

## OBTAINING INVARIANT ESTIMATORS FOR MULTIPLE RELATIVE RATES FROM THEIR GIVEN PAIRWISE ESTIMATORS

Zhang, Zhong-Zhang  
Institute of Statistics, Beijing Polytechnic University

Yanagawa, Takashi  
Graduate School of Mathematics, Kyushu University

<https://doi.org/10.5109/13470>

---

出版情報 : Bulletin of informatics and cybernetics. 30 (1), pp.69-77, 1998-03. Research  
Association of Statistical Sciences

バージョン :

権利関係 :

# OBTAINING INVARIANT ESTIMATORS FOR MULTIPLE RELATIVE RATES FROM THEIR GIVEN PAIRWISE ESTIMATORS

By

Zhong-Zhan ZHANG \* and Takashi YANAGAWA †

## Abstract

This article is concerned with the estimation of relative rates of population characteristics of multiple populations when the estimators of the relative rates have been given from the relevant two populations. We generalize the invariance of relative rates estimators and the projection method introduced in previous articles, and propose a class of estimators obtained from the generalized projection. It is shown that the estimators are invariant. The consistency and the asymptotic normality of the estimators are proved.

**Key words:** *Asymptotic covariance matrix; Begun-Reid estimator; Consistency; Mantel-Haenszel estimator; Projection method; Ridge estimator.*

## 1. INTRODUCTION

Consider  $I+1$  populations indexed by  $0, 1, \dots, I$  ( $I > 1$ ), and let  $\mu_i$  be a population characteristic of population  $i$ ,  $i = 0, 1, \dots, I$ . Frequently, the relative rate parameters  $\theta_{i'}^{(i)} = \mu_{i'}/\mu_i$ ,  $i, i' = 0, 1, \dots, I$ , such as relative risks and rate ratios in epidemiology and hazard rates in survival analysis, are the main concern of researchers.

It is easy to see from the definition of relative rate that

$$\theta_{i'}^{(i)} = \theta_{i'}^{(k)}/\theta_i^{(k)}, \quad i, i', k = 0, 1, \dots, I, \quad (1.1)$$

which we call the invariance of the relative rates with respect to the selection of a base population.

Introducing mathematical models one may estimate  $\{\theta_{i'}^{(i)}\}$  using all populations by conventional methods, such as likelihood methods (unconditional likelihood, conditional likelihood, partial likelihood) (Cox 1975, Breslow et al. 1978, Anderson et al. 1994) and estimating equations (Liang 1987). However, the blind faith of these estimates with little awareness of strong assumption implicit to the mathematical models leads to the possibility of obtaining misleading results if the assumptions are violated. On the other hand,  $\theta_{i'}^{(i)}$  may be estimated from population  $i$  and population  $i'$ . We call

---

\* Graduate School of Mathematics, Kyushu University, Fukuoka, 812, Japan; Institute of Statistics, Beijing Polytechnic University, Beijing, 100022, China

† Graduate School of Mathematics, Kyushu University, Fukuoka, 812, Japan

the estimators pairwise estimators. They are naive and could be used as a diagnostic tool for the conventional models. In fact, as shown by Yanagawa and Fujii (1995), and Zhang and Yanagawa (1997), many pairwise estimators are informative, and may be combined to get alternative estimators which are competing with the estimators from the conventional methods. A common disadvantage of many pairwise estimators is that they do not satisfy relationship (1.1). This implies that the inference based on the pairwise estimators could depend on the base population selected for the reference of comparison. For example, it is well known that Mantel-Haenszel type estimators have no invariance property (Yanagawa and Fujii 1995, Zhang and Yanagawa 1997).

Any estimators which do not satisfy relationship (1.1) have been called the lack of invariance with respect to the selection of a base population (Yanagawa and Fujii 1995, Zhang and Yanagawa 1997). In this article, we furthermore note that even if estimators of  $\{\theta_{i'}^{(i)}\}$  from a base population satisfy (1.1), they may still depend on the base population. For example, suppose that we have got pairwise estimators  $\{b_i^{(0)}\}$  of  $\{\theta_i^{(0)}\}$  taking population 0 as the base, and also  $\{b_i^{(1)}\}$  of  $\{\theta_i^{(1)}\}$  taking population 1 as the base. Then we may construct two estimators  $\hat{\theta}_{i'}^{(i)}$  and  $\tilde{\theta}_{i'}^{(i)}$  of  $\theta_{i'}^{(i)}$  by

$$\begin{aligned}\hat{\theta}_{i'}^{(i)} &= b_{i'}^{(0)}/b_i^{(0)} \quad (i \neq 0), & \hat{\theta}_i^{(0)} &= b_i^{(0)}, \quad i = 1, 2, \dots, I; \\ \tilde{\theta}_{i'}^{(i)} &= b_{i'}^{(1)}/b_i^{(1)} \quad (i \neq 1), & \tilde{\theta}_i^{(1)} &= b_i^{(1)}, \quad i = 0, 2, \dots, I.\end{aligned}$$

It is clear that the two estimators satisfy (1.1), but  $\hat{\theta}_{i'}^{(i)} \neq \tilde{\theta}_{i'}^{(i)}$  if  $\{b_{i'}^{(i)}\}$  do not satisfy (1.1). We generalize the concept of invariance as follows.

**DEFINITION 1.1.** Denote the relative rate vector by  $\theta = (\theta_1^{(0)}, \dots, \theta_I^{(0)}, \theta_2^{(1)}, \dots, \theta_I^{(1)}, \dots, \theta_I^{(I-1)})^\tau$ . Suppose that we have an estimator  $(\hat{\theta}_0^{(j)}, \hat{\theta}_1^{(j)}, \dots, \hat{\theta}_I^{(j)})$  of  $(\theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_I^{(j)})$  taking population  $j$  as the base population. Define an estimator of  $\theta$

$$\hat{\theta}(j) = (\hat{\theta}_1^{(0)}(j), \dots, \hat{\theta}_I^{(0)}(j), \hat{\theta}_2^{(1)}(j), \dots, \hat{\theta}_I^{(1)}(j), \dots, \hat{\theta}_I^{(I-1)}(j))^\tau$$

by

$$\hat{\theta}_{i'}^{(i)}(j) = \hat{\theta}_{i'}^{(j)}/\hat{\theta}_i^{(j)}, \quad i < i' = 1, \dots, I.$$

If  $\hat{\theta}(j)$  does not depend on  $j$ , we call the estimator *invariant with respect to the selection of a base population*.

This article generalizes the projection method introduced by Yanagawa and Fujii (1995) and considers a class of relative rates estimators obtained from the generalized projection. It is shown that the estimators are invariant with respect to the selection of a base population. The paper is organized as follows. Section 2 gives the projection method estimators and discusses their properties; Section 3 shows the invariance; in Section 4, examples are presented to illustrate the usefulness of the projection method estimators.

## 2. PROJECTION METHOD ESTIMATORS

Suppose that pairwise estimators  $\{b_{i'}^{(i)}\}$  of  $\{\theta_{i'}^{(i)}\}$  are given, and set population 0 as the base population. Consider the following model to estimate  $\{\theta_{i'}^{(0)}\}$  from  $\{b_{i'}^{(i)}\}$ :

$$\log b_{i'}^{(i)} = \log \theta_{i'}^{(0)} - \log \theta_{i'}^{(i)} + \varepsilon_{i'}^{(i)}, \quad i < i', \quad i' = 1, \dots, I. \quad (2.1)$$

Denoting

$$\begin{aligned} \boldsymbol{\alpha} &= (\theta_1^{(0)}, \dots, \theta_I^{(0)})^\tau, & \boldsymbol{\beta} &= \log \boldsymbol{\alpha}, \\ \mathbf{b} &= (b_1^{(0)}, \dots, b_I^{(0)}, b_2^{(1)}, \dots, b_I^{(1)}, \dots, b_I^{(I-1)})^\tau, & \mathbf{y} &= \log \mathbf{b}, \\ D &= \begin{pmatrix} E_I & -\mathbf{1}_{I-1}^\tau & 0 & \dots & 0 \\ & E_{I-1} & -\mathbf{1}_{I-2}^\tau & \dots & -1 \\ & & E_{I-2} & & 1 \\ & & & & \\ & & & & \end{pmatrix}^\tau, & \boldsymbol{\eta} &= \log \boldsymbol{\theta}, \\ \boldsymbol{\varepsilon} &= (\varepsilon_1^{(0)}, \dots, \varepsilon_I^{(0)}, \varepsilon_2^{(1)}, \dots, \varepsilon_I^{(1)}, \dots, \varepsilon_I^{(I-1)})^\tau, \end{aligned} \quad (2.2)$$

where  $E_q$  stands for the identity matrix of order  $q \times q$ , and  $\mathbf{1}_q$  for the column  $q$  vector of 1, (2.1) can be rewritten as

$$\mathbf{y} = D\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.3)$$

and also  $\boldsymbol{\eta}$  may be represented as  $\boldsymbol{\eta} = D\boldsymbol{\beta}$ .

**DEFINITION 2.1.** Let  $\rho(\cdot, \cdot): \mathcal{R}^h \times \mathcal{R}^h \rightarrow [0, \infty)$  be a continuous function such that  $\rho(\mathbf{u}, \mathbf{u}) = 0$  for each  $\mathbf{u} \in \mathcal{R}^h$ , where  $h = I(I+1)/2$ . When  $\mathbf{y}$ , or equivalently  $\mathbf{b}$ , is given, we define the *projection method estimator of  $\boldsymbol{\beta}$  with respect to  $\rho$*  as

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \rho(\mathbf{y}, D\boldsymbol{\beta}) \quad (2.4)$$

provided that it exists. Accordingly the *projection method estimators* of  $\boldsymbol{\alpha}$  and  $\boldsymbol{\theta}$  are defined by  $\hat{\boldsymbol{\alpha}} = \exp \hat{\boldsymbol{\beta}}$  and  $\hat{\boldsymbol{\theta}} = \exp D\hat{\boldsymbol{\beta}}$ , respectively.

Usually  $\rho(\mathbf{u}, \mathbf{v})$  is taken as a function expressing some kind of distance between  $\mathbf{u}$  and  $\mathbf{v}$ . An example is the weighted sum of squared errors of (2.3)

$$L_{2,Q} = (\mathbf{y} - D\boldsymbol{\beta})^\tau Q(\mathbf{y} - D\boldsymbol{\beta}), \quad (2.5)$$

where  $Q$  is a given positive definite matrix. This leads to the estimator

$$\hat{\boldsymbol{\beta}}_Q = (D^\tau Q D)^{-1} D^\tau Q \mathbf{y}. \quad (2.6)$$

We shall discuss this estimator further and also other estimators in Section 4.

The following theorem shows the consistency and asymptotic distributions of the projection method estimators.

$$\text{Denote } \frac{\partial \rho(\mathbf{y}, \mathbf{z})}{\partial \mathbf{u}} = \frac{\partial \rho(\mathbf{u}, \mathbf{v})}{\partial \mathbf{u}} \Big|_{\mathbf{u}=\mathbf{y}, \mathbf{v}=\mathbf{z}} \quad \text{and} \quad \frac{\partial^2 \rho(\mathbf{y}, \mathbf{z})}{\partial \mathbf{u} \partial \mathbf{v}} = \frac{\partial^2 \rho(\mathbf{u}, \mathbf{v})}{\partial \mathbf{u} \partial \mathbf{v}} \Big|_{\mathbf{u}=\mathbf{y}, \mathbf{v}=\mathbf{z}}.$$

**THEOREM 2.2.** *Suppose that the function  $\rho(\mathbf{y}, D\boldsymbol{\beta})$  in definition 2 is strictly convex with respect to  $\boldsymbol{\beta}$ , and that  $\rho(\cdot, \cdot)$  is twice continuously differentiable in a neighbourhood of  $(\boldsymbol{\eta}, \boldsymbol{\eta})$ , i.e.  $\partial^2 \rho(\mathbf{u}, \mathbf{v})/\partial \mathbf{u}^2$ ,  $\partial^2 \rho(\mathbf{u}, \mathbf{v})/\partial \mathbf{u} \partial \mathbf{v}$  and  $\partial^2 \rho(\mathbf{u}, \mathbf{v})/\partial \mathbf{v}^2$  exist and are continuous with respect to  $\mathbf{u}$  and  $\mathbf{v}$ .*

(1) *If the pairwise estimator  $\mathbf{b}$  satisfies  $\mathbf{b} \xrightarrow{P} \boldsymbol{\theta}$  ( $n \rightarrow \infty$ ), where  $n$  is the total sample size, then*

$$\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta} \quad (n \rightarrow \infty).$$

(2) *If  $\mathbf{b}$  satisfies that  $\sqrt{n}(\mathbf{y} - \boldsymbol{\eta}) \xrightarrow{L} N(0, V_y)$  ( $n \rightarrow \infty$ ), then*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{L} N(0, V),$$

where

$$V = \left( D^\tau \Gamma_{22} D \right)^{-1} \Gamma_{12} V_y \Gamma_{12}^\tau \left( D^\tau \Gamma_{22} D \right)^{-1},$$

$$\Gamma_{12} = D^\tau \frac{\partial^2 \rho(\boldsymbol{\eta}, \boldsymbol{\eta})}{\partial \mathbf{u} \partial \mathbf{v}},$$

and

$$\Gamma_{22} = \frac{\partial^2 \rho(\boldsymbol{\eta}, \boldsymbol{\eta})}{\partial \mathbf{v}^2}.$$

**PROOF.** (1) By the Taylor expansion,

$$\begin{aligned} \rho(\mathbf{y}, D\boldsymbol{\gamma}) &= \rho(\mathbf{y}, \boldsymbol{\eta}) + D^\tau \frac{\partial \rho(\mathbf{y}, \boldsymbol{\eta})}{\partial \mathbf{v}} (\boldsymbol{\gamma} - \boldsymbol{\beta}) \\ &\quad + (\boldsymbol{\gamma} - \boldsymbol{\beta})^\tau \left( D^\tau \frac{\partial^2 \rho(\mathbf{y}, \boldsymbol{\eta})}{\partial \mathbf{v}^2} D \right) (\boldsymbol{\gamma} - \boldsymbol{\beta}) + R, \end{aligned}$$

where  $R = o(\|\boldsymbol{\gamma} - \boldsymbol{\beta}\|^2)$ . The right hand side except the reminder  $R$  is a quadratic function in  $\boldsymbol{\gamma}$ , achieving its minimum at

$$\boldsymbol{\gamma} = \boldsymbol{\beta} - \frac{1}{2} \left( D^\tau \frac{\partial^2 \rho(\mathbf{y}, \boldsymbol{\eta})}{\partial \mathbf{v}^2} D \right)^{-1} D^\tau \frac{\partial \rho(\mathbf{y}, \boldsymbol{\eta})}{\partial \mathbf{v}}.$$

Thus it follows that  $\rho(\mathbf{y}, D\boldsymbol{\gamma})$  has a local minimum at

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} - \frac{1}{2} \left( D^\tau \frac{\partial^2 \rho(\mathbf{y}, \boldsymbol{\eta})}{\partial \mathbf{v}^2} D \right)^{-1} D^\tau \frac{\partial \rho(\mathbf{y}, \boldsymbol{\eta})}{\partial \mathbf{v}} + o\left( D^\tau \frac{\partial \rho(\mathbf{y}, \boldsymbol{\eta})}{\partial \mathbf{v}} \right).$$

The strict convexity of  $\rho(\mathbf{y}, D\boldsymbol{\gamma})$  implies that the local minimum is the global minimum, i.e.  $\hat{\boldsymbol{\beta}}$  is the projection method estimator of  $\boldsymbol{\beta}$ . Since

$$D^\tau \frac{\partial \rho(\mathbf{y}, \boldsymbol{\eta})}{\partial \mathbf{v}} \xrightarrow{P} D^\tau \frac{\partial \rho(\boldsymbol{\eta}, \boldsymbol{\eta})}{\partial \mathbf{v}} = 0$$

and

$$D^\tau \frac{\partial^2 \rho(\mathbf{y}, \boldsymbol{\eta})}{\partial \mathbf{v}^2} D \xrightarrow{P} D^\tau \frac{\partial^2 \rho(\boldsymbol{\eta}, \boldsymbol{\eta})}{\partial \mathbf{v}^2} D,$$

the consistency is established.

(2) may be proved by using the Taylor expansion of  $D^\tau \frac{\partial \rho(\mathbf{y}, D\boldsymbol{\gamma})}{\partial \mathbf{v}}$  at  $\boldsymbol{\gamma} = \boldsymbol{\beta}$ .  $\square$

### 3. INVARIANCE OF PROJECTION METHOD ESTIMATORS

We first define the following terminology for temporary use.

**DEFINITION 3.1.** A linear transformation of  $\mathcal{R}^h$  into itself is called the *axes-kept* if its matrix is represented by the products of transformation matrices which exchange any two elements of  $h$  column vector and(or) change the signs of any elements.

Obviously, an axes-kept transformation is orthogonal. As an example, in  $\mathcal{R}^2$  the image of a point  $(x, y)$  for an axes-kept transformation will be one of the eight points:  $(x, y), (-x, y), (-x, -y), (x, -y), (y, x), (-y, x), (-y, -x), (y, -x)$ .

In this section we employ the following notation to establish the invariance property. Denote  $\mathcal{P}$  be the symmetric group of  $(0, 1, \dots, I)$ . Let  $\mathbf{p} = (i_0, \dots, i_I) \in \mathcal{P}$ ,  $\boldsymbol{\theta}_p = (\theta_{i_1}^{(i_0)}, \dots, \theta_{i_I}^{(i_0)}, \theta_{i_2}^{(i_1)}, \dots, \theta_{i_I}^{(i_1)}, \dots, \theta_{i_I}^{(i_{I-1})})^\tau$ ,  $\boldsymbol{\beta}_p = (\log \theta_{i_1}^{(i_0)}, \dots, \log \theta_{i_I}^{(i_0)})^\tau$ , and  $\mathbf{b}_p = (b_{i_1}^{(i_0)}, \dots, b_{i_I}^{(i_0)}, b_{i_2}^{(i_1)}, \dots, b_{i_I}^{(i_1)}, \dots, b_{i_I}^{(i_{I-1})})^\tau$  be a pairwise estimator of  $\boldsymbol{\theta}_p$ .

With this notation it follows that

$$D\boldsymbol{\beta}_p = \log \boldsymbol{\theta}_p.$$

Furthermore, for  $\boldsymbol{\eta}$  defined in (2.2), we may represent  $\log \boldsymbol{\theta}_p$  as

$$\log \boldsymbol{\theta}_p = A_p \boldsymbol{\eta} = A_p \log \boldsymbol{\theta}, \quad (3.1)$$

where  $A_p$  is an axes-kept transformation matrix. Therefore,  $\boldsymbol{\theta} = \exp A_p^{-1} D\boldsymbol{\beta}_p$ . Now the projection method estimator of  $\boldsymbol{\beta}_p$  is given by  $\hat{\boldsymbol{\beta}}_p = \arg \min_{\boldsymbol{\beta}} \rho(\log \mathbf{b}_p, D\boldsymbol{\beta}_p)$ . Thus the estimator of  $\boldsymbol{\theta}$  corresponding to  $\mathbf{p} \in \mathcal{P}$  may be obtained as  $\hat{\boldsymbol{\theta}}(\mathbf{p}) = \exp A_p^{-1} D\hat{\boldsymbol{\beta}}_p$ . Recall Definition 1.1. Then to prove the invariance of the estimator, it is sufficient to prove that  $\hat{\boldsymbol{\theta}}(\mathbf{p})$  does not depend on  $\mathbf{p}$ .

**THEOREM 3.2.** *Suppose that the following conditions are satisfied.*

- (a) *The pairwise estimator  $\mathbf{b}_p$  of  $\boldsymbol{\theta}_p$  may be represented by  $\log \mathbf{b}_p = A_p \mathbf{y}$  for every  $\mathbf{p} \in \mathcal{P}$ , where  $\mathbf{y}$  is defined in (2.2).*
- (b) *For each value of  $\mathbf{y}$ , the solution of (2.4) exists uniquely.*
- (c) *For any axes-kept transformation  $A$  and any  $\mathbf{u}, \mathbf{v} \in \mathcal{R}^h$*

$$\rho(\mathbf{u}, \mathbf{v}) = \rho(A\mathbf{u}, A\mathbf{v}).$$

*Then the estimator  $\hat{\boldsymbol{\theta}}(\mathbf{p})$  does not depend on  $\mathbf{p}$ , hence it is invariant with respect to the selection of a base population.*

**PROOF.** From condition (a) and (c), we have

$$\rho(\log \mathbf{b}_p, D\boldsymbol{\beta}_p) = \rho(A_p \mathbf{y}, D\boldsymbol{\beta}_p) = \rho(\mathbf{y}, A_p^{-1} D\boldsymbol{\beta}_p).$$

On the other hand, from (3.1) and (c)

$$\rho(\log \mathbf{b}_p, D\boldsymbol{\beta}_p) = \rho(A_p \mathbf{y}, \log \boldsymbol{\theta}_p) = \rho(A_p \mathbf{y}, A_p D\boldsymbol{\beta}) = \rho(\mathbf{y}, D\boldsymbol{\beta}).$$

Thus putting  $\hat{\boldsymbol{\beta}}_p = \arg \min_{\boldsymbol{\beta}_p} \rho(\mathbf{y}, A_p^{-1} D\boldsymbol{\beta}_p)$  and  $\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \rho(\mathbf{y}, D\boldsymbol{\beta})$ , we have

$$A_p^{-1} D\hat{\boldsymbol{\beta}}_p = \hat{\boldsymbol{\beta}}, \quad (3.2)$$

since the solution of (2.4) is unique from assumption (b).  $\square$

Noting that the right-hand side of (3.2) does not depend on  $\mathbf{p}$ , the theorem is obtained.

Many useful distance functions satisfy condition (c). For example,  $L_q(\mathbf{u}, \mathbf{v}) = \sum |u_i - v_i|^q, 0 < q \leq \infty$  have this property. Condition (a) is equivalent to  $b_i^{(i)} = (b_i^{(i')})^{-1}, i, i' = 0, 1, \dots, I$ . Many natural pairwise estimators satisfy the condition. For example, the Mantel-Haenszel estimators of odds ratios (Mantel and Haenszel 1959), Begun-Reid estimators of hazard rates (Begun and Reid 1983), pairwise Cox partial likelihood estimators of hazard rates (Crowley, Liu and Veolkel 1982) have this property.

#### 4. APPLICATIONS

To apply the projection method, we must suppose that pairwise estimators have been given. The selection of pairwise estimator depends on both the background of the problem and the favour of researchers. In the following, we present several available distance functions and also some comments on the selection of pairwise estimator.

##### 4.1. Linear projection estimators

Taking the weighted sum of squared errors (2.5) as  $\rho(\mathbf{y}, D\boldsymbol{\beta})$ , we get estimator  $\hat{\boldsymbol{\beta}}_Q$  given in (2.6), which is a linear function of  $\mathbf{y}$ . We call the estimator of  $\boldsymbol{\theta}$  defined by  $\hat{\boldsymbol{\theta}}_Q = \exp D\hat{\boldsymbol{\beta}}_Q$  the linear projection estimator. When pairwise estimators satisfy the conditions of Theorem 2.2,  $\hat{\boldsymbol{\beta}}_Q$  has the asymptotic covariance matrix  $(D^T Q D)^{-1} D^T Q V_y Q D (D^T Q D)^{-1}$ .

The following corollary gives the invariance property of the linear projection estimator.

**COROLLARY 4.1.** *Suppose that the given pairwise estimators satisfy condition (a) of Theorem 3.2. For any  $\mathbf{p} \in \mathcal{P}$ , putting*

$$\hat{\boldsymbol{\beta}}_{p,Q} = \arg \min_{\boldsymbol{\beta}_p} (A_p \mathbf{y} - D\boldsymbol{\beta}_p)^T A_p Q A_p^{-1} (A_p \mathbf{y} - D\boldsymbol{\beta}_p), \quad (4.1)$$

where  $A_p$  is the axes-kept linear transformation defined in (3.1). Then the projection method estimator of  $\boldsymbol{\theta}$  given by  $\hat{\boldsymbol{\theta}}_Q(\mathbf{p}) = \exp A_p^{-1} D\hat{\boldsymbol{\beta}}_{p,Q}$  does not depend on  $\mathbf{p}$ . Thus the linear projection estimator is invariant with respect to the selection of a base population.

A proper selection of the weight matrix  $Q$  is  $\hat{V}_y^{-1}$ , the inverse of the estimated covariance matrix of  $\mathbf{y}$ , if it is available. This weight matrix minimizes the asymptotic covariance matrix of  $\hat{\beta}_Q$  with respect to  $Q$ . Another simple choice is  $Q = E_h$ , the  $h \times h$  identity matrix, which is associated to orthogonal linear projection. By the corollary both weight matrices lead to invariant estimators.

In the following, we present two examples of linear projection estimators.

### Example 1

Consider the following multiple population proportional hazards model in survival analysis

$$\lambda_{i'}(t) = \theta_{i'}^{(i)} \lambda_i(t), \quad i, i' = 0, 1, \dots, I,$$

where  $\lambda_i(t)$  is the unknown hazard function of population  $i$  and  $\theta_{i'}^{(i)} \in (0, \infty)$  is the unknown hazard rate of population  $i'$  with respect to population  $i$ , which is assumed independent of time  $t$ . Suppose that the Begun-Reid estimator of  $\theta_{i'}^{(i)}$  is given from population  $i$  and  $i'$ ,  $0 \leq i < i' \leq I$ , and selected for  $b_{i'}^{(i)}$  involved in (2.6) (Begun and Reid 1983), we get a generalized Begun-Reid estimator of the hazard rate vector. Zhang and Yanagawa (1997) investigated the asymptotic distribution of the estimator and given an estimator of its covariance matrix. It has been shown by simulations that the estimator  $\hat{\beta}_{\hat{V}_y^{-1}}$  (i.e. the estimator taking  $\hat{V}_y^{-1}$  as  $Q$ ) may have smaller bias and covariance matrix than the Cox partial likelihood estimator for small and medium sample sizes.

### Example 2

The Mantel-Haenszel type estimators (1959) of relative risk have been widely used in various studies (Greenland 1985, Sato 1992, Yamagimoto 1992, Breslow 1996 and Austin *et al.* 1997). These estimators are for  $2 \times 2$  tables in stratified data, easy to compute and have simple variance estimators (Robins *et al.* 1986), and furthermore, they are dually consistent and highly efficient. Suppose that the exposure has  $I + 1$  levels, and that the relative risk of level  $i'$  with respect to level  $i$ , denoted by  $\theta_{i'}^{(i)}$ , is common throughout strata,  $i, i' = 0, 1, \dots, I$ . Let  $\{b_{i'}^{(i)}\}$  be one of pairwise Mantel-Haenszel type estimators of  $\{\theta_{i'}^{(i)}\}$ . Taking  $\rho$  as  $L_{2,Q}$  in (2.5), we may obtain generalized Mantel-Haenszel estimators. In particular, for  $L_{2,E}$ , i.e.  $Q = E$  in (2.5), we get

$$\hat{\beta}_{MH} = \frac{1}{I+1} \left[ \sum_{i=1}^I \log b_i^{(0)} \mathbf{1}_I - \left( \sum_{i \neq 1} \log b_i^{(1)}, \dots, \sum_{i \neq I} \log b_i^{(I)} \right)^\tau \right]$$

by using the relationship  $(D^\tau D)^{-1} = (E_I + \mathbf{1}_I \mathbf{1}_I^\tau)/(I+1)$ . The resulted estimators inherit all characteristics of the pairwise Mantel-Haenszel type estimators mentioned above (Yanagawa and Fujii 1995).

If the size of each entry in a contingency table is large, the optimum Begun-Reid estimator (that is, the estimator with estimated optimum score, see Zhang and Yanagawa 1997) might be applicable. The efficiency of the resulted estimator would be slightly high than above generalized Mantel-Haenszel estimator, but it is not consistent for sparse data.



## 4.2. Ridge estimators

Adding a penalty term to the weighted sum of squared errors, we may consider the estimator defined by

$$\hat{\beta}_R = \arg \min_{\beta} \left[ (\mathbf{y} - D\beta)^T \hat{V}_y^{-1} (\mathbf{y} - D\beta) + K(D\beta)^T D\beta \right],$$

where  $K$  is a given positive number, and  $\hat{V}_y$  is an estimator of asymptotic covariance matrix of  $\mathbf{y}$  satisfying  $\sqrt{n}(\hat{V}_y - V_y) = O_p(1)$ , and the other notations are the same as in Section 2. It is easy to show that the estimator is given by

$$\hat{\beta}_R = [D^T \hat{V}_y^{-1} D + K D^T D]^{-1} D^T \hat{V}_y^{-1} \mathbf{y}.$$

This kind of estimators is well known as ridge estimators. The estimator  $\hat{\beta}_R$  with  $K = I/(n\mathbf{y}^T \mathbf{y})$  (Ryan 1997) has the same asymptotic distribution as  $\hat{\beta}_{\hat{V}_y^{-1}}$ . It is not difficult to show the corresponding result as the corollary; thus the estimator of  $\theta$  derived by  $\hat{\beta}_R$  is invariant with respect to the selection of a base population.

## References

- Andersen, P.K., Borgan,  $\phi.$ , Gill, R.D. and Keiding, N.(1993), *Statistical models based on counting processes*. New York: Springer-Verlag.
- Begun, J.M. and Reid, N.(1983), "Estimating the relative risk with censored data". *Journal of the American Statistical Association*, 78, 337-341.
- Breslow, N.E.(1996), "Statistics in epidemiology: the case-control study". *Journal of the American Statistical Association*, 91, No.433, 14-28.
- Breslow, N.E., Day, N.E., Halvorsen, K.T. and Prentice, R.L. (1978), "Estimation of multiple relative risk functions in matched case-control studies". *American Journal of Epidemiology*, 108, 299-307.
- Cox, D.R.(1975), "Partial likelihood". *Biometrika* 62, 2, 269-276.
- Crowley, J., Liu, P.Y. and Veolkel, J.G.(1982), "Estimation of the ratio of hazard functions". In *Survival Analysis: IMS Lecture Notes-Monograph Series, Vol.2*, J. Crowley and R.A.Johnson(eds), Hayward, California: The Institute of Mathematical Statistics, pp.56-73.
- Greenland, S.(1986) "Adjustment of risk ratios in case-base studies(hybrid epidemiological designs)". *Statistics in Medicine*, 5, 579-584.
- Liang, K.Y.(1987), "Extended Mantel-Haenszel estimating procedure for multivariate logistic regression models". *Biometrics*, 43, 289-299.
- Mantel, N. and Haenszel, W.(1959), "Statistical aspects of the analysis of data from retrospective studies of disease". *Journal of the National Cancer Institute*, 22, 719-748.

- Robins, J., Breslow, N. and Greenland, S.(1986), "Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models". *Biometrics*, 42, 311-323.
- Ryan, T.P.(1997), *Modern Regression Methods*. New York: John Wiley and Sons.
- Sato, T.(1991), "Elementary methods of cohort analysis with several exposure levels". *Biometrics*, 47, 1165-1170.
- Sato, T.(1992), "Estimation of a common risk ratio in stratified case-cohort studies". *Statistics in Medicine*, 11, 1599-1605.
- Yanagawa, T. and Fujii, Y.(1995), "Projection-method Mantel-Haenszel estimator for  $K \times J$  tables". *Journal of the American Statistical Association*, 90, 649-656.
- Yanagimoto, T.(1992), "The Mantel-Haenszel statistics for the extended odds ratio in the negative binomial distribution". *Journal of the Japan Statistical Society*, 22(1), 7-17.
- Zhang, Z.Z. and Yanagawa, T.(1997), "Generalized Begun-Reid estimators for multiple hazard rates". To be published.

*Received December 27, 1997*