AISM (2006) 58: 475–497
DOI 10.1007/s10463-005-0022-8

Konstantinos Fokianos · Irene Kaimi

On the effect of misspecifying the density ratio model

Received: 17 February 2004 / Revised: 31 May 2005 / Published online: 17 June 2006

Abstract The density ratio model specifies that the log-likelihood ratio of two unknown densities is of known linear form which depends on some finite dimensional parameters. The model can be broadened to allow for m-samples in a quite natural way. Estimation of both parametric and nonparametric part of the model is carried out by the method of empirical likelihood. However the assumed linear form has an impact on the estimation and testing for the parametric part. The goal of this study is to quantify the effect of choosing an incorrect linear form and its impact to inference. The issue of misspecification is addressed by embedding the unknown linear form to some parametric transformation family which yields ultimately to its identification. Simulated examples and data analysis integrate the presentation.

Keywords Biased sampling · Empirical likelihood · Box–Cox transformation · Mean square error · Bias · Power

1 Introduction

The subject matter of this study is the so called *density ratio model* which is specified by assuming that the log-ratio of two *unknown* probability density functions is linear in some parameters. The model is motivated by considering a binary random variable Y, which assumes two values, say 1 and 2—where "1" denotes success—and X, a p-dimensional vector of covariates, see Cox and Snell (1989) for example.

K. Fokianos (✉)
Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
E-mail: fokianos@ucy.ac.cy

I. Kaimi
Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK

Then the logistic regression model expresses the probability of the event $\{Y = 1\}$ as a function of X by

$$\mathrm{P}[Y = 1 \mid x] = \frac{\exp\left(\alpha_1^{\star} + \boldsymbol{\beta}_1' x\right)}{1 + \exp\left(\alpha_1^{\star} + \boldsymbol{\beta}_1' x\right)}, \tag{1}$$

where $\alpha_1^{\star}$ is a scalar parameter and $\boldsymbol{\beta}_1$ is a $p \times 1$ vector of regression coefficients. Eq. 11 leads to the density ratio model when considering case–control or retrospective sampling, Prentice and Pyke (1979). Suppose that $X_{11}, \ldots, X_{1n_1}$ is a random sample from $F(x \mid Y = 1)$ and $X_{21}, \ldots, X_{2n_2}$ is another independent sample from $F(\mathbf{x} \mid Y = 2)$. Set $\pi = \mathrm{P}[Y = 1]$ and $f(\mathbf{x} \mid Y = i) = dF(\mathbf{x} \mid Y = i)/d\mathbf{x}$, for the conditional probability density function of X given $Y = i$, $i = 1, 2$. Bayes' theorem yields to

$$f(x \mid Y = 1) = \frac{\mathrm{P}[Y = 1 \mid x]\, f_X(x)}{\pi},$$

and

$$f(x \mid Y = 2) = \frac{(1 - \mathrm{P}[Y = 1 \mid x])\, f_X(x)}{1 - \pi},$$

where the marginal density of X, say $f_X(x)$, is left unspecified. The last two equations, when combined with (1), show that

$$\begin{aligned} \frac{f(x \mid Y = 1)}{f(x \mid Y = 2)} &= \frac{1 - \pi}{\pi} \exp\left(\alpha_1^{\star} + \boldsymbol{\beta}_1' x\right) \\ &= \exp\left(\alpha_1 + \boldsymbol{\beta}_1' x\right), \end{aligned}$$

with $\alpha_1 = \alpha_1^{\star} + \log\{(1 - \pi)/\pi\}$. The preceding display justifies the term density ratio model: the densities of the observations are related by a parametric exponential tilt, but otherwise are unknown. Switch notation by setting $g_i(x) \equiv f(x \mid Y = i)$, $i = 1, 2$ to derive the following two independent samples semiparametric problem

$$\begin{array}{ll} X_{11}, \ldots, X_{1n_1} & \text{is a random sample from } g_1(x) = \exp\left(\alpha_1 + \boldsymbol{\beta}_1' \mathbf{h}(\mathbf{x})\right) g_2(x), \\ X_{21}, \ldots, X_{2n_2} & \text{is a random sample from } g_2(x) \end{array} \tag{2}$$

by employing analogous arguments (Qin and Zhang 1997, Qin 1998) and inserting a more general linear form in Eq. 1. In what follows, assume that X is univariate but $\mathbf{h}(.)$ is a p-dimensional vector function which is assumed to be *known* and consists of functions of X—a fact that is of central importance in the rest of the article.

Example 1 The Normal Case: further insight on the density ratio model can be gained by some concrete examples. Assume that $g_1(.)$ and $g_2(.)$ denote densities of normal random variables with unequal means, say μ_1 and μ_2 respectively, but with equal variance σ^2. A straightforward calculation shows that Eq. 2 holds with $\alpha_1 = (\mu_2^2 - \mu_1^2)/2\sigma^2$, $\boldsymbol{\beta}_1 = (\mu_1 - \mu_2)/\sigma^2$ and $\mathbf{h}(x) = x$. Similarly if $g_1(.)$ and

$g_2(.)$ stand for densities of normal random variables with different parameters, say $(\mu_1, \sigma_1^2)'$ and $(\mu_2, \sigma_2^2)'$, then the density ratio model holds again but with

$$\alpha_1 = \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\mu_2^2}{2\sigma_2^2} - \frac{\mu_1^2}{2\sigma_1^2},$$
$$\boldsymbol{\beta}_1 = \left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}, \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2}\right)',$$
$$\mathbf{h}(x) = \left(x, x^2\right)'.$$

An important observation is that when the model holds, and if $\boldsymbol{\beta}_1 = \mathbf{0}$, then the two samples are identically distributed. We conclude that model of Eq. 2 is useful to the semiparametric comparison of two samples in the sense that the densities $g_i(.)$, $i = 1, 2$ are left completely unspecified but the weight function $\exp\left(\alpha_1 + \boldsymbol{\beta}_1'\mathbf{h}(x)\right)$ depends on some finite dimensional parameter. The last remark connects the density ratio model and biased sampling theory, see Vardi (1982, 1985), Gill et al. (1988), Gilbert et al. (1999) and Gilbert (2000). Some other related literature associated with the density ratio model is on testing its goodness of fit. Specifically, Qin and Zhang (1997) propose a bootstrap method, Fokianos et al. (1999) study a generalized moments test statistic and more recently Zhang (2001) proposes an information matrix test for model of Eq. 2. The density ratio model has been applied to both environmental (Fokianos et al. (1998), Kedem et al. (2004)) and biomedical data (Qin et al. (2002)).

Following the same reasoning, it is feasible to generalize Eq. 2 to an m–samples problem. Avoiding unnecessary repetition, consider m unknown densities which are related by an exponential tilt of the following form

$$\begin{aligned} X_{11}, \ldots, X_{1n_1}, \text{random sample from } g_1(x) &= \exp\left(\alpha_1 + \boldsymbol{\beta}_1'\mathbf{h}(x)\right) g_m(x), \\ X_{21}, \ldots, X_{2n_2}, \text{random sample from } g_2(x) &= \exp\left(\alpha_2 + \boldsymbol{\beta}_2'\mathbf{h}(x)\right) g_m(x), \\ \cdots \quad \cdots & \quad \cdots \\ X_{m1}, \ldots, X_{mn_m}, \text{random sample from } g_m(x). & \end{aligned} \tag{3}$$

Estimation of $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{m-1}$ as well as inference regarding the cumulative distribution functions that correspond to $g_1(.), \ldots, g_m(.)$ has been considered by Fokianos et al. (2001) who also propose some test statistics for the hypotheses $\boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_{m-1} = \mathbf{0}$—that is all the samples are identically distributed. In this sense, model of Eq. 3 is also referred as a *semiparametric one way ANOVA*. The density ratio model for two and m samples avoids the normal theory by specifying that the log ratio of two unknown densities is of some parametric form. Hence it provides another way of testing the equality of several distributions without resorting to transformations or any other techniques. The last comment is particular useful since there are examples of data which show that populations follow skewed distributions and therefore classical estimation theory might yield questionable results. The suggested model accommodates skewed data and provides desirable results such as consistent estimators of means, test statistics and so on-see, for example, White and Thompslon (2003) regarding the UK700 clinical trial for more.

A drawback of Eq. 2 and more generally Eq. 3, is their dependence on the assumed known function $\mathbf{h}(.)$. It is anticipated therefore that incorrect choice of

$\mathbf{h}(.)$ affects estimation and testing—this point is demonstrated in Sect. 3 by some simulated examples and theory. The aim of this contribution is to quantify the effect of specifying an incorrect linear form on both estimation and testing and the objective is met by assuming that the true $\mathbf{h}(.)$ belongs to the Box–Cox parametric family of transformations (see Box and Cox 1964). This in turn implies that the parameter λ associated with the well known transformation can be estimated by standard methods—a fact which yields ultimately to the *identification of the function itself from data*. The line of attack is as follows: Sect. 2 reviews some basic facts about estimation for the two—and more general m-sample problem and Sect. 3 addresses the problem of misspecifying $\mathbf{h}(.)$. Sect. 4 discusses estimation of the Box–Cox transforation parameter while Sect. 5 illustrates the methodology to real data. The paper concludes with some comments and discussion.

2 Estimation

Consider first the two samples density ratio model which is easier to follow. Recall Eq. 2 and denote by $\mathbf{x} = (x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{2n_2})'$ the combined sample based on $n = n_1 + n_2$ observations. Consider the following associated inferential problems:

(1) Estimation of the finite dimensional parameters α_1 and $\boldsymbol{\beta}_1$ and
(2) Estimation of the cdf of $g_2(x)$, say $G_2(x)$,

based on the given sample $\mathbf{x}$. Needless to mention that if there exists estimator for all parameters α_1, $\boldsymbol{\beta}_1$ and $G_2(x)$, then it is straightforward to construct an estimator of $G_1(x)$.

Both of these problems are attacked cleverly by the method of empirical likelihood as outlined in Owen (1988) and recently summarized in his monograph, Owen (2001). Accordingly, let p_{ij} denote the size of the jump at the observed datum x_{ij}, that is $p_{ij} = \mathrm{d}G_2(x_{ij}) = G_2(x_{ij}^{+}) - G_2(x_{ij}^{-})$, $j = 1, 2, \ldots, n_i$, $i = 1, 2$ and consider the following empirical likelihood given the data,

$$\begin{aligned} L(\alpha_1, \boldsymbol{\beta}_1, G_2 \mid \mathbf{x}) &= \left\{ \prod_{j=1}^{n_1} \exp\left(\alpha_1 + \boldsymbol{\beta}_1' \mathbf{h}(x_{1j})\right) \mathrm{d}G_2(x_{1j}) \right\} \left\{ \prod_{j=1}^{n_2} \mathrm{d}G_2(x_{2j}) \right\} \\ &= \left\{ \prod_{i=1}^{2} \prod_{j=1}^{n_i} p_{ij} \right\} \left\{ \prod_{j=1}^{n_1} \exp\left(\alpha_1 + \boldsymbol{\beta}_1' \mathbf{h}(x_{1j})\right) \right\}. \end{aligned} \tag{4}$$

Hence the log empirical likelihood equals to

$$l \equiv \log L(\alpha_1, \boldsymbol{\beta}_1, G_2 \mid \mathbf{x}) = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \log p_{ij} + \sum_{j=1}^{n_1} \left(\alpha_1 + \boldsymbol{\beta}_1' \mathbf{h}(x_{1j})\right). \tag{5}$$

Following Qin and Zhang (1997) inference for the finite dimensional parameters α_1 and $\boldsymbol{\beta}_1$ is based on the following parametric likelihood

$$l = -\sum_{i=1}^{2}\sum_{j=1}^{n_i} \log\left[1 + \rho_1 \exp(\alpha_1 + \boldsymbol{\beta}_1' \mathbf{h}(x_{ij}))\right] + \sum_{j=1}^{n_1}\left(\alpha_1 + \boldsymbol{\beta}_1' \mathbf{h}(x_{1j})\right)$$
$$-n \log n_2. \tag{6}$$

where $\rho_1 = n_1/n_2$. Assume that $\hat{\alpha}_1$ and $\hat{\boldsymbol{\beta}}_1$ denote the unique solutions of the score equations

$$\frac{\partial l}{\partial \alpha_1} = -\sum_{i=1}^{2}\sum_{j=1}^{n_i} \frac{\rho_1 \exp(\alpha_1 + \boldsymbol{\beta}_1' \mathbf{h}(x_{ij}))}{1 + \rho_1 \exp(\alpha_1 + \boldsymbol{\beta}_1' \mathbf{h}(x_{ij}))} + n_1 = 0, \tag{7}$$

and

$$\frac{\partial l}{\partial \boldsymbol{\beta}_1} = -\sum_{i=1}^{2}\sum_{j=1}^{n_i} \frac{\rho_1 \exp(\alpha_1 + \boldsymbol{\beta}_1' \mathbf{h}(x_{ij}))\mathbf{h}(x_{ij})}{1 + \rho_1 \exp(\alpha_1 + \boldsymbol{\beta}_1' \mathbf{h}(x_{ij}))} + \sum_{j=1}^{n_1} \mathbf{h}(x_{1j}) = 0. \tag{8}$$

In addition,

$$\hat{p}_{ij} = \frac{1}{n_2} \frac{1}{1 + \rho_1 \exp(\hat{\alpha}_1 + \hat{\boldsymbol{\beta}}_1' \mathbf{h}(x_{ij}))}. \tag{9}$$

Hence the cdf of $g_2(.)$, say $G_2(.)$ is estimated simply by

$$\hat{G}_2(x) = \sum_{i=1}^{2}\sum_{j=1}^{n_i} \hat{p}_{ij} I(X_{ij} \le x)$$

while

$$\hat{G}_1(x) = \sum_{i=1}^{2}\sum_{j=1}^{n_i} \hat{p}_{ij} \exp(\hat{\alpha}_1 + \hat{\boldsymbol{\beta}}_1' \mathbf{h}(x_{ij})) I(X_{ij} \le x),$$

where $I(.)$ denotes the indicator function.

Remark 1 Identifiability of α_1, $\boldsymbol{\beta}_1$ and $G_2(x)$ as well as existence and uniqueness of $\hat{\alpha}_1$, $\hat{\boldsymbol{\beta}}_1$ and $\hat{G}_2(x)$ is guaranteed by the work of Gilbert et al. (1999). In particular, Theorem 2 of this article shows that if there exist a value x_0 such that $\mathbf{h}(x_0) = \mathbf{0}$, then the density ratio model (Eq. 2) is identifiable.

Remark 2 Maximum likelihood estimation for the m-sample problem proceeds along the previous lines (see Fokianos et al. 2001, Sect.2). For clarity and completeness of the presentation, consider the profile log likelihood with respect to $\alpha_1, \ldots, \alpha_{m-1}$ and $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{m-1}$ which is given up to a constant by

$$l = -\sum_{i=1}^{m}\sum_{j=1}^{n_i} \log\left(1 + \sum_{k=1}^{m-1} \rho_k \exp(\alpha_k + \boldsymbol{\beta}_k' \mathbf{h}(x_{ij}))\right)$$
$$+\sum_{i=1}^{m-1}\sum_{j=1}^{n_i}\left(\alpha_i + \boldsymbol{\beta}_i' \mathbf{h}(x_{ij})\right) \tag{10}$$

with $\rho_i = n_i/n_m$.

2.1 Asymptotics

Asymptotic inference for the two sample problem is based on the limiting distribution of $\hat{\boldsymbol{\beta}}_1$. To avoid lengthy formulas and complicated expressions which naturally merge as p increases, we focus on the case $p=1$—that is both $\boldsymbol{\beta}_1$ and $\mathbf{h}(x)$ are univariate. For the rest of the paper assume that $\boldsymbol{\beta}_1 = \beta_1$ and $\mathbf{h}(x) = h(x)$ so that the fact $p=1$ is emphasized. Similar notation applies for the m-sample density ratio model (Eq. 3) with the necessary modifications.

Denote by

$$a_{11}(t) = \int_{-\infty}^{t} \frac{\exp(\alpha_1+\beta_1 h(x))}{1+\rho_1\exp(\alpha_1+\beta_1 h(x))} \mathrm{d}G_2(x), \qquad a_{11} = a_{11}(\infty),$$

$$a_{21}(t) = \int_{-\infty}^{t} \frac{\exp(\alpha_1+\beta_1 h(x))}{1+\rho_1\exp(\alpha_1+\beta_1 h(x))} h(x)\mathrm{d}G_2(x), \qquad a_{21} = a_{21}(\infty),$$

$$a_{22}(t) = \int_{-\infty}^{t} \frac{\exp(\alpha_1+\beta_1 h(x))}{1+\rho_1\exp(\alpha_1+\beta_1 h(x))} h^2(x)\mathrm{d}G_2(x), \qquad a_{22} = a_{22}(\infty),$$

and set

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{21} \\ a_{21} & a_{22} \end{bmatrix}, \quad \boldsymbol{\Sigma}^2 = \frac{1+\rho_1}{\rho_1}\left(\mathbf{A}^{-1} - \begin{bmatrix} 1+\rho_1 & 0 \\ 0 & 0 \end{bmatrix}\right), \tag{11}$$

where the superscript "2" in $\boldsymbol{\Sigma}^2$ emphasizes the fact that this 2×2 matrix is associated with the two sample density ratio model. Then under suitable regularity conditions, the maximum likelihood estimators of α_1 and β_1 tend to a two dimensional normal distribution with covariance matrix $\boldsymbol{\Sigma}^2$, that is

$$\sqrt{n}\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\beta}_1 - \beta_1 \end{pmatrix} \to \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}^2), \tag{12}$$

in distribution, as $n \to \infty$ (Qin and Zhang 1997). Consistent estimators of the asymptotic covariance matrix can be obtained by substituting the maximum likelihood estimators $\hat{\alpha}_1$, $\hat{\beta}_1$ and $\hat{G}_2(x)$ in (Eq. 11).

Remark 3 Assertion (Eq. 12) has been generalized for the m-sample density ratio problem (Eq. 3) but the corresponding asymptotic covariance matrix is more complicated (see Fokianos et al. 2001, Eq. 12). Recall that p has been fixed to unity so that the parameters are univariate and define the vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{m-1})'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{m-1})'$, $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{m-1})'$, $\hat{\boldsymbol{\beta}} = (\beta_1, \ldots, \beta_{m-1})'$ so that result (12) can be restated as follows:

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \end{pmatrix} \to \mathcal{N}_{2(m-1)}(\mathbf{0}, \boldsymbol{\Sigma}^m), \tag{13}$$

in distribution, as $n \to \infty$, where the $\boldsymbol{\Sigma}^m$ is the associated covariance matrix with the m–samples density ratio model.

2.2 Testing

Recalling the two sample density ratio model (Eq. 2) when $p = 1$, interest is focused on testing the hypothesis $\beta_1 = 0$ so that the two samples are identically distributed. Two procedures have been proposed for carrying out the task and they are described below. The first is based on the large sample properties of the maximum likelihood estimator of β_1. More specifically consider the following test procedure

$$Z = \frac{\hat{\beta}_1}{\sqrt{\widehat{\mathrm{Var}(\hat{\beta}_1)}}} \tag{14}$$

where $\widehat{\mathrm{Var}(\hat{\beta}_1)}$ denotes the estimated asymptotic variance of $\hat{\beta}_1$ which is obtained by result (Eq. 12). Then, reject the hypothesis $\beta_1 = 0$ when $\mid Z \mid > c^\star$ where the critical value $c^\star$ is determined by the standard normal distribution.

An alternative procedure is based on the following quantity, Fokianos et al. (2001),

$$\mathcal{X}_1^2 = \frac{n\mathrm{Var}[h(x)]\rho_1}{(1+\rho_1)^2}\hat{\beta}_1^2 \tag{15}$$

which is asymptotically distributed as a chi–square random variable with 1 degree of freedom under the hypothesis. Expression $\mathrm{Var}[h(x)]$ is computed under the reference sample and it can be estimated consistently by

$$\widehat{\mathrm{Var}[h(x)]} = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\hat{p}_{ij}h^2(x_{ij}) - \left(\sum_{i=1}^{2}\sum_{j=1}^{n_i}\hat{p}_{ij}h(x_{ij})\right)^2,$$

where $\hat{p}_{ij}$, $j = 1, 2, \ldots, n_i$, $i = 1, 2$ are given by (9) and $\rho_1 = n_1/n_2$.

Turning to the m-sample problem, consider the hypotheses $\beta_1 = \cdots = \beta_{m-1} = 0$ which are of central importance since they imply that all the samples are identically distributed. To develop a testing procedure it is necessary to modify both Eqs.(14) and (15) so that they conform with the multivariate aspects of the problem. Recall that $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_{m-1})'$ and suppose that the asymptotic covariance matrix of Eq. (13) is partitioned as follows:

$$\boldsymbol{\Sigma}^m = \begin{bmatrix} \boldsymbol{\Sigma}^m_{\boldsymbol{\alpha\alpha}} & \boldsymbol{\Sigma}^m_{\boldsymbol{\alpha\beta}} \\ \boldsymbol{\Sigma}^{m\prime}_{\boldsymbol{\alpha\beta}} & \boldsymbol{\Sigma}^m_{\boldsymbol{\beta\beta}} \end{bmatrix}$$

Then, standard arguments show that the asymptotic distribution of

$$\chi^2_{m;1} = n\hat{\boldsymbol{\beta}}'\left(\hat{\boldsymbol{\Sigma}}^m_{\boldsymbol{\beta\beta}}\right)^{-1}\hat{\boldsymbol{\beta}} \tag{16}$$

approximates a chi square random variable with $m-1$ degrees of freedom. Hence the hypotheses $\beta_1 = \cdots = \beta_{m-1} = 0$ is rejected for large values of $\chi^2_{m;1}$.

In a similar manner, test statistic (Eq. 15) takes on the following form (see Fokianos et al. 2001, Eq. 16)

$$\chi^2_{m,2} = n\mathrm{Var}\left[h(x)\right]\hat{\boldsymbol{\beta}}'\mathbf{A}_{11}\hat{\boldsymbol{\beta}} \tag{17}$$

where $\mathbf{A}_{11}$ is the $(m-1) \times (m-1)$ matrix whose jth diagonal element is

$$\frac{\rho_j[1+\sum_{k \neq j}^{m-1} \rho_k]}{[1+\sum_{k=1}^{m-1} \rho_k]^2}$$

and otherwise for $j \neq j'$, the jj' element is

$$\frac{-\rho_j \rho_j'}{[1+\sum_{k=1}^{m-1} \rho_k]^2},$$

upon recalling that $\rho_i = n_i/n_m, i = 1, 2, \ldots, m-1$. As it turns out, the limiting distribution of $\chi^2_{m,2}$ is again a chi square random variable with $m-1$ degrees of freedom. Hence both test statistics (Eq. 16) and (Eq. 17) have the same rejection area. The performance of all these tests will be studied next under the assumption of an incorrectly specified model. Definitely both Eqs.(16) and (17) reduces to Eqs.(14) and (15) respectively when $m = 2$ but it is rather interesting to examine each case separately.

Summarizing the above discussion, we conclude that the density ratio model–for two or more samples—is amenable to standard estimation techniques by means of empirical likelihood methodology. Furthermore, the problem of examining whether or not the samples are identically distributed reduces to a parametric hypothesis which can be tested by the aforementioned techniques.

3 Effect of misspecified linear form

A major drawback of the density ratio model is that it depends on the known function $h(.)$—recall that p has been fixed to 1. Hence all the inferential output relies on the choice of $h(.)$ and therefore departures from the true underlying model have an impact on the statistical analysis. To study the effect of incorrect choice of $h(.)$ it is useful to consider a parametric family, say $h_\lambda(x)$, $\lambda \in R$ so that the true linear form belongs to this class. The Box–Cox family of transformations

$$h(x,\lambda) = \begin{cases} \frac{|x|^\lambda sgn(x)-1}{\lambda} & \text{when } \lambda \neq 0 \\ \log|x| & \text{when } \lambda = 0, \end{cases} \tag{18}$$

where $sgn(x)$ denotes the sign function, provides a sound methodological framework of quantifying the effect of using the incorrect $h(.)$ and estimating the function itself from data. The last remark is especially useful since it suggests a method of inference for biased sampling models when the weights are completely unknown. Notice that if the data are positive, then (Eq. 18) reduces to the pioneering transformation of Box and Cox (1964). However, if the data assume real values, then (Eq. 18) is an appropriate choice for the analysis of such observations—see Example 1, for instance. An alternative analysis for real valued data would be based on shifting all the values to the right by adding the minimum to each observation. Then the data are positive and analysis can proceed along the lines of Box and Cox (1964).

The following examples demonstrate effectively the impact of $h(.)$ on estimation and testing by means of Eq. (18).

Example 2 Lognormal distribution: consider two independent samples from lognormal distributions, that is

$$g(x;\mu,\sigma^2)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}(\log x-\mu)^2\right),\quad x>0,$$

with $\mu_1\neq\mu_2$ but $\sigma_1^2=\sigma_2^2=\sigma^2$. It is easy to see that

$$\log\left[\frac{g(x;\mu_1,\sigma^2)}{g(x;\mu_2,\sigma^2)}\right]=\frac{\mu_2^2-\mu_1^2}{2\sigma^2}+\frac{\mu_1-\mu_2}{\sigma^2}\log x.$$

Hence, the density ratio model (Eq. 2) is rediscovered with $\alpha_1=(\mu_2^2-\mu_1^2)/2\sigma^2$, $p=1$, $\beta_1=(\mu_1-\mu_2)/\sigma^2$ and $h(x)=\log x$. A direct comparison of the above results with that of the Gaussian case—see Example 1— reveals that the parameters α_1 and β_1 have the same functional form but the function $h(.)$ differs. Moreover, it is straightforward to verify that when $\sigma_1^2\neq\sigma_2^2$ the density ratio model is true with the same parameters as those of Example 1 but with $\mathbf{h}(x)=(\log x,(\log x)^2)'$—that is $p=2$.

Table 1 reports empirical results regarding the effect of misspecification when the data follow lognormal distributions with $\mu_1\neq\mu_2$ but $\sigma_1^2=\sigma_2^2$, for different sample sizes. In particular, it reports the value of the estimate $\hat{\beta}_1$, its bias and mean square error and the simulated power of both Eq. (14) and (15). All the output is based on 1,000 simulations and the computations were carried out using the `R` system. Transformation (Eq. 18) was used to assess the consequences of choosing the wrong $h(.)$ with the parameter value λ varying between -1 and 1, by 0.2. The conclusion drawn by examining the bias and mean square error results of Table 1 is that an incorrect specification of $h(.)$ leads to biased estimators with large deviations—a uniform result for all sample sizes. In regard to the power of the Z test (Eq. 14) notice that values near $\lambda=0$—the true $h(.)$—lead to optimal performance. In contrast, the chi-square test (Eq. 15) possesses this property for large sample sizes. In addition, its power is superior than the corresponding power of the Z test when λ assumes negative values and it is approximately similar around the true model.

In the same vein, consider three lognormal samples with different values of the μ parameter, say μ_1, μ_2 and μ_3 but the same σ^2. Similar to two sample problem, we obtain a three sample problem with $\alpha_1=(\mu_3^2-\mu_1^2)/2\sigma^2, \alpha_2=(\mu_3^2-\mu_2^2)/2\sigma^2$, $\beta_1=(\mu_1-\mu_3)/\sigma^2$, $\beta_2=(\mu_2-\mu_3)/\sigma^2$, $p=1$ and $h(x)=\log x$ so that Eq. 3 is satisfied. Table 2 illustrates the same information with Table 1 augmented with two columns which correspond to the estimate of $\hat{\beta}_2$ and its bias. Notice that the bias is of the same sign as in the case of two samples while the mean square error takes on its minimum value at the true $\lambda=0$. A comparison between the power of (16) and Eqs. (17) shows that the former assumes larger values for negative λ while the latter takes on its maximum value around 0 for large sample sizes–a conclusion along the lines of the two sample problem.

Example 3 Gamma distribution:

Consider two independent samples from Gamma distributions, i.e.

$$g(x;\mu,\nu)=\frac{1}{\nu^{\mu}\Gamma(\mu)}x^{\mu-1}\exp(-x/\nu),\quad x>0,$$

Table 1 Effect of misspecifying $h(.)$ when the data follow the lognormal distribution with $\mu_1 = 1$, $\mu_2 = 0.5$ and $\sigma^2 = 1$ using transformation (Eq. 18)

Sample Size	λ	$\hat{\beta}_1$	Bias	MSE	Power of Z	Power of $\mathcal{X}_1^2$
	−1.0	0.583	0.083	6.979	0.283	0.596
	−0.8	0.625	0.125	15.633	0.370	0.599
	−0.6	0.646	0.146	21.521	0.413	0.604
	−0.4	0.638	0.138	19.075	0.495	0.607
$n_1 = 40$	−0.2	0.608	0.108	11.672	0.504	0.596
	0.0	0.526	0.026	0.683	0.507	0.545
$n_2 = 30$	0.2	0.448	−0.052	2.645	0.487	0.492
	0.4	0.354	−0.146	21.130	0.445	0.412
	0.6	0.291	−0.209	43.375	0.440	0.370
	0.8	0.209	−0.291	84.311	0.355	0.266
	1.0	0.152	−0.348	120.414	0.282	0.203
	−1.0	0.600	0.100	10.041	0.459	0.711
	−0.8	0.617	0.117	13.727	0.505	0.726
	−0.6	0.633	0.133	17.723	0.590	0.743
	−0.4	0.624	0.124	15.425	0.647	0.757
$n_1 = 50$	−0.2	0.584	0.084	7.073	0.667	0.731
	0.0	0.522	0.022	0.520	0.693	0.710
$n_2 = 50$	0.2	0.441	−0.059	3.379	0.680	0.667
	0.4	0.350	−0.150	22.329	0.649	0.602
	0.6	0.253	−0.247	60.925	0.567	0.494
	0.8	0.181	−0.318	101.590	0.511	0.379
	1.0	0.134	−0.366	133.315	0.460	0.327
	−1.0	0.552	0.052	2.786	0.799	0.911
	−0.8	0.589	0.089	8.061	0.836	0.913
	−0.6	0.611	0.111	12.380	0.901	0.944
	−0.4	0.598	0.098	9.758	0.920	0.953
$n_1 = 100$	−0.2	0.574	0.074	5.565	0.939	0.951
	0.0	0.514	0.014	0.213	0.949	0.953
$n_2 = 100$	0.2	0.426	−0.074	5.345	0.936	0.932
	0.4	0.339	−0.161	25.766	0.914	0.898
	0.6	0.291	−0.209	43.375	0.889	0.848
	0.8	0.173	−0.327	106.850	0.824	0.741
	1.0	0.118	−0.382	145.171	0.760	0.658

The true $h(.)$ corresponds to $\lambda = 0$ and the results are based on 1,000 simulations.

with $\mu_1 \neq \mu_2$ but $\nu_1 = \nu_2 = \nu$. Then

$$\log\left[\frac{g(x;\mu_1,\nu)}{g(x;\mu_2,\nu)}\right] = \log\frac{\Gamma(\mu_2)}{\Gamma(\mu_1)} + (\mu_2 - \mu_1)\log\nu + (\mu_1 - \mu_2)\log x.$$

Hence, Eq. 2 holds again with $p = 1$, $\alpha_1 = \log(\Gamma(\mu_2)/\Gamma(\mu_1)) + (\mu_2 - \mu_1)\log\nu$, $\beta_1 = (\mu_1 - \mu_2)$ and $h(x) = \log x$.

Table 3 reports the same results as those of Table 1 but only for moderate samples. Similar conclusions can be drawn again–the estimates are biased with increasing mean square error as we move away from the true $h(.)$. Furthermore the power of test statistic (Eq. 15) is superior to that of (14), for $\lambda<0$. Notice that for this example, $\lambda>0$ shows that $\hat{\beta}_1$ has positive bias as opposed to the lognormal case—Example 2—where the estimate had negative bias for the same range of λ-values.

Table 2 Effect of misspecifying $h(.)$ when the data follow the lognormal distribution with $\mu_1 = 1$, $\mu_2 = 1.5$, $\mu_3 = 0.5$ and $\sigma^2 = 1$ using transformation (Eq. 18)

Sample size	λ	$\hat{\beta}_1$	Bias of $\hat{\beta}_1$	$\hat{\beta}_2$	Bias of $\hat{\beta}_2$	MSE	Power of $\mathcal{X}^2_{m;1}$	Power of $\mathcal{X}^2_{m;2}$
	−1.0	0.691	0.191	2.103	1.103	1254.843	0.416	0.895
	−0.8	0.662	0.162	1.926	0.926	885.645	0.513	0.902
	−0.6	0.689	0.189	1.672	0.672	487.719	0.588	0.888
	−0.4	0.608	0.108	1.521	0.521	283.969	0.683	0.888
$n_1 = 20$	−0.2	0.623	0.123	1.332	0.332	125.454	0.746	0.882
	0.0	0.538	0.038	1.095	0.095	10.528	0.754	0.819
$n_2 = 20$	0.2	0.468	−0.032	0.896	−0.104	11.773	0.749	0.760
	0.4	0.395	−0.105	0.690	−0.310	106.620	0.685	0.633
$n_3 = 20$	0.6	0.333	−0.167	0.540	−0.460	239.284	0.629	0.501
	0.8	0.244	−0.256	0.385	−0.615	443.209	0.629	0.501
	1.0	0.201	−0.299	0.293	−0.707	587.656	0.456	0.217
	−1.0	0.533	0.033	1.645	0.645	418.002	0.818	0.975
	−0.8	0.573	0.073	1.614	0.614	383.181	0.893	0.974
	−0.6	0.581	0.081	1.540	0.540	298.664	0.936	0.986
	−0.4	0.594	0.094	1.403	0.403	171.523	0.960	0.984
$n_1 = 40$	−0.2	0.564	0.064	1.237	0.237	60.402	0.981	0.988
	0.0	0.512	0.012	1.032	0.032	1.214	0.978	0.982
$n_2 = 50$	0.2	0.467	−0.033	0.851	−0.149	23.061	0.976	0.976
	0.4	0.390	−0.110	0.676	−0.324	116.716	0.959	0.937
$n_3 = 30$	0.6	0.314	−0.186	0.505	−0.495	279.437	0.931	0.854
	0.8	0.252	−0.248	0.376	−0.624	449.765	0.888	0.745
	1.0	0.182	−0.318	0.265	−0.735	640.882	0.827	0.602

The true $h(.)$ corresponds to $\lambda = 0$ and the results are based on 1,000 simulations.

Example 4 Example 1 continued: Consider the Gaussian case which was discussed earlier in Example 1. Table 4 reports the empirical output of 1,000 simulations when the means of both normals are $\mu_1 = 1$ and $\mu_2 = 2$ respectively, while the standard deviation has been set to $\sigma = 2$ for both distributions. In this case, the parameter λ of (Eq. 18) varies from 0 to 2 by 0.2 and its true value is equal to 1. In principle, the results are in agreement with the previous analysis but there are some features that deserve further attention. First notice that in most of the cases the bias of the estimate of β_1 is positive and the mean square error is minimized near values of the true λ. Test (Eq. 14) seems more powerful than the chi-square test (Eq. 15) when λ assumes values less than 1. Both of tests appear to have less power when compared with Tables 1 and 3—this is a consequence of the fact that the variance is rather substantial.

Table 5 reports results for three normally distributed samples where the means are different but the variances are all equal—in fact $\sigma^2 = 1$ for this case. The true values of the parameters can be easily calculated—see Example 1 for more details. As the simulated results illustrate, values far from $\lambda = 1$ yield to biased estimators while the behavior of both test statistics is similar to the results obtained from the case of two normal samples.

Before proceeding on calculating the bias of the maximum empirical likelihood estimate under the incorrect model it is worth pointing out that transformation should be used cautiously since there might exist values of λ that do not introduce a proper probability distribution so that $\sum_{i,j} \hat{p}_{ij} = 1$ according to Eq. (9). For further discussion on existence issues, see Owen (2001). Furthermore, according

Table 3 Effect of misspecifying $h(.)$ when the data follow the Gamma distribution with $\mu_1 = 1$, $\mu_2 = 0.5$ and $\nu = 1$ using transformation (18).

Sample size	λ	$\hat{\beta}_1$	Bias	MSE	Power of Z	Power of $\mathcal{X}_1^2$
	−1.0	0.048	−0.452	203.678	0.274	0.920
	−0.8	0.081	−0.419	174.982	0.408	0.915
	−0.6	0.144	−0.355	126.472	0.565	0.938
	−0.4	0.228	−0.272	73.502	0.687	0.948
$n_1 = 40$	−0.2	0.371	−0.129	16.439	0.863	0.967
	0.0	0.533	0.033	1.089	0.902	0.947
$n_2 = 30$	0.2	0.710	0.210	44.132	0.902	0.910
	0.4	0.841	0.341	116.942	0.861	0.848
	0.6	0.936	0.436	190.264	0.812	0.758
	0.8	0.954	0.454	206.399	0.704	0.626
	1.0	0.967	0.467	218.111	0.638	0.516
	−1.0	0.040	−0.460	211.468	0.357	0.972
	−0.8	0.074	−0.425	181.129	0.519	0.974
	−0.6	0.134	−0.366	133.949	0.698	0.971
	−0.4	0.228	−0.272	73.836	0.833	0.977
$n_1 = 50$	−0.2	0.369	−0.131	16.932	0.935	0.987
	0.0	0.513	0.013	0.186	0.969	0.988
$n_2 = 50$	0.2	0.687	0.187	35.297	0.977	0.983
	0.4	0.805	0.305	93.252	0.961	0.958
	0.6	0.890	0.390	152.212	0.927	0.908
	0.8	0.917	0.417	174.646	0.882	0.828
	1.0	0.881	0.381	145.318	0.814	0.721

The true $h(.)$ corresponds to $\lambda = 0$ and the results are based on 1,000 simulations.

Table 4 Effect of misspecifying $h(.)$ when the data follow Gaussian distribution with $\mu_1 = 1$, $\mu_2 = 2$ and $\sigma = 2$ using transformation (Eq. 18).

Sample size	λ	$\hat{\beta}_1$	Bias	MSE	Power of Z	Power of $\mathcal{X}_1^2$
	0.0	−0.313	−0.063	4.083	0.171	0.145
	0.2	−0.103	0.147	21.490	0.372	0.241
	0.4	−0.202	0.048	2.303	0.451	0.372
	0.6	−0.252	−0.002	0.006	0.490	0.450
$n_1 = 40$	0.8	−0.270	−0.020	0.413	0.496	0.499
	1.0	−0.270	−0.020	0.417	0.526	0.573
$n_2 = 30$	1.2	−0.237	0.013	0.162	0.487	0.570
	1.4	−0.209	0.041	1.639	0.483	0.590
	1.6	−0.178	0.072	5.117	0.457	0.602
	1.8	−0.146	0.104	10.738	0.445	0.615
	2.0	−0.120	0.230	16.763	0.413	0.595
	0.0	−0.296	−0.046	2.167	0.279	0.228
	0.2	−0.098	0.152	23.042	0.520	0.413
	0.4	−0.187	0.063	3.911	0.603	0.529
	0.6	−0.240	0.010	0.097	0.647	0.607
$n_1 = 50$	0.8	−0.268	−0.018	0.324	0.708	0.702
	1.0	−0.254	−0.004	0.021	0.658	0.687
$n_2 = 50$	1.2	−0.239	0.011	0.113	0.687	0.728
	1.4	−0.204	0.045	2.075	0.653	0.724
	1.6	−0.180	0.070	4.881	0.640	0.757
	1.8	−0.142	0.108	11.477	0.591	0.731
	2.0	−0.118	0.132	17.447	0.567	0.733

The true $h(.)$ corresponds to $\lambda = 1$ and the results are based on 1,000 simulations.

Table 5 Effect of misspecifying $h(.)$ when the data follow the Gaussian distribution with $\mu_1=1$, $\mu_2=1$, $\mu_3=0.5$ and $\sigma^2=1$ using transformation (Eq. 18).

Sample size	λ	$\hat{\beta}_1$	Bias of $\hat{\beta}_1$	$\hat{\beta}_2$	Bias of $\hat{\beta}_2$	MSE	Power of $\mathcal{X}^2_{m;1}$	Power of $\mathcal{X}^2_{m;2}$
	0.0	0.259	−0.241	0.266	−0.234	112.120	0.161	0.325
	0.2	0.108	−0.392	0.106	−0.394	308.163	0.383	0.538
	0.4	0.238	−0.262	0.238	−0.261	136.410	0.451	0.585
	0.6	0.364	−0.136	0.358	−0.142	38.341	0.506	0.626
	0.8	0.449	−0.051	0.449	−0.051	5.090	0.504	0.569
$n_1=40$	1.0	0.540	0.040	0.533	0.033	2.714	0.526	0.585
	1.2	0.566	0.066	0.565	0.065	8.652	0.489	0.511
$n_2=50$	1.4	0.564	0.064	0.569	0.069	9.077	0.440	0.431
	1.6	0.566	0.066	0.567	0.067	8.958	0.436	0.414
$n_3=30$	1.8	0.551	0.051	0.544	0.044	4.618	0.383	0.348
	2.0	0.518	0.018	0.524	0.024	0.980	0.339	0.283
	0.0	0.280	−0.220	0.275	−0.225	98.846	0.244	0.424
	0.2	0.107	−0.393	0.111	−0.389	305.092	0.552	0.690
	0.4	0.242	−0.258	0.236	−0.264	135.924	0.631	0.741
	0.6	0.364	−0.136	0.365	−0.135	36.562	0.697	0.786
	0.8	0.456	−0.044	0.464	−0.036	3.208	0.715	0.771
$n_1=50$	1.0	0.496	−0.004	0.500	0.000	0.011	0.655	0.690
	1.2	0.548	0.048	0.547	0.047	4.630	0.689	0.704
$n_2=50$	1.4	0.550	0.050	0.543	0.043	4.488	0.653	0.645
	1.6	0.525	0.025	0.535	0.035	1.937	0.628	0.579
$n_3=50$	1.8	0.505	0.005	0.508	0.008	0.099	0.585	0.519
	2.0	0.484	−0.016	0.472	−0.028	1.009	0.555	0.467

The true $h(.)$ corresponds to $\lambda=1$ and the results are based on 1,000 simulations.

to Remark 1, the density ratio model when using transformation the (Eq. 18) is identifiable since $h_\lambda(1)=0$, independently of the choice of λ, provided that $\beta\neq 0$.

3.1 Asymptotic bias

The empirical results show that a substantial amount of bias is introduced by misspecifying the function $h(.)$ when the latter belongs to the Box–Cox family of transformations. The following result shows that there exists a simple formula for assessing the asymptotic bias in a small neighborhood of the true model. Its proof is postponed to the Appendix.

Proposition 3.1 *Suppose that the two samples density ratio model (2) holds for $p=1$, $\beta_1\neq 0$ and λ_0 denotes the true linear form which belongs to the parametric family (Eq. 18). In addition assume the regularity conditions of Qin & Lawless (1994, Theorem 1) and suppose further that $E[|\ h_\lambda(X,\lambda)\ |\ (1+\rho_1\pi(X,\lambda_0))]$, $E[|\ h_\lambda(X,\lambda)h(X,\lambda)\ |\ (1+\rho_1\pi(X,\lambda_0))]$ and $E[|\ h_\lambda(X,\lambda)h^2(X,\lambda)\ |\ (1+\rho_1\pi(X,\lambda_0))]$ are all finite when expectation is taken with respect to $G_2(x)$ in a $n^{-1/2}$ neighborhood of the true value λ_0. If $\hat{\boldsymbol{\theta}}_{1;\lambda}=(\hat{\alpha}_{1;\lambda},\hat{\beta}_{1;\lambda})'$ are the maximum likelihood estimators of the parameters $\boldsymbol{\theta}=(\alpha_1,\beta_1)'$ which are calculated under a misspecified model for λ in the same neighborhood of the true value λ_0, then*

$$\sqrt{n}\left(\boldsymbol{\Sigma}^2_\lambda\right)^{-1/2}\left(\hat{\boldsymbol{\theta}}_{1;\lambda}-\boldsymbol{\theta}_1-\mathbf{b}_{1n}\right)\to\mathcal{N}_2(\mathbf{0},\mathbf{I}),$$

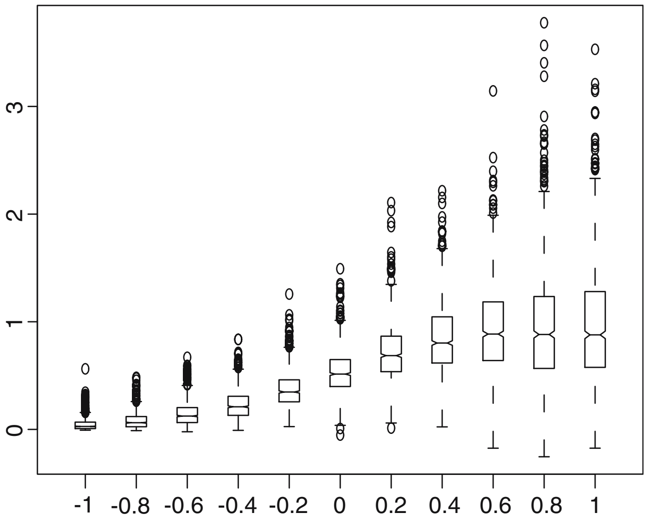

Fig. 1 Boxplots of 1,000 maximum likelihood estimators for the two sample density ratio model when the data follow the Gamma distribution with $\mu_1 = 1$, $\mu_2 = 0.5$ and $\nu = 1$ using transformation (Eq. 18) for different values of λ (*horizontal axis*). The sample sizes equal to $n_1 = 40$, $n_2 = 30$ and λ varies from -1 to 1 with step equal to 0.2

in distribution, as n_1, $n_2 \to \infty$ such that $n_1/n_2 \to \rho_1$. Here $\mathbf{I}$ is the two dimensional identity matrix, the 2×1 vector $\mathbf{b}_{1n}$ is given by

$$\mathbf{b}_{1n} = -(\lambda - \lambda_0)\beta_1 \mathbf{A}_\lambda^{-1} E[h_\lambda(x, \lambda_0) Y(1, h(x, \lambda_0))']$$

where Y, $h_\lambda(X, \lambda)$ and $\pi(X, \lambda_0)$ are defined by (Eq. 22) and the 2×2 matrix $\mathbf{A}_\lambda$ is defined by (Eq. 24). In addition

$$\mathbf{\Sigma}_\lambda^2 = \mathbf{A}_\lambda^{-1} \mathbf{V}_\lambda (\mathbf{A}_\lambda^{-1})'$$

where $\mathbf{V}_\lambda$ is defined by Eq. 27.

Proposition 3.1 shows that the bias can be represented in terms of the two dimensional vector $\mathbf{b}_{1n}$ and its sign has somehow been confirmed empirically at least by the limited simulation results. Indeed Table 1 shows that for $\lambda_0 = 0$ and $\beta_1 > 0$ the bias is positive if $\lambda < 0$ and negative otherwise. Similarly, Table 3 illustrates the opposite result since for this case $\beta_1 < 0$. The other terms appearing in the definition of $\mathbf{b}_{1n}$ certainly influence the final form of the bias but there seems to be positive at least for all simulations considered. Furthermore, Proposition 3.1 facilitates the calculation of the asymptotic distribution of the maximum likelihood estimators $\hat{\alpha}_{1;\lambda}$, $\hat{\beta}_{1;\lambda}$ calculated under the misspecified model and shows the asymptotic normality of the estimates even if the model is misspecified. However for values in the neighborhood of the true λ the approximation is more accurate compared to the values that fall far from λ_0. The point is illustrated in Fig. 1 where boxplots of simulated $\hat{\beta}_{1;\lambda}$ are shown for relative small sample sizes ($n_1 = 40$ and $n_2 = 30$) of Gamma random variables according to Example 3 and Table 3. Notice that the variance increases as λ tends to positive values and the approximation is more accurate for positive values of λ-especially near $\lambda = 0$.

Proposition 3.1 can be extended along the previous lines to the m-sample density ratio model (Eq. 3). Indeed, following the same reasoning of the appendix and recalling the notation of (Eq. 13), we obtain the following generalized version:

Proposition 3.2 *Suppose that the m samples density ratio model (3) holds for $p = 1$, $\beta_i \neq 0$, $i = 1, 2, \ldots, m-1$ and λ_0 denotes the true linear form which belongs to the parametric family (Eq. 18). In addition assume the regularity conditions of Qin & Lawless (1994, Theorem 1) and suppose that $E[| h_\lambda(X, \lambda) | \sum_{i=1}^{m-1}(1 + \rho_i \pi_i(X, \lambda_0))]$, $E[| h_\lambda(X, \lambda) h(X, \lambda) | \sum_{i=1}^{m-1}(1 + \rho_i \pi_i(X, \lambda_0))]$ and $E[| h_\lambda(X, \lambda) h^2(X, \lambda) | \sum_{i=1}^{m-1}(1 + \rho_i \pi_i(X, \lambda_0))]$ are all finite when expectation is taken with respect to $G_m(x)$ in a $n^{-1/2}$ neighborhood of the true value λ_0. If $\hat{\boldsymbol{\theta}}_{1;\lambda} = (\hat{\alpha}_{1;\lambda}, \ldots, \hat{\alpha}_{m-1,\lambda}, \hat{\beta}_{1;\lambda}, \ldots, \hat{\beta}_{m;\lambda})'$ are the maximum likelihood estimators of the parameters $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_{m-1}, \beta_1, \ldots, \beta_{m-1})'$ which are calculated under a misspecified model for λ in the same neighborhood of the true value λ_0, then*

$$\sqrt{n}\left(\boldsymbol{\Sigma}^2_{\lambda;m}\right)^{-1/2}\left(\hat{\boldsymbol{\theta}}_{1;\lambda} - \boldsymbol{\theta}_1 - \mathbf{b}_{mn}\right) \rightarrow \mathcal{N}_{2(m-1)}(\mathbf{0}, \mathbf{I}_{2(m-1)}),$$

in distribution, as for every $i = 1, 2, \ldots, m$, $n_i \rightarrow \infty$ with $n_i/n_m \rightarrow \rho_i$. Here $\mathbf{I}_{2(m-1)}$ is the $2(m-1)$ dimensional identity matrix and the $2(m-1)\times 1$ vector $\mathbf{b}_{mn}$ is given by

$$\begin{aligned}\mathbf{b}_{mn} = & -(\lambda - \lambda_0)\mathbf{A}^{-1}_{\lambda,m}\\ & \times \left\{\sum_{i=1}^{m-1} \beta_i E[h_\lambda(X, \lambda_0) Y_i \mathbf{h}_i(X, \lambda_0)(1 + \sum_{j\neq i}(1 + \rho_j \pi_j(X, \lambda_0)))]\right\}\end{aligned}$$

where all the quantities are defined by (30) and the $2(m-1) \times 2(m-1)$ matrix $\mathbf{A}_{\lambda;m}$ is defined by (31). In addition

$$\boldsymbol{\Sigma}^2_\lambda = \mathbf{A}^{-1}_{\lambda;m}\mathbf{V}_{\lambda;m}(\mathbf{A}^{-1}_{\lambda;m})'$$

where $\mathbf{V}_{\lambda;m}$ is defined by Eq. (32).

To conclude the section, consider the following resume of the above analysis:

- Misspecification of $h(.)$ in Eq. (2) and more generally (Eq. 3) leads to biased estimators with inflated variance.
- Test statistic (Eq. 14) (respectively (Eq. 16)) is preferable to test statistic (Eq. 15) (respectively (Eq. 17)) when $\lambda > \lambda_0$.
- Misspecification of $h(.)$ in (Eq. 2) and more generally (Eq. 3) does not appear to reduce the power of all test statistics considered especially for large sample sizes and when λ assumes values within a neighborhood of the true value.

4 Estimation of λ

The previous results show that the choice of function $h(.)$ is of considerable importance in both estimation and testing. Therefore it is quite natural to investigate the problem of estimating itself from data—see Fokianos (2003). Consider again the two sample density ratio model (Eq. 2) with $p = 1$, $\beta_1 \neq 0$ and let us examine the problem of estimating the parameter λ in (Eq. 18). A standard profiling procedure can be used where different trial values of the parameter λ are considered, say $\lambda_1, \ldots, \lambda_m$, together with the associated log likelihood (Eq. 6), say l_λ. Then the estimator of λ, say $\hat{\lambda}$ is that value where the maximum log likelihood occurs among $l_{\lambda_1}, \ldots, l_{\lambda_m}$. Alternatively a graphical display of $\lambda_1, \ldots, \lambda_m$ versus $l_{\lambda_1}, \ldots, l_{\lambda_m}$ can be used so that $\hat{\lambda}$ can be located visually. An $(1-\alpha)\%$ confidence interval for λ is given by

$$\left\{\lambda : \quad l_{\hat{\lambda}} - l_\lambda < \frac{1}{2}\chi^2_{1,\alpha}\right\}, \tag{19}$$

where $\chi^2_{1,\alpha}$ is the percentage point of the chi–squared distribution with one degree of freedom which leaves an area of α in the upper tail of the distribution. The same methodology can be applied to the m samples density ratio model but the corresponding log likelihood l is given by Eq. (10). The point is illustrated in Fig. 2 where the upper plot corresponds to a three sample problem from the Lognormal distribution with corresponding sample sizes $n_1 = n_2 = n_3 = 50$ while the lower plot shows the log likelihood from three normal samples of the same size, that is $n_1 = n_2 = n_3 = 50$. For the lognormal case, the confidence interval (19) is given by $(-0.490, 0.735)$ and for the normal example is $(0.438, 1.794)$. In both cases the maximum is achieved at the true value of λ.

Concluding this section we point out that another way of obtaining an estimate of λ—for the two samples problem—is to maximize the profile likelihood

$$l = -\sum_{i=1}^{2}\sum_{j=1}^{n_i} \log\left[1 + \rho_1 \exp(\alpha_1 + \beta_1 h(x_{ij}, \lambda))\right] + \sum_{j=1}^{n_1}\left(\alpha_1 + \beta_1 h(x_{1j}, \lambda)\right)$$

by recalling Eqs.(18) and (6) and the fact that p has been fixed to 1. However asymptotic approximation is valid for very large sample sizes; the same remarks hold true for the m samples density ratio model.

5 A two sample problem

The Current Population Survey (CPS) is used to supplement census information between census years. These data consist of a random sample of 534 persons from the CPS, with information on wages and other characteristics of the workers, including sex, number of years of education, years of work experience and other variables of interest and can be obtained from `http://lib.stat.cmu.edu/datasets/CPS_85_Wages`. We examine whether or not there are differences on wages between males and females, without adjusting for covariates, by the method outlined in this manuscript—there are 245 female subjects and 289 males. The boxplots

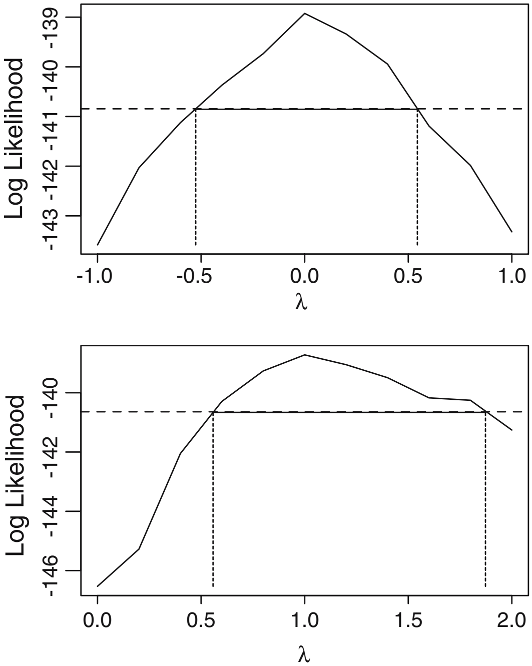

Fig. 2 Estimation of λ via profiling. The *upper plot* corresponds to a three sample problem from the Lognormal distribution with corresponding sample sizes $n_1 = n_2 = n_3 = 50$, $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 0.5$ and $\sigma^2 = 1$ and the true model holds for $\lambda = 0$. The lower plot shows the log likelihood from three normal samples with $n_1 = n_2 = n_3 = 50$, $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 0.5$ and $\sigma^2 = 1$ and the true model holds for $\lambda = 1$. Dotted lines show the 95% confidence interval (19). All results are based on 1,000 simulations

of the raw data are illustrated in Fig. 3 where it is seen that assessing differences between the means of the two samples via the t-test is possible only after transformation since both distributions appear skewed with unequal variances. A square root transformation of the raw data (right panel of Fig.3) shows resemblance to normality, yet the comparison need to be made in the *transformed scale*.

Turning to the method outlined earlier, Fig. 4 illustrates the maximized log likelihood (Eq. 5) as a function of λ when using transformation (18) for positive data. The parameter λ varies from −2 to 2 by 0.1 while the dashed lines indicate the 95% confidence interval obtained by means of (19). The resulting confidence interval is (−0.502, 1.177) and includes the values 0 (logarithmic transformation), 0.5 (square root transformation) and 1 (no transformation).

Specifically, the model applied to those data has the following form

$$\log \frac{g(x \mid \text{male})}{g(x \mid \text{female})} = \alpha_1 + \beta_1 h(x, \lambda),$$

and as the results show the value of $\hat{\beta}_1$ is always positive indicating that males earn more than females without adjusting any covariate effect. In fact for $\lambda = 0$, $\hat{\beta}_{1;0} = 0.873717$ with $\sqrt{\widehat{\mathrm{Var}[\hat{\beta}_{1;0}]}} = 0.17672$, for $\lambda = 0.5$, $\hat{\beta}_{1;0.5} = 0.30060$ with

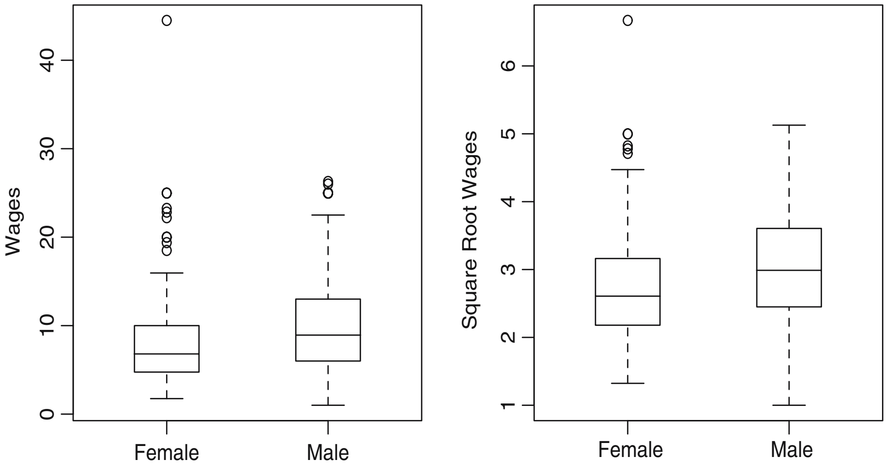

Fig. 3 Boxplots of wage data for males and females from the Current Population Survey. The *left panel* shows the boxplots for the raw data while the *right panel* shows the same information for the square root transformed data

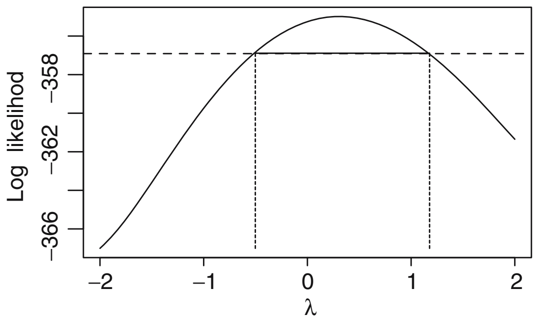

Fig. 4 Estimation of λ via profiling for the wage data. Dotted lines show the 95% confidence interval (Eq. 19)

Table 6 Testing for the wage data

Procedure	Test statistic	p-value
t-test after square root transformation	-5.1582	3.524e-07
t-test after log transformation	-5.1658	3.390e-07
Wilcoxon	26025	1.304e-07
Z with $\lambda = 0$	4.941	7.77e-07
$\mathcal{X}_1^2$ with $\lambda = 0$	26.023	3.37e-07
Z with $\lambda = 0.5$	4.915	8.87e-07
$\mathcal{X}_1^2$ with $\lambda = 0.5$	22.868	1.74e-06

$\sqrt{\widehat{\mathrm{Var}[\hat{\beta}_{1;0.5}]}} = 0.06116$ and for $\lambda = 1$, $\hat{\beta}_{1;1} = 0.09099$ with $\sqrt{\widehat{\mathrm{Var}[\hat{\beta}_{1;1}]}} = 0.01978$. Table 6 summarizes the results of testing procedures (14) and (15) for $\lambda = 0, 0.5$ including standard textbook tests for the two sample problem. In all cases considered notice the small magnitude of the p-values while the results of our analysis are consistent with standard inferential output from well known procedures. It is worth repeating the point that the new methodology does not depend on the transformation in the sense that the t-test does.

6 Conclusions

The results indicate the wide applicability of the density ratio model for the analysis of two, or more general m samples. Inference can be carried out in a direct way and the output can be used for semiparametric comparison of independent samples. However, as we showed by both theory and examples, the results are affected by the assumed linear form in the following sense:

1. Introduction of bias and large standard errors
2. Power loss when the fitted model falls far from the true model.

Therefore it is necessary to estimate the linear form and a natural framework is the Box–Cox family of transformations which allows for real valued data to be analyzed in a coherent way:

– First we estimate λ upon resorting to a standard profiling procedure.
– For the chosen λ, fit the associated density ratio model.

The methodology can be extended quite naturally to include further powers of x as in

$$\log \frac{g_1(x)}{g_2(x)} = \alpha_1 + \beta_{11} h(x, \lambda_1) + \beta_{12} h(x, \lambda_2)$$

and so on, with $\lambda_1 \neq \lambda_2$. The estimation procedure should be slightly modified so that estimation of both λ_1, λ_2 can be carried out by profiling for the allowable range of values.

Appendix

A.1 Proof of Proposition 3.1

Recall that $p = 1$ and set $\boldsymbol{\theta}_1 = (\alpha_1, \beta_1)'$ and

$$S(\boldsymbol{\theta}_1, \lambda) = \begin{bmatrix} \partial l(\boldsymbol{\theta}_1)/\partial \alpha_1 \\ \partial l(\boldsymbol{\theta}_1)/\partial \beta_1 \end{bmatrix}$$

for the two dimensional vector which consists of Eqs.(7) and (8). By the mean value theorem

$$S(\boldsymbol{\theta}_1, \lambda) = S(\boldsymbol{\theta}_1, \lambda_0) + (\lambda - \lambda_0) \begin{bmatrix} \partial^2 l(\boldsymbol{\theta}_1)/\partial \lambda \partial \alpha_1 \\ \partial^2 l(\boldsymbol{\theta}_1)/\partial \lambda \partial \beta_1 \end{bmatrix}_{\lambda = \lambda^\star} \tag{20}$$

where $\lambda^\star$ lies in the line segment connecting λ and λ_0. Let $\hat{\boldsymbol{\theta}}_{1,\lambda} = (\hat{\alpha}_{1;\lambda}, \hat{\beta}_{1,\lambda})'$ be the resulting maximum likelihood estimators under an incorrectly specified model. Then a new expansion of the score around the true value leads to

$$\mathbf{0} = S(\hat{\boldsymbol{\theta}}_{1,\lambda}, \lambda) = S(\boldsymbol{\theta}_1, \lambda) + \frac{\partial^2 l(\boldsymbol{\theta}^\star, \lambda)}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1'} \left(\hat{\boldsymbol{\theta}}_{1,\lambda} - \boldsymbol{\theta}_1 \right)$$

where $\boldsymbol{\theta}^\star$ lies in the line segment connecting $\boldsymbol{\theta}_1$ and $\hat{\boldsymbol{\theta}}_{1,\lambda}$. It follows that

$$\begin{aligned}\left(\hat{\boldsymbol{\theta}}_{1,\lambda}-\boldsymbol{\theta}_1\right) &= -\left[\frac{\partial^2 l(\boldsymbol{\theta}^\star,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]^{-1} S(\boldsymbol{\theta}_1,\lambda)\\ &= -\left[\frac{\partial^2 l(\boldsymbol{\theta}^\star,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]^{-1}[S(\boldsymbol{\theta}_1,\lambda)-\mathrm{E}(S(\boldsymbol{\theta}_1,\lambda))]\\ &\quad -\left[\frac{\partial^2 l(\boldsymbol{\theta}^\star,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]^{-1}\mathrm{E}(S(\boldsymbol{\theta}_1,\lambda))\end{aligned}$$

The last equation can be rewritten as

$$\begin{aligned}&\left(\hat{\boldsymbol{\theta}}_{1,\lambda}-\boldsymbol{\theta}_1\right)+\left[\frac{\partial^2 l(\boldsymbol{\theta}_1,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]^{-1}\mathrm{E}(S(\boldsymbol{\theta}_1,\lambda))\\ &\quad= -\left[\frac{\partial^2 l(\boldsymbol{\theta}^\star,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]^{-1}[S(\boldsymbol{\theta}_1,\lambda)-\mathrm{E}(S(\boldsymbol{\theta}_1,\lambda))] \qquad (21)\\ &\quad\quad -\left\{\left[\frac{\partial^2 l(\boldsymbol{\theta}^\star,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]^{-1}-\left[\frac{\partial^2 l(\boldsymbol{\theta}_1,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]^{-1}\right\}\mathrm{E}(S(\boldsymbol{\theta}_1,\lambda)).\end{aligned}$$

Define the following quantities

$$\begin{aligned}Y &= \frac{\rho_1\exp(\alpha_1+\beta_1 h(X,\lambda))}{1+\rho_1\exp(\alpha_1+\beta_1 h(X,\lambda))},\\ \mathbf{H}(X,\lambda) &= \begin{bmatrix}1 & h(X,\lambda)\\ h(X,\lambda) & h^2(X,\lambda)\end{bmatrix}, \qquad (22)\\ \pi(X,\lambda_0) &= \exp(\alpha_1+\beta_1 h(X,\lambda_0)),\\ h_\lambda(X,\lambda) &= \frac{dh(X;\lambda)}{d\lambda}.\end{aligned}$$

Then the following facts hold upon noticing that expectation is calculated componentwise with respect to $G_2(x)$ and $n_1,\ n_2\to\infty$ such that $n_1/n_2\to\rho_1$ in a neighborhood of λ such that $\mid\lambda-\lambda_0\mid< n^{-1/2}$:

$$\frac{1}{n}\mathrm{E}(S(\boldsymbol{\theta}_1,\lambda))-\left(-(\lambda-\lambda_0)\frac{\beta_1}{1+\rho_1}\mathrm{E}[h_\lambda(X,\lambda_0)Y(1,h(X,\lambda_0))']\right)=o(1), \qquad (23)$$

a fact that follows from Eq. (20). Furthermore, denote by

$$\mathbf{A}_\lambda\equiv\frac{1}{1+\rho_1}\mathrm{E}\left[Y(1-Y)(1+\rho_1\pi(X,\lambda_0)\mathbf{H}(X,\lambda)\right] \qquad (24)$$

to obtain

$$-\frac{1}{n}\left[\frac{\partial^2 l(\boldsymbol{\theta}_1,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]-\mathbf{A}_\lambda=o_p(1) \qquad (25)$$

and

$$\frac{1}{\sqrt{n}}\mathbf{V}_{\lambda}^{-1/2}\left[S(\boldsymbol{\theta}_1,\lambda)-\mathrm{E}(S(\boldsymbol{\theta}_1,\lambda))\right]\to\mathcal{N}_2(\mathbf{0},\mathbf{I}_2) \tag{26}$$

where

$$\begin{aligned}\mathbf{V}_{\lambda}=\frac{1}{1+\rho_1}&\left\{\mathrm{E}\left[\left(Y^2+\rho_1(1-Y)^2\pi(X,\lambda_0)\right)\mathbf{H}(X,\lambda)\right]\right.\\&-\rho_1\mathrm{E}\left[(1-Y)\pi(X;\lambda_0)(1,h(X,\lambda))'\right]\\&\quad\mathrm{E}\left[(1-Y)\pi(X;\lambda_0)(1,h(X,\lambda))\right]\\&\left.-\mathrm{E}\left[Y(1,h(X,\lambda))'\right]\mathrm{E}\left[Y(1,h(X,\lambda))\right]\right\}\end{aligned} \tag{27}$$

Moreover

$$\left\{\left[\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta}^{\star},\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]^{-1}-\left[\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta}_1,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]^{-1}\right\}=o_p(1) \tag{28}$$

and

$$\frac{1}{\sqrt{n}}\mathrm{E}[S(\boldsymbol{\theta}_1,\lambda)]=O(1) \tag{29}$$

when λ is such that $\mid\lambda-\lambda_0\mid<\delta=cn^{-1/2}$. For instance, to show (Eq. 28) consider the following Taylor expansion

$$\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta},\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}=\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta},\lambda_0)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}+(\lambda-\lambda_0)\frac{\partial}{\partial\lambda}\left[\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta},\lambda_0)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]+o(n^{-1/2}).$$

But, for the second summand, $\mid\lambda-\lambda_0\mid<cn^{-1/2}$ and the other term converges to a finite limit in probability provided that $\mathrm{E}[\mid h_{\lambda}(X,\lambda)\mid(1+\rho_1\pi(X,\lambda_0))]$, $\mathrm{E}[\mid h_{\lambda}(X,\lambda)h(X,\lambda)\mid(1+\rho_1\pi(X,\lambda_0))]$ and $\mathrm{E}[\mid h_{\lambda}(X,\lambda)h^2(X,\lambda)\mid(1+\rho_1\pi(X,\lambda_0))]$ are all finite when expectation is taken with respect to $G_2(x)$ in a $n^{-1/2}$ neighborhood of the true value λ_0. Hence

$$\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta}^*,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}-\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta}_1,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}=\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta}^*,\lambda_0)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}-\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta}_1,\lambda_0)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}+o_p(1)\to\mathbf{0},$$

in probability from Eq. (25) and the consistency of $\hat{\boldsymbol{\theta}}_{1,\lambda_0}$, see Qin and Zhang (1997). The result follows by combining Eq. 21 with Eqs. 23–29.

A.2 Proof of Proposition 3.2

The generalization to the m'sample problem follows by analogous arguments. Recall the associated log–likelihood (Eq. 10), $p=1$ and that $\rho_i=n_i/n_m$ and define the following quantities

$$
\begin{aligned}
Y_i &= \frac{\rho_i \exp(\alpha_i + \beta_i h(X,\lambda))}{1+\sum_{k=1}^{m-1} \rho_k \exp(\alpha_k + \beta_k h(X,\lambda))}, \quad i = 1,2,\dots,m-1,\\
\mathbf{Y} &= (Y_1, Y_2, \dots, Y_{m-1})',\\
\mathbf{h}_i(X,\lambda) &= (1, h(X,\lambda))' \oplus \mathbf{e}_i, \quad i = 1,2,\dots,m-1,\\
\tilde{Y}_{ij} &= \begin{cases} Y_i(1-Y_i), & \text{if } i = j\\ -Y_iY_j, & \text{if } i \neq j \end{cases}\\
\tilde{\mathbf{Y}} &= (\tilde{Y}_{ij}), i, j = 1,2,\dots,m-1,\\
\pi_i(X;\lambda_0) &= \exp(\alpha_i + \beta_i h(X;\lambda_0)) \quad i = 1,2,\dots,m-1,
\end{aligned}
\tag{30}
$$

where $\oplus$ stands for the Kronecker product and $\mathbf{e}_i$ denotes the unit vector with 1 at the i'th position and 0 otherwise, $i = 1, 2, \dots, m-1$. Hence $\mathbf{h}_i(X,\lambda)$ is an $2(m-1)\times 1$ vector.

By redefining $\boldsymbol{\theta}_1 = (\alpha_1, \dots, \alpha_{m-1}, \beta_1, \dots, \beta_{m-1})'$, a $2(m-1)\times 1$ vector, expressions (23,24,25,26,27) become

$$
\begin{aligned}
&\frac{1}{n}\mathrm{E}(S(\boldsymbol{\theta}_1,\lambda)) - \left(-\frac{(\lambda-\lambda_0)}{1+\sum_{i=1}^{m-1}\rho_i}\left\{\sum_{i=1}^{m-1}\beta_i \mathrm{E}[h_\lambda(X,\lambda_0)Y_i\mathbf{h}_i(X,\lambda_0)\right.\right.\\
&\quad\left.\left.\times\ (1+\sum_{j\neq i}(1+\rho_j\pi_j(X,\lambda_0)))]\right\}\right) = o(1),
\end{aligned}
$$

$$
\mathbf{A}_{\lambda;m} \equiv \frac{1}{1+\sum_{i=1}^{m-1}\rho_i}\mathrm{E}\left[(\tilde{\mathbf{Y}}\oplus\mathbf{H}(X,\lambda))(1+\sum_{i=1}^{m-1}\rho_i\pi_i(X,\lambda_0)\right], \tag{31}
$$

$$
-\frac{1}{n}\left[\frac{\partial^2 l(\boldsymbol{\theta}_1,\lambda)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right] - \mathbf{A}_{\lambda;m} = o_p(1)
$$

and

$$
\frac{1}{\sqrt{n}}\mathbf{V}_{\lambda;m}^{-1/2}\left[S(\boldsymbol{\theta}_1,\lambda) - \mathrm{E}(S(\boldsymbol{\theta}_1,\lambda))\right] \to \mathcal{N}_{2(m-1)}(\mathbf{0}, \mathbf{I}_{2(m-1)})
$$

where

$$
\begin{aligned}
\mathbf{V}_{\lambda;m} &= \frac{1}{1+\sum_{i=1}^{m-1}\rho_i}\\
&\quad\times\left\{\mathrm{E}\left[\mathbf{Y}\mathbf{Y}' + \left(\sum_{i=1}^{m-1}(\mathbf{e}_i-\mathbf{Y})(\mathbf{e}_i-\mathbf{Y})'\rho_i\pi_i(X,\lambda_0)\right)\oplus\mathbf{H}(X,\lambda)\right]\right.\\
&\quad-\sum_{i=1}^{m-1}\rho_i\mathrm{E}[(\mathbf{e}_i-\mathbf{Y})\oplus(1,h(X,\lambda))'\pi_i(x,\lambda)]\\
&\quad\times\mathrm{E}[(1,h(X,\lambda))\oplus(\mathbf{e}_i-\mathbf{Y})'\pi_i(x,\lambda)]\\
&\quad\left.-\ \mathrm{E}[\mathbf{Y}\oplus(1,h(X,\lambda))']\mathrm{E}[(1,h(X,\lambda))\oplus\mathbf{Y}]\right\}
\end{aligned}
\tag{32}
$$

The desired result follows by the same arguments as in the case of the two samples problem.

Acknowledgements Thanks are due to both reviewers whose constructive and helpful criticism improved the presentation.

References

Box, G. E. P., D. R. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, *26*, 211–252. with discussion.
Cox, D. R. and E. J. Snell (1989). *The analysis of binary data* (2nd ed). London: Chapman & Hall
Fokianos, K. (2003). *Foundations of statistical inference* In: Y. Haitovsky, H. R. Lerche and Y. Ritov (eds.), Haidelberg:Physica–Verlag, pp 131–140
Fokianos, K., B. Kedem, J. Qin, J. Haferman, & D. A. Short (1998). On combining instruments. *Journal of Applied Meteorology*, *37*, 220–226.
Fokianos, K., B. Kedem, J. Qin, & D. A. Short (2001). A semiparametric approach to the one-way layout, *Technometrics*, *43*, 56–64.
Fokianos, K., A. Peng, & J. Qin (1999). A generalized-moments specification test for the logistic link. *The Canadian Journal of Statistics*, *27*, 735–750.
Gilbert, P. B. (2000). Large sample theory of maximum likelihood estimates in semiparametric biased sampling models. *The Annals of Statistics*, *28*, 151–194.
Gilbert, P. B., S. R. Lele, & Y. Vardi (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*, *86*, 27–43.
Gill, R. D., Y. Vardi, & J. A. Wellner (1988). Large sample theory of empirical distributions in biased sampling models. *The Annals of Statistics*, *16*, 1069–1112.
Kedem, B., D. B. Wolff, & K. Fokianos (2004). Statistical comparisons of algorithms. *IEEE Transactions on Instrumantetion and Measurement*, *53*, 770–776.
Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*, 237–249.
Owen, A. B. (2001). *Empirical likleihood*. Boca Raton, Florida: Chapman and Hall/CRC
Prentice, R. L. & R. Pyke (1979). Logistic disease incidence models and case–control studies. *Biometrika*, *66*, 403–411.
Qin, J. (1998). Inferences for case-control data and semiparametric two–sample density ratio models. *Biometrika*, *85*, 619–630.
Qin, J., M. Barwick, R. Ashbolt, & T. Dwyer (2002). Quantifying the change of melanoma incidence by Breslow thickness, *Biometrics*, *58*, 665–670.
J. Qin, & J. F. Lawless (1994). Empirical likelihood and general estimating functions. *Annals of Statistics*, *22*, 300–325.
Qin, J. & B. Zhang (1997). A goodness of fit test for the logistic regression model based on case–control data. *Biometrika*, *84*, 609–618.
Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *The Annals of Statistics*, *10*, 616–620.
Vardi, Y. (1985). Empirical distribution in selection bias models. *The Annals of Statistics*, *13*, 178–203.
White, I. R. & S. G. Thompson (2003). Choice of test for comparing two groups, with particular application to skewed outcomes. *Statistics in Medicine*, *22*, 1205–1215.
Zhang, B. (2001). An information matrix test for logistic regression models based on case-control data. *Biometrika 88*, 921–932.