table S3. This is not surprising and agrees with the results of Section 3. Note here that the p-values vary with $\phi$. This is expected, because the power of the test depends on how close our choice of $\phi$ is to the correct function. However, in our simulation study, they all are insignificant.

### 6.2. Effects of nitrogen and phosphorus on maize crop

The effect of nitrogen and phosphorus on the yield of maize was investigated at the Ludhiana Centre of All India Coordinated Research Project. This particular dataset comes from an experiment conducted in 1998. This is a balanced factorial experiment with three factors,
namely, nitrogen (40, 80 and 120 kg/ha), phosphorus (0, 40 and 80 kg/ha) and potassium (0 and 40 kg/ha). By considering only the first two factors, we obtain a two-way layout dataset with \(3 \times 3 = 9\) treatment combinations and \(n_{ij} = 8\) observations per cell; see Table S4 and Figure S3. The complete data can be found at http://www.iasri.res.in/design.

A preliminary analysis using a standard two-way ANOVA reveals that there is no significant interaction \((p \approx .142)\) and that both factors affect crop yield (both \(p < 10^{-6}\)). However, the residuals display some heteroscedasticity (Levene’s test \(p\)-value \(\approx .046\)); therefore, we apply our method to these data.

By fitting model (1) with \(\phi\) the identity function and testing \(H_g\) versus any alternative among \(A_m, A_r\) and \(A_c\), we find in all cases the \(p\)-value to be approximately zero. This implies that there are both row and column ordered effects. We then proceed with the methodology described in Section 4. Because the sample sizes are small, we first test for interaction as follows. We fit the (unordered) additive model (21) with \(\phi\) the identity function and estimate the underlying distributions \(G_{ij}\) by \(\hat{G}_{ij}(x) = \sum_{k=1}^{72} \hat{g}_{ij}(Y_k)1(Y_k \leq x), i, j = 1, 2, 3\). The estimated parameter values are

\[
\hat{A} = \begin{pmatrix}
0 & -12.70 & -17.84 \\
-9.38 & -23.83 & -30.07 \\
-22.50 & -40.52 & -47.81
\end{pmatrix}, \quad \hat{\beta} = (0, 1.447, 3.214), \quad \hat{\gamma} = (0, 1.926, 2.627).
\]

As described in Section 4, we bootstrap \(B = 1000\) datasets with eight observations per cell from \(\hat{G}_{ij}\)’s, \(i, j = 1, 2, 3\), and record the corresponding values of \(\Lambda_i = \{\sup_l(A, A) - \sup_{H_l} l(A, A)\}\). The bootstrap \(p\)-value for the test for interaction is defined as the percentage of simulated values of \(\Lambda_i\) that exceed the observed one. This is found equal to \(.466\), so we conclude that the additive model fits adequately to the data.

We now proceed to test for order. Observe that the estimated parameters are already ordered. This is an indication that favours the hypothesis of ordering. Indeed, the estimated \(p\)-value of the permutation test for \(H_g\) versus \(A_m\) was zero because for all simulated datasets, \(\Lambda_{gm}\) was found smaller than the observed one. We conclude that increasing either nitrogen or phosphorus level has a growing effect to the crop and that the two chemicals do not interact.

In closing this example, we note that we repeated the aforementioned procedure by using the functions \(\phi(x) = \log x, \sqrt{x}, x^2, x/(x + 1), -1/x\) as well. The corresponding \(p\)-values for the interaction test ranged from \(.39\) to \(.90\), whereas for the ordering test were also found equal to zero under both general and additive models. Note that the \(p\)-values for the interaction test vary. Nevertheless, they are all nonsignificant. Thus, we obtained the same inferences with all of the aforementioned \(\phi\) functions.

7. Discussion

In this paper, we have shown that the density ratio model provides a general semiparametric modelling framework for analysing data from two-way layouts without resorting to fully parametric assumptions. The density ratio model is especially well suited for order restricted inference. We start by showing that general hypothesis concerning order relations can be reduced to simple inequalities on the model parameters. Estimation procedures with and without order restrictions are developed. A simple, easy-to-implement algorithm for constrained estimation is presented. Simulations show that the constrained estimators improve, often substantially, on the unrestricted ones. It is worth noting that our estimators are also highly efficient relative to the full MLEs. Likelihood ratio type tests are developed, and their large sampling distributions are found. The proposed methodology performed well in our
simulation study. Mean square errors were improved, and the power of the test for (8) was comparable with that of standard parametric tests.

Besides its simplicity, model (1) has a very nice property. As shown in Section 3 and illustrated by the simulations, it captures the ordering of the distributions even if the function φ is misspecified. This is extremely important because φ is typically unknown and difficult to determine. In fact, our analysis shows that the tests are consistent even if the density ratio model does not hold.

The proposed methodology can be further extended and generalized in several directions. A natural first step would be an extension to the multiway setting. This seems straightforward. In this paper, we have focused on the simple ordering, that is, ordering treatment along rows and/or columns. However, many other orderings can be considered. For instance, there may be some level \( j^* \) of the column factor with \( 1 < j^* < C \), such that \( \lambda_{i1} \leq \cdots \leq \lambda_{i,j^*+1} \leq \lambda_{ij^*} \) and \( \lambda_{ij^*} \geq \lambda_{i,j^*+1} \geq \cdots \geq \lambda_{iC} \) for all \( i \). This corresponds to the umbrella ordering (cf. Silvapulle and Sen, 2005) with respect to the column factor. Moreover, the ‘top’ level \( j^* \) could depend on the row level \( i \) as well. In fact, any testing problem that can be formulated as \( \Delta \lambda = 0 \) versus \( \Delta \lambda \geq 0 \) can be treated similarly.

We have focused on testing for the likelihood ratio ordering. Tests against the likelihood ratio order can also be developed. For example, one could consider testing hypothesis of the form \( \Delta \lambda \geq 0 \) versus \( \Delta \lambda \neq 0 \). This is a situation where the null hypothesis contains inequalities. Such problems can be handled as described in Silvapulle and Sen (2005, Section 3.8).

References


Received October 2012, in final form September 2013

George Iliopoulos, Department of Statistics and Insurance Science, University of Piraeus, Karaoli & Dimitriou 80, 18534 Piraeus, Greece.
E-mail: geh@unipi.gr

Supporting information

Additional supporting information may be found in the online version of this article at the publisher’s web site.

Appendix A. Some results on the unordered case, including Theorem A.1, which states that, in some cases, misspecification of the density ratio model can lead to inconsistent tests.
Appendix B. Proofs of theoretical results.
Appendix C. Additional tables and figures to Sections 5 and 6 as well as numerical results.
Supplementary material to the paper
Semiparametric inference for the two-way layout
under order restrictions
Ori Davidov       Konstantinos Fokianos       George Iliopoulos

Appendices

A The unordered case

A.1 The profiling procedure

The empirical likelihood in (10) has unique maximizer if there is no nontrivial partition
\( \{M_1, M_2\} \) of \( M \) such that

\[
\left[ \bigcup_{(i,j) \in M_1} (\min S_{ij}, \max S_{ij}) \right] \cap \left[ \bigcup_{(i,j) \in M_2} (\min S_{ij}, \max S_{ij}) \right] = \emptyset,
\]

see Davidov and Iliopoulos (2009). Maximizing (10) for fixed \((A, \Lambda)\) is equivalent to finding

\[
\max \left\{ \prod_{i=1}^{n} p_k : p \succ 0, \sum_{k=1}^{n} p_k \exp(\alpha_{ij} + \lambda_{ij} Z_k) = 1, \forall (i,j) \in M \right\},
\]

where \( p \succ 0 \) means that all \( p_k \)'s must be positive. Using Lagrange multipliers we obtain the Langrange function

\[
\sum_{k=1}^{n} \log p_k + \sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij}(\alpha_{ij} + \lambda_{ij} Z_{ij}) + \nu_{11} \left( \sum_{k=1}^{n} p_k - 1 \right) + \sum_{(i,j) \neq (1,1)} \nu_{ij} \left( \sum_{k=1}^{n} p_k e^{\alpha_{ij} + \lambda_{ij} Z_k} - 1 \right).
\]

By equating the derivative with respect to \( \alpha_{ij} \) to zero we get \( n_{ij} + \nu_{ij} \sum_{k=1}^{n} p_k e^{\alpha_{ij} + \lambda_{ij} Z_k} = 0 \) which together with the corresponding constraint shows that \( \nu_{ij} = -n_{ij} \) for all \( (i, j) \neq (1,1) \). Next, equating the derivative with respect to \( p_k \) to zero yields

\[
0 = \frac{1}{p_k} + \nu_{11} + \sum_{(i,j) \neq (1,1)} \nu_{ij} e^{\alpha_{ij} + \lambda_{ij} Z_k}
\]

which together with the corresponding constraint shows that \( \nu_{ij} = -n_{ij} \) for all \( (i, j) \neq (1,1) \).
By adding over \( k \) and using the constraints we get that \( \sum \nu_{ij} = -n \). This implies that \( \nu_{11} = -n_{11} \) as well. Replacement of all Lagrange multiplies in (A.1) yields

\[
p_k = \left\{ \sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij} \exp(\alpha_{ij} + \lambda_{ij} Z_k) \right\}^{-1}, \quad \text{for } k = 1, \ldots, n
\]

(recall that \( \alpha_{11} = \lambda_{11} = 0 \)). Finally, by substituting these \( p_k \)'s back to (10) and taking the logarithm leads to (11). The MLE of \((A, A)\) is obtained by equating the partial derivatives of (11), namely,

\[
\frac{\partial}{\partial \alpha_{ij}} l(A, A) = n_{ij} \left\{ 1 - \frac{\sum_{k=1}^{n} \exp(\alpha_{ij} + \lambda_{ij} Z_k)}{\sum_{k=1}^{n} \sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij} \exp(\alpha_{ij} + \lambda_{ij} Z_k)} \right\},
\]

\[
\frac{\partial}{\partial \lambda_{ij}} l(A, A) = n_{ij} \left\{ \bar{Z}_{ij} - \frac{\sum_{k=1}^{n} Z_k \exp(\alpha_{ij} + \lambda_{ij} Z_k)}{\sum_{k=1}^{n} \sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij} \exp(\alpha_{ij} + \lambda_{ij} Z_k)} \right\}
\]
to zero.

**A.2 Testing**

As mentioned in the Introduction, \( H_g, H_c, \) and \( H_r \) can be expressed in terms of equalities between \( \lambda_{ij} \)'s. In fact, all hypotheses of interest are of exactly the same form as in a standard two-way ANOVA with normal errors but expressed in terms of \( \lambda_{ij} \)'s instead of the population means. Since \( \lambda \) has an asymptotic normal distribution, see (13), all asymptotic (unordered) tests against \( H_g, H_c, \) and \( H_r \) may be performed similarly to the well-known normal case. Furthermore, the parameters \( \lambda_{ij} \) can be decomposed using standard models as that in (19) so that one can test for interaction and/or additive effects.

In order to calculate the test statistics, the asymptotic covariance matrix \( \Psi \) in (13) is needed. Fortunately, the elements of the matrices \( S \) and \( V \) are expectations of functions with respect to the baseling distribution \( f \) and can be consistently estimated by the data. However, due to the complicated notations we omit the presentation of their explicit forms. They can be derived as in Fokianos et al. (2001) and Davidov et al. (2009). On the other hand, in case the sample sizes are moderate to small one can use either bootstrap or permutation tests.

**A.3 Model misspecification**

In the unordered case, inference based on the density ratio model can lead to erroneous results when the model is misspecified even in large samples. This important issue has not been recognized so far in the literature. Consider for example the situation where \( G_{ij} \neq G_{i'j'} \) but neither \( G_{ij} \succ_{lr} G_{i'j'} \) nor \( G_{ij} \prec_{lr} G_{i'j'} \) holds. Suppose now that we fit model (1) with some (not necessarily monotone) function \( \psi \).

**Theorem A.1.** Suppose that the alternative hypothesis \( A_g \) in (5) holds. If \( \mathbb{E}_{G_{ij}} \{ \psi(X) \} \) is constant for all \((i, j) \in M\), then the asymptotic power of the test equals its asymptotic size.
Figure S.1: Simulated distribution functions of p-value when $G_{11} = G_{12} \sim G(2, 1)$ and $G_{21} = G_{22} \sim G(1, 2)$ for three different choices of $\phi$, namely, $\phi_0(x) = \log x - x/2$ (solid line), $\phi_1(x) = \log x$ (dashed line) and $\phi_2(x) = x$ (dotted line). The table on the right presents the corresponding powers for the 5%- and 10%-level tests.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\phi_0(x)$</th>
<th>$\phi_1(x)$</th>
<th>$\phi_2(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.05</td>
<td>.307</td>
<td>.116</td>
<td>.063</td>
</tr>
<tr>
<td>.10</td>
<td>.436</td>
<td>.193</td>
<td>.101</td>
</tr>
</tbody>
</table>

Theorem A.1 implies that in some situations the power to detect a difference between the distributions equals the nominal level for large sample sizes.

This phenomenon is illustrated graphically in Figure S.1. Here we consider a $2 \times 2$ design where $G_{11} = G_{12} \sim G(2, 1)$, i.e. the gamma distribution with shape 2 and scale 1 and $G_{21} = G_{22} \sim G(1, 2)$, that is, the exponential distribution with scale 2. Notice that in this case the distributions $G_{1j}$ and $G_{2j}$ are unordered yet satisfy the density ratio model with the function $\phi_0(x) = \log x - x/2$. We explore the power of the test also for the misspecified functions $\phi_1(x) = \log x$ and $\phi_2(x) = x$. Note that the expected value under $G_{ij}$ of $\phi_2$ equals 2 for all $i, j$. Thus, under Theorem A.1 we expect the power of the $\Lambda_{gg}$-test (that is, the test for $H_g$ against $A_g$) based on $\phi_2$ to be close to the nominal level. Let $\hat{F}_i$ denote the distribution function of the p-value of the test based on $\phi_i$, $i = 0, 1, 2$. In Figure S.1 we plot these distribution functions. The results are based on 1000 simulations with $n_{ij} = 15$ for all $i, j$. The powers of the corresponding 5% and 10%-level $\Lambda_{gg}$-tests are displayed next to the figure.

Note that $\hat{F}_2(u)$ is very close to $u$ for all $u \in (0, 1)$. This means that the p-values from the test based on $\phi_2$ are approximately uniformly distributed. In particular, $\hat{F}_2(.05) = .063$ and $\hat{F}_2(.10) = .101$ confirming the statement of Theorem A.1. This simulation study confirms empirically and in a reasonable setting, that it is possible to have very low power for distinguishing among unordered distributions under the density ratio model.

Note further that $\hat{F}_0(u) \geq \hat{F}_1(u) \geq \hat{F}_2(u)$ for all $u \in (0, 1)$. This ordering among the distribution functions corresponds to the ordering among the powers of the tests.
A.4 An extension

Model (1) can be generalized considerably. For example, consider

\[ g_{ij}(x) = f(x) \exp\{\alpha_{ij} + \beta_i(x) + \gamma_j(x) + \theta_{ij}(x)\}, \quad (i, j) \in \mathcal{M}. \]  

(A.2)

This offers more flexibility since, under (A.2), the row, column and interaction effects are measured in possibly different scales. Nevertheless, model (A.2) is more suitable for detecting differences among factor levels rather than testing for order. At present, it is not clear how to determine the \( \phi \)-functions. As we already saw in Sections 3 and A.3, choosing a single function \( \phi \) is a nontrivial task. Obviously, the determination of three functions is a much more difficult problem.
B  Proofs

Lemma B.1. Assume that \( g_{ij}(y) = f(y) \exp\{\alpha_{ij} + \lambda_{ij}\phi(y)\} \) for \( i \in \mathcal{R} \) and \( j \in \mathcal{C} \), where \( \phi \) is some given function. Then, for any fixed \( \Lambda = (\lambda_{ij}; i \in \mathcal{R}, j \in \mathcal{C}) \), the function

\[
l_1(A) = \sum_{i=1}^{R} \sum_{j=1}^{C} \sum_{Y_k \in S_{ij}} \left( \alpha_{ij} + \lambda_{ij}\phi(Y_k) \right) - \sum_{k=1}^{n} \log \left[ \sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij} \exp\{\alpha_{ij} + \lambda_{ij}\phi(Y_k)\} \right]
\]

has a unique maximizer \( \tilde{A}(\Lambda) \), say, provided there is no nontrivial partition \((\mathcal{M}_1, \mathcal{M}_2)\) of \( \mathcal{M} = \{(i, j) : i \in \mathcal{R}, j \in \mathcal{C}\} \) such that

\[
\bigcap_{t=1}^{2} \bigcup_{(i, j) \in \mathcal{M}_t} \left( \min\{\phi(Y_k) : Y_k \in S_{ij}\}, \max\{\phi(Y_k) : Y_k \in S_{ij}\} \right) = \emptyset.
\]

Moreover, \( \tilde{A} \) is continuous in \( \Lambda \).

Proof. This is a consequence of the standard theory of nonparametric maximum likelihood estimation with several biased samples; see Vardi (1985), and Davidov and Iliopoulos (2009). Indeed, for any fixed \( \Lambda \), we observed \( RC \) independent random samples from corresponding biased versions of \( F \) with the biasing function for \( G_{ij} \) being \( w_{ij}(y) = \exp\{\lambda_{ij}\phi(y)\} \). The continuity of \( \tilde{A} \) follows by the concavity of the function \( l_1 \).

Remark B.1. Lemma B.1 implies that we are allowed to further profile the log likelihood of any of the previously mentioned models with respect to \( A \). In particular, notice that \( \tilde{A}(\hat{A}) = \hat{A} \), \( \tilde{A}(A^*) = A^* \), and \( \tilde{A}(A^{**}) = A^{**} \).

Algorithm for estimation under order restrictions

Algorithm 1.

1. Find first the isotonic regression of \( \overline{Z} = (Z_{ij}) \) with weights \( n = (n_{ij}) \) with respect to the matrix order. In other words, find

\[
U^* = \arg \min \left\{ \sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij}(U_{ij} - Z_{ij})^2 : U_{ij} \leq U_{i'j'}, \forall (i, j) \leq (i', j') \right\}. \tag{B.1}
\]

Note that (B.1) is a quadratic programming problem which can be solved by a Pool Adjacent Violators-type algorithm (cf. Robertson et al., 1988).

2. Let \( L_k, k = 1, \ldots, K \), where \( 1 \leq K \leq RC \) be the level sets of \( \mathcal{M} \) derived from the isotonic regression (B.1). Note that \((i, j)\) and \((i', j')\) are in the same level set if \( U_{ij}^* = U_{i'j'}^* \). Merge the samples belonging to the same level sets, i.e., for \( k = 1, \ldots, K \) let \( S_k^0 = \bigcup_{(i, j) \in L_k} S_{ij} \)
3. If \( K = 1 \) set \( A^* = \Lambda^* = 0 \) otherwise fit the unconstrained one-way model to the samples \( S_0^1, \ldots, S_K^k \) and denote the corresponding maximizer by \((\alpha^0, \lambda^0)\).

4. Set \((\alpha_{ij}^*, \lambda_{ij}^*) = (\alpha_k^0, \lambda_k^0)\) if \((i, j) \in L_k\).

**Theorem B.1.** Algorithm 1 yields the restricted MLE \((A_m^*, \Lambda_m^*)\) of \((A_m, \Lambda_m)\).

**Proof.** The proof of this theorem extends Theorem 1 in Davidov et al. (2010) from the one-way to the multiway setup. Briefly, in (16) and (B.1) we seek the extremes of the corresponding functions under some inequality restrictions. Moreover, it is easy to see that in both cases it suffices to identify the samples to be merged. Using the score equations of \(\Lambda\),

\[
\frac{\partial}{\partial \lambda_{ij}} l(A, \Lambda) = n_{ij} \left\{ Z_{ij} - \sum_{k=1}^{n} \frac{Z_k \exp(\alpha_{ij} + \lambda_{ij} Z_k)}{\sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij} \exp(\alpha_{ij} + \lambda_{ij} Z_k)} \right\} = 0, \quad (i, j) \in \mathcal{M},
\]

it can be seen that the conditions of Kuhn–Tucker Theorem (see for example Sundaram, 1996) completely match for the two problems which means that they must have the same solution.

The restricted MLE and the projection estimator given by (17) are related by the following:

**Theorem B.2.** If \( A \in \mathcal{L}_m \) then both \((A_m^*, \Lambda_m^*)\) and \((A_m^{**}, \Lambda_m^{**})\) are consistent for \((A, \Lambda)\).

**Proof.** Since \( \tilde{\Lambda}_m^{**} = \Pi_{\Sigma}(\tilde{\Lambda}_m | \mathcal{L}_m) \) it holds

\[
0 \leq (\lambda^{**}_m - \tilde{\lambda}_m)^T \Sigma^{-1} (\lambda^{**}_m - \tilde{\lambda}_m) \leq (\lambda - \tilde{\lambda}_m)^T \Sigma^{-1} (\lambda - \tilde{\lambda}_m)
\]

for all \( \lambda \in \mathcal{L}_m \). By the consistency of \( \tilde{\lambda}_m \) the last quadratic form is \( o_P(1) \), so \( \lambda_m^{**} \) is consistent too. Now, it can be shown (cf. Silvapulle and Sen, 2005) that the likelihood admits a quadratic approximation and this implies that the distance between \( \lambda_m^* \) and \( \lambda_m^{**} \) is also \( o_P(1) \) which means that \( \lambda_m^* \) is also consistent. Finally, the consistency of \( A^*, A^{**} \) follows by the continuity of \( \tilde{A} \) (see Lemma B.1 as well as Remark B.1.)

Algorithm 1 can be suitably modified to find the constrained MLEs under hypotheses (6) and (7). For example, under (6), we minimize \( \sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij} (U_{ij} - Z_{ij})^2 \) subject to the constraint that \( U_{ij} \leq U_{ij'} \) for all \( i \in \mathcal{R} \) and \( j \leq j' \in \mathcal{C} \). A similar result holds for (7). In both cases proceed with steps (2)–(4) as outlined above. Analogous results to Theorems B.1 and B.2 can be established.
Proof of Theorem 1

Consider first the LRTs when the null hypothesis is $H_0 : \mathbf{A} = \mathbf{0}$. Let $\tilde{l}(\mathbf{A}) = l(\hat{\mathbf{A}}(\mathbf{A}), \mathbf{A})$ be the log profile likelihood after having further profiled with respect to $\mathbf{A}$. Then the LRT can be expressed as

$$\Lambda_{\mu \nu} = 2\{\sup_{\lambda \in \mathcal{L}_\nu} \tilde{l}(\lambda) - \tilde{l}(0)\}$$

(see also Silvapulle and Sen, 2005, Proposition 4.2.2(C)). It follows from (B.2) and (B.3) that

$$\Lambda_{\mu \nu} = n\hat{\lambda}^T \Sigma_0^{-1} \hat{\lambda} - \min_{\lambda \in \mathcal{L}_\nu} \{n(\hat{\lambda} - \lambda)^T \Sigma_0^{-1}(\hat{\lambda} - \lambda)\} + o_P(1)$$

where $\sqrt{n}(\hat{\lambda} - \lambda) \Rightarrow \mathcal{N}_{RC-1}(0, \Sigma_0)$ as $n \to \infty$. Now, standard theory shows that $\Lambda_{\mu \nu} \Rightarrow \chi^2(\Sigma_0, C_\nu)$ where $C_\nu = \{\lambda : M\lambda \geq 0\}$ is the convex cone defined by $A_\nu$, and $\chi^2(\Sigma_0, C_\nu)$ denotes the chi-bar-square distribution. The formula in the statement of the proof follows by applying Theorem 3.7.2 in Silvapulle and Sen (2005) noting that $\text{rank}(\mathbf{M}) = RC - 1$. Similar arguments can be made when $\mu \neq g$ using the fact that $\text{rank}(\mathbf{R}_0) = RC - 1, \text{rank}(\mathbf{R}_1) = R(C - 1), \text{rank}(\mathbf{C}_0) = RC - 1$, and $\text{rank}(\mathbf{C}_1) = C(R - 1)$.

Lemma B.2. Let $H_i, i = 1, 2$, be distributions with support $\mathcal{X}$ satisfying $H_1 \prec_{\mu} H_2$. Let $g_i(x) = f(x)e^{\alpha(\beta) + \beta \psi(x)}$, $i = 1, 2$, be their best approximating densities with respect to the Kullback–Leibler divergence in the family $g(x) = f(x)e^{\alpha(\beta) + \beta \psi(x)}$, where $f$ is some fixed density function with support $\mathcal{X}$ and $\psi$ is a strictly increasing function in $\mathcal{X}$. Then $\beta_1 < \beta_2$.

Proof. By differentiating both sides of the equation $\int g(x)dx = 1$ with respect to $\beta$, we get that $\alpha'(\beta) = -\mathbb{E}_g\{\psi(X)\}$. Since the family $g$ has the monotone likelihood ratio property with respect to $\beta$ and $\psi$ is a strictly increasing function, it follows that $\mathbb{E}_g\{\psi(X)\}$ is a strictly increasing function of $\beta$ and thus $\alpha'$ is strictly decreasing in $\beta$.

The best approximating values $\beta_i$, $i = 1, 2$, with respect to the Kullback–Leibler (KL) divergence maximize $\int \log g(x)dH_i(x)$, $i = 1, 2$, respectively, with respect to $\beta$ (see, for example, Claeskens and Hjort, 2008, p. 25). Setting the derivatives of these integrals with respect to $\beta$ equal to zero, we obtain the equations $\alpha'(\beta) + \mathbb{E}_{H_i}\{\psi(X)\} = 0$, $i = 1, 2$. Now, the ordering of $H_i$’s implies that $\mathbb{E}_{H_1}\{\psi(X)\} < \mathbb{E}_{H_2}\{\psi(X)\}$ and so, $\alpha'(\beta_1) > \alpha'(\beta_2)$ which means that $\beta_1 < \beta_2$.

Proof of Theorem 2

Suppose that we fit the model

$$g_{ij}(x; \psi) = f_0(x) \exp\{\tau_{ij} + \omega_{ij}\psi(x)\}, \quad (i, j) \in \mathcal{M}, \ (i, j) \neq (1, 1),$$

(B.4)
where \( \psi \) is a monotone increasing function. Let \( \hat{\Omega} = (\hat{\omega}_{ij}) \) be the matrix containing the maximizers of the likelihood with respect to \( \omega_{ij} \)'s (with \( \hat{\omega}_{11} \equiv 0 \) and \( \hat{\omega} = \text{vec}(\hat{\Omega}) \) be as before the corresponding \((RC - 1)\)-dimensional vector with \( \hat{\omega}_{11} \) dropped. A standard argument based on the derivations of Qin and Zhang (1997) and Claeskens and Hjort (2008) shows that
\[
\sqrt{n}(\hat{\omega} - \omega) \Rightarrow N_{RC-1}(0, \Phi_\omega),
\]
where \( \Phi_\omega \) is some positive definite covariance matrix depending on \( \omega \). Here, \( \omega \) is the vector containing the best approximating parameter values with respect to the KL divergence, i.e., \( \omega_{ij} \) is the value that minimizes the KL divergence between \( g_{ij}(\cdot; \psi) \) and the true density \( g_{ij} \); see also Lemma B.2. It now follows from Lemma B.2 that if \( G_{ij} = G_{i'j'} \) then \( \omega_{ij} = \omega_{i'j'} \) and that if \( G_{ij} \preceq_{lr} G_{i'j'} \) (or \( G_{ij} \succeq_{lr} G_{i'j'} \)) then \( \omega_{ij} < \omega_{i'j'} \) (or \( \omega_{ij} > \omega_{i'j'} \)). This implies that the hypotheses (2) and (8) are equivalent to the corresponding hypotheses in terms of the \( \omega_{ij} \)'s.

It can be shown (cf. Silvapulle and Sen, 2005) that the LRT admits the quadratic approximation
\[
\Lambda_{gm} = n\hat{\omega}^T \Phi^{-1} \hat{\omega} - \min_{\omega \in L_m} \{ n(\hat{\omega} - \omega)^T \Phi^{-1}(\hat{\omega} - \omega) \} + o_P(1)
\]
Under the null hypothesis, the usual theory shows that \( \Lambda_{gm} \Rightarrow \chi^2(\Phi_0, L_m) \), i.e., a chi-bar square distribution (cf. Silvapulle and Sen, 2005). This means that the asymptotic size of the LRT does not depend on \( \psi \). On the other hand, under \( A_m \) we have \( \lim_{n \to \infty} \mathbb{P}(\hat{\omega} \in L_m) = 1 \). Consequently, the second term on the RHS above converges almost surely to zero and therefore \( \Lambda_{gm} \to \infty \) with probability one. This implies that the LRT is consistent even under misspecification. \( \square \)

**Proof of Theorem A.1**

Zhang (2006) proved that for the unrestricted density ratio model, the likelihood ratio test statistic is asymptotically equivalent to the score statistic in the two sample case. His result readily extends to any number of groups. Therefore, in order to prove the theorem, it is sufficient to show that the statement holds for the test based on the score.

Suppose that we fit the model (1) using the function \( \psi \), i.e., model (B.4). Let \( U = \text{vec}(U_{ij}) \), \((i, j) \neq (1, 1)\), be the score function of \( \Omega \) at \( \Omega = 0 \) under the misspecified model that is, \( U_{ij} = \frac{1}{n} \frac{\partial}{\partial \omega_{ij}} \ell(T, \Omega) \big|_{T=\Omega=0} \). It can be easily verified that \( U_{ij} = n_{ij} (Z_{ij} - Z)/n \), where \( Z_{ij} = n^{-1} \sum_{k \in S_{ij}} \psi(Y_k) \) and \( Z \) is the corresponding grand mean. Then, by the central limit theorem, \( n^{1/2} (U - \mu) \Rightarrow N_{RC-1}(0, \Psi) \), where \( \mu = \text{vec}(\mu_{ij}) \), \((i, j) \neq (1, 1)\), with \( \mu_{ij} = \mathbb{E}_{G_{ij}} \{ \psi(X) \} - \sum_{s=1}^{R_i} \sum_{l=1}^{C} \rho_{sl} \mathbb{E}_{G_{is}} \{ \psi(X) \} \), and \( \Psi \) is some positive definite matrix. It follows that the test statistic for \( H_g \) versus \( A_g \) based on the score is \( S = n U^T \hat{\Psi}_0 U \), where \( \hat{\Psi}_0 \) is a consistent estimator of \( \Psi_0 \). Clearly, under \( H_g \), \( S \sim \chi^2_{RC-1} \). However, \( \mu = 0 \) by assumption. This implies that \( S \sim \chi^2_{RC-1} \) under \( A_g \) as well. Hence, the asymptotic power of the test equals its asymptotic size. \( \square \)
C  Additional numerical results, figures and tables

This appendix contains additional material to Sections 5 and 6 of the paper. Subsection C.1 contains Table S.1 which shows the mean square errors (estimated by simulation) of unconstrained and constrained estimators of the parameters $\lambda$ and the corresponding relative efficiency for two underlying families of distributions and Table S.2 where the results of a simulation study when the underlying family of distributions is exponential are presented. Subsection C.2 contains tables and figures related to the real data examples we analyze in the paper.

C.1  Simulations

Table S.1: Componentwise mean square error of unconstrained and constrained estimators for a $3 \times 2$ design in which each cell contains $n_{ij}$ observations drawn from the normal $N(\lambda_{ij}, 1)$ or the beta $B(1 + \lambda_{ij}, 2)$ distributions, for $i = 1, 2, 3, j = 1, 2$. Model (1) is fitted to the data with $\phi(x) = x$ and $\phi(x) = \log x$, respectively. The third and sixth columns are the ratios of the mean square error of the unconstrained estimators to the mean square error of the constrained estimators. Results are based on 1000 simulations.

<table>
<thead>
<tr>
<th>$A$</th>
<th>$n_{ij} = 20$</th>
<th></th>
<th></th>
<th>$n_{ij} = 50$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE($\lambda_{ij}$)</td>
<td>MSE($\lambda_{ij}^*$)</td>
<td>Efficiency</td>
<td>MSE($\lambda_{ij}$)</td>
<td>MSE($\lambda_{ij}^*$)</td>
<td>Efficiency</td>
</tr>
<tr>
<td>$\lambda_{12} = 0$</td>
<td>.105</td>
<td>.041</td>
<td>2.561</td>
<td>.042</td>
<td>.017</td>
<td>2.471</td>
</tr>
<tr>
<td>$\lambda_{21} = 0$</td>
<td>.108</td>
<td>.030</td>
<td>3.600</td>
<td>.043</td>
<td>.012</td>
<td>3.583</td>
</tr>
<tr>
<td>$\lambda_{22} = 0$</td>
<td>.111</td>
<td>.059</td>
<td>1.881</td>
<td>.043</td>
<td>.024</td>
<td>1.792</td>
</tr>
<tr>
<td>$\lambda_{31} = 0$</td>
<td>.119</td>
<td>.063</td>
<td>1.889</td>
<td>.046</td>
<td>.003</td>
<td>1.840</td>
</tr>
<tr>
<td>$\lambda_{32} = 0$</td>
<td>.116</td>
<td>.102</td>
<td>1.137</td>
<td>.043</td>
<td>.039</td>
<td>1.103</td>
</tr>
<tr>
<td>$\lambda_{12} = 1$</td>
<td>.313</td>
<td>.235</td>
<td>1.332</td>
<td>.093</td>
<td>.087</td>
<td>1.069</td>
</tr>
<tr>
<td>$\lambda_{21} = .5$</td>
<td>.192</td>
<td>.153</td>
<td>1.255</td>
<td>.065</td>
<td>.061</td>
<td>1.066</td>
</tr>
<tr>
<td>$\lambda_{22} = 1.5$</td>
<td>.455</td>
<td>.330</td>
<td>1.379</td>
<td>.147</td>
<td>.127</td>
<td>1.157</td>
</tr>
<tr>
<td>$\lambda_{31} = 1$</td>
<td>.316</td>
<td>.289</td>
<td>1.093</td>
<td>.089</td>
<td>.087</td>
<td>1.023</td>
</tr>
<tr>
<td>$\lambda_{32} = 2$</td>
<td>.509</td>
<td>.488</td>
<td>1.043</td>
<td>.179</td>
<td>.173</td>
<td>1.035</td>
</tr>
<tr>
<td>$\lambda_{12} = 0$</td>
<td>.112</td>
<td>.030</td>
<td>3.733</td>
<td>.038</td>
<td>.011</td>
<td>3.455</td>
</tr>
<tr>
<td>$\lambda_{21} = 0$</td>
<td>.113</td>
<td>.023</td>
<td>4.913</td>
<td>.032</td>
<td>.007</td>
<td>4.571</td>
</tr>
<tr>
<td>$\lambda_{22} = 0$</td>
<td>.106</td>
<td>.048</td>
<td>2.208</td>
<td>.035</td>
<td>.018</td>
<td>1.944</td>
</tr>
<tr>
<td>$\lambda_{31} = 0$</td>
<td>.105</td>
<td>.054</td>
<td>1.944</td>
<td>.033</td>
<td>.017</td>
<td>1.941</td>
</tr>
<tr>
<td>$\lambda_{32} = 0$</td>
<td>.115</td>
<td>.100</td>
<td>1.150</td>
<td>.036</td>
<td>.035</td>
<td>1.029</td>
</tr>
<tr>
<td>$\lambda_{12} = 1$</td>
<td>.307</td>
<td>.237</td>
<td>1.295</td>
<td>.105</td>
<td>.095</td>
<td>1.105</td>
</tr>
<tr>
<td>$\lambda_{21} = .5$</td>
<td>.198</td>
<td>.148</td>
<td>1.338</td>
<td>.060</td>
<td>.057</td>
<td>1.053</td>
</tr>
<tr>
<td>$\lambda_{22} = 1.5$</td>
<td>.414</td>
<td>.303</td>
<td>1.366</td>
<td>.151</td>
<td>.129</td>
<td>1.171</td>
</tr>
<tr>
<td>$\lambda_{31} = 1$</td>
<td>.295</td>
<td>.271</td>
<td>1.089</td>
<td>.096</td>
<td>.094</td>
<td>1.021</td>
</tr>
<tr>
<td>$\lambda_{32} = 2$</td>
<td>.610</td>
<td>.581</td>
<td>1.050</td>
<td>.187</td>
<td>.180</td>
<td>1.039</td>
</tr>
</tbody>
</table>
In order to evaluate further the performance of our method under model misspecification we additionally considered the case where the underlying distributions are exponential. Here, the baseline distribution is exponential with scale parameter 1 while the alternative distributions are exponential with scale parameters $1 + (i - 1)/R + (j - 1)/C$, $i = 1, \ldots, R$, $j = 1, \ldots, C$. The results are displayed in Table S.2.

Table S.2: Estimated size and power of $\Lambda_{gm}$-test as functions of the Box-Cox transformation parameter $\zeta$ when the underlying distribution is exponential based on balanced (b) and unbalanced (u) samples. The number of simulated samples is 1000 and the number of random permutations is 500. The correct choice is $\phi(x) = x$ which corresponds to $\zeta = 1$. The last column displays the corresponding results of the permutation test based on the likelihood ratio test statistic.

<table>
<thead>
<tr>
<th>Design</th>
<th>$\zeta = 0$</th>
<th>$\zeta = 1/2$</th>
<th>$\zeta = 1$</th>
<th>$\zeta = 3/2$</th>
<th>$\zeta = 2$</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>$C$</td>
<td>.05</td>
<td>.10</td>
<td>.05</td>
<td>.10</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>Estimated size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2 (b)</td>
<td>.059</td>
<td>.108</td>
<td>.057</td>
<td>.122</td>
<td>.054</td>
<td>.109</td>
</tr>
<tr>
<td>2 2 (u)</td>
<td>.044</td>
<td>.092</td>
<td>.048</td>
<td>.104</td>
<td>.053</td>
<td>.113</td>
</tr>
<tr>
<td>2 3 (b)</td>
<td>.048</td>
<td>.094</td>
<td>.038</td>
<td>.086</td>
<td>.037</td>
<td>.085</td>
</tr>
<tr>
<td>2 3 (u)</td>
<td>.047</td>
<td>.102</td>
<td>.043</td>
<td>.100</td>
<td>.040</td>
<td>.089</td>
</tr>
<tr>
<td>3 3 (b)</td>
<td>.047</td>
<td>.094</td>
<td>.040</td>
<td>.097</td>
<td>.046</td>
<td>.091</td>
</tr>
<tr>
<td>3 3 (u)</td>
<td>.045</td>
<td>.083</td>
<td>.043</td>
<td>.089</td>
<td>.034</td>
<td>.086</td>
</tr>
<tr>
<td>Estimated power</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2 (b)</td>
<td>.196</td>
<td>.319</td>
<td>.248</td>
<td>.403</td>
<td>.262</td>
<td>.404</td>
</tr>
<tr>
<td>2 3 (b)</td>
<td>.336</td>
<td>.474</td>
<td>.424</td>
<td>.567</td>
<td>.443</td>
<td>.588</td>
</tr>
<tr>
<td>2 3 (u)</td>
<td>.396</td>
<td>.542</td>
<td>.480</td>
<td>.635</td>
<td>.503</td>
<td>.667</td>
</tr>
<tr>
<td>3 3 (b)</td>
<td>.482</td>
<td>.623</td>
<td>.587</td>
<td>.740</td>
<td>.621</td>
<td>.781</td>
</tr>
<tr>
<td>3 3 (u)</td>
<td>.573</td>
<td>.702</td>
<td>.693</td>
<td>.823</td>
<td>.733</td>
<td>.838</td>
</tr>
</tbody>
</table>
C.2 Real data examples

C.2.1 Days absent from school

Figure S.2: Boxplots of the number of days absent from the school versus ethnicity and gender.

Table S.3: Estimates of the parameters $\lambda_{ij}$ along with their standard errors and the corresponding $z$-scores ($= \text{estimate/se}$) as well as estimated p-values for testing $H_c$ vs $A_c$ under models (1) and (21), respectively, for several functions $\phi$.

<table>
<thead>
<tr>
<th>$\phi(x)$</th>
<th>$\lambda_{12}$</th>
<th>$\lambda_{21}$</th>
<th>$\lambda_{22}$</th>
<th>$\theta_{22}$</th>
<th>$\Lambda_{cc}$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>.0324</td>
<td>.0559</td>
<td>.0579</td>
<td>-.0304</td>
<td>.122</td>
<td>.130</td>
</tr>
<tr>
<td></td>
<td>(.022)</td>
<td>(.019)</td>
<td>(.019)</td>
<td>(.026)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.473</td>
<td>2.942</td>
<td>3.047</td>
<td>-1.169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-1/x + 1$</td>
<td>-.524</td>
<td>2.487</td>
<td>5.187</td>
<td>3.225</td>
<td>.188</td>
<td>.429</td>
</tr>
<tr>
<td></td>
<td>(1.007)</td>
<td>(1.533)</td>
<td>(1.903)</td>
<td>(2.454)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.520</td>
<td>1.622</td>
<td>2.726</td>
<td>1.314</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sqrt{x + 1}$</td>
<td>.153</td>
<td>.413</td>
<td>.482</td>
<td>-.084</td>
<td>.261</td>
<td>.134</td>
</tr>
<tr>
<td></td>
<td>(.165)</td>
<td>(.141)</td>
<td>(.137)</td>
<td>(.217)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.927</td>
<td>2.929</td>
<td>3.518</td>
<td>-.387</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log($x + 2$)</td>
<td>.117</td>
<td>.712</td>
<td>.975</td>
<td>.146</td>
<td>.311</td>
<td>.168</td>
</tr>
<tr>
<td></td>
<td>(.289)</td>
<td>(.277)</td>
<td>(.277)</td>
<td>(.432)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.405</td>
<td>2.570</td>
<td>3.520</td>
<td>.338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-1/\log(x + 2)$</td>
<td>-.579</td>
<td>1.740</td>
<td>4.687</td>
<td>3.526</td>
<td>.153</td>
<td>.483</td>
</tr>
<tr>
<td></td>
<td>(.714)</td>
<td>(1.453)</td>
<td>(1.995)</td>
<td>(2.407)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.811</td>
<td>1.198</td>
<td>2.349</td>
<td>1.465</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log log($x + 2$)</td>
<td>-.122</td>
<td>1.285</td>
<td>2.221</td>
<td>1.059</td>
<td>.244</td>
<td>.327</td>
</tr>
<tr>
<td></td>
<td>(.517)</td>
<td>(.649)</td>
<td>(.718)</td>
<td>(1.005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.236</td>
<td>1.980</td>
<td>3.093</td>
<td>1.054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sqrt{\log(x + 2)}$</td>
<td>.070</td>
<td>1.972</td>
<td>2.980</td>
<td>0.938</td>
<td>.292</td>
<td>.233</td>
</tr>
<tr>
<td></td>
<td>(.800)</td>
<td>(.854)</td>
<td>(.885)</td>
<td>(1.325)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.088</td>
<td>2.309</td>
<td>3.367</td>
<td>.708</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
C.2.2  Effects of nitrogen and phosphorus on maize crop

Table S.4: Maize crop data.

<table>
<thead>
<tr>
<th>N</th>
<th>0</th>
<th>40</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>40</td>
<td>7.79</td>
<td>6.12</td>
<td>6.30</td>
</tr>
<tr>
<td></td>
<td>5.94</td>
<td>6.05</td>
<td>4.95</td>
</tr>
<tr>
<td></td>
<td>6.35</td>
<td>6.00</td>
<td>6.65</td>
</tr>
<tr>
<td>120</td>
<td>8.44</td>
<td>7.78</td>
<td>7.86</td>
</tr>
<tr>
<td></td>
<td>9.00</td>
<td>8.36</td>
<td>7.15</td>
</tr>
</tbody>
</table>

Figure S.3: Boxplots of the maize crop data.
References


