97-F-32

Two Transformation Models and Rank Estimation

Hideatsu Tsukahara
University of Tokyo

October 1997

Two transformation models and Rank estimation

Hideatsu Tsukahara *

Abstract

In the first half of the paper, we shall investigate the relation of two models both of which have been called a transformation model. One model is in terms of distribution functions and the other is in terms of random variables. We shall show that the former class is larger than the latter and we give an explicit relation between these models. The second half deals with estimation procedures for the regression parameters in the transformation model in terms of distribution functions. After reviewing and extending previously proposed estimators for the model, we derive a new estimator based on ranks. Monte Carlo simulation is performed to compare the empirical properties of several estimators for the Cox model, which is a particular case of our transformation model.

* Auther's affiliation: Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

1. Introduction

In the literature the following two models have been studied under the name of "transformation model". One is expressed in terms of distribution function(df):

$$(1.1) \qquad X \sim G_\theta = D(F(\cdot);\theta),$$

where $D(\,\cdot\,;\theta)$ is a known continuous df on $(0,1)$ and F is an arbitrary baseline df. θ is a parameter whose values are in some parameter space $\Theta \subset \mathbb{R}$. This model for the two-sample problem is studied in Dabrowska, Doksum and Miura(1989) and, with censored data, in Tsukahara(1991). The above two papers considered semiparametric estimation of θ based on ranks.

The other is expressed in terms of random variable(rv):

$$(1.2) \qquad h(X) = \nu(\theta) + \epsilon,$$

where h is an unknown strictly monotone increasing function and ϵ is distributed according to Ψ which is a known df. $\nu(\theta)$ is a function of θ such as $\log\theta$ and is often connected by the linear model $\nu(\theta) = \boldsymbol{\beta}'\boldsymbol{z}$. With this linear form, various methods of estimation of regression parameters $\boldsymbol{\beta}$ are suggested. See Dabrowska and Doksum(1988b), Doksum(1987), Pettitt(1982,1983,1987) and Cuzick(1988).

We call (1.1) *Lehmann alternative transformation* (abbreviated LAT) model and (1.2) *location transformation* (LOT) model. The same word "transformation" is used, but in LOT model, transformation acts on the sample space, while in LAT model it acts on the space of probability distributions. So these models are essentially different in this point. In Section 2, we consider the problem of relating these two models. It will be proved that a LOT model is always rewritten as a LAT model. The family of the models of this kind includes several important models such as the proportional hazards model. Moreover we derive a necessary condition and a sufficient condition for a given LAT model to be re-expressible as a LOT model. Using this condition, it is shown that some of LAT models cannot be re-expressible as LOT models.

In Section 3, we give a short review of previous work for LOT model and extend those methods to LAT model framework. And we derive an estimator of regression parameter for LAT model (1.1) with the form $D(F;\psi(\boldsymbol{\beta},\boldsymbol{z}))$. In the model which satisfies the sufficient condition mentioned above, this model includes a linear connection $\nu(\theta)=\boldsymbol{\beta}'\boldsymbol{z}$ in (1.2). Our estimator is a generalization of the rank approximate M(RAM) estimator obtained in Dabrowska, Doksum and Miura(1989) for two-sample problem and in Cuzick(1988) for LOT model.

2. Two models

First we state the precise definitions and assumptions of the models we consider in this section.

(A.1) For LAT model (1.1), assume that $D(t;\theta)$ is a df on (0,1) for each $\theta\in\Theta\subset\mathbb{R}$ and is strictly increasing in t , strictly decreasing in θ and continuous in both variables. Also assume that F is an arbitrary strictly increasing df on $A\subset\mathbb{R}$.

(A.2) For LOT model (1.2), assume that h is an unknown strictly increasing function on A , and Ψ is a known strictly increasing df on the whole real line.

Next we define the following proportional model for later use:

$$\alpha(G(x))=\frac{1}{\theta}\alpha(F(x)), \tag{2.1}$$

where F and G are df's. α is a quantity determined by df such as cumulative hazard and odds ratio.

(A.3) For the proportional model (2.1), assume that α is a strictly increasing function on $[0,1)$ into $[0,\infty)$ which satisfies $\alpha(0)=0$ and $\lim_{t\to 1}\alpha(t)=\infty$.

We now show that a LOT model is always reexpressible as a LAT model. For a given LOT model (1.2), set

$$D(t;\theta)=\Psi(\Psi^{-1}(t)-\nu(\theta)).$$

Then, if $X \sim D(F(\cdot);\theta)$, we find that

$$\Psi^{-1}(F(X))=\nu(\theta)+\epsilon,$$

where ϵ has a df Ψ . Since F is an arbitrary df $\Psi^{-1}(F(x))$ is an unknown monotone increasing function, so that we can regard this as h . This proves that a LOT model is always reexpressible as a LAT model.

Here we illustrate some examples.

Example 1. (proportional hazards model) The proportional hazards model is expressed in the form of a LAT model by taking $D(t;\theta)=1-(1-t)^{1/\theta}$. Thus, according to the above argument, we see that this model is rewritten as

$$\log\Lambda_F(X)=\log\theta+\epsilon,$$

where Λ_F is a cumulative hazard function corresponding to F and ϵ has the extreme value distribution whose df is $1-\exp\{-e^x\}$. This model was introduced in Cox(1972). It is shown that the maximum partial likelihood estimates is best for estimating θ for two-sample problem in the nonparametric sense in Begun and Wellner(1983). See also Begun(1987) and Begun and Reid(1983) for another construction of the best estimate.

Example 2. (proportional odds model) Taking $D(t;\theta)=t[(1-t)\theta+t]^{-1}$ in LAT model yields the proportional odds model and rewritten as

$$\log\frac{F(X)}{1-F(X)}=\log\theta+\epsilon,$$

where ϵ has the logistic distribution whose df is $1/(1+e^{-x})$. This model was introduced and studied in McCullagh(1980,1984) for ordinal data. Bennett(1983) and Pettitt(1984) studied this model in the survival analysis context.

Example 3. (proportional γ -odds model) Taking

$$D(t;\theta) = \begin{cases} 1 - \left[\frac{\theta(1-t)^\gamma}{1-(1-t)^\gamma+\theta(1-t)^\gamma}\right]^{1/\gamma} & \text{if } \gamma > 0 \\ 1-(1-t)^\theta & \text{if } \gamma = 0 \end{cases},$$

in LAT model yields the proportional γ -odds model and rewritten as

$$\log \frac{(1-F(X))^{-\gamma} - 1}{\gamma} = \log\theta + \epsilon,$$

where e^ϵ has the Pareto distribution whose df is $1-(1+\gamma x)^{-1/\gamma}$. Clayton and Cuzick(1986) investigated this model for regression problems, and Dabrowska and Doksum(1988a) for the two-sample problem.

Remark. These three examples are typical cases of the proportional model (2.1); for instance, $\alpha(t) = -\log(1-t)$ gives the proportional hazards model and $\alpha(t) = t/(1-t)$ gives the proportional odds model.

Now let a proportional model (2.1) be given. If for any df F we define

$$h(x) = \log\alpha(F(x)),$$

then letting $X \sim G$, we obtain

$$h(X) = \log\theta + \epsilon, \tag{2.2}$$

where $\epsilon \sim \Psi$ and Ψ is determined by the relation $\Psi^{-1}(t) = \log\alpha(t)$. This is a special case of a LOT model with $\nu(\theta) = \log\theta$.

Conversely, if (1.4) is given, then defining $\alpha(t)$ by

$$\alpha(t) = \exp[\Psi^{-1}(t)]$$

yields the proportional model which is equivalent to (1.4). Moreover this becomes a form of LAT model;

$$\log \alpha(G(x)) = \log \alpha(F(x)) - \log \theta$$

$$\Psi^{-1}(G(x)) = \Psi^{-1}(F(x)) - \log \theta,$$

so that we have

$$G(x) = \Psi\left[\Psi^{-1}(F(x)) - \log \theta\right]. \tag{2.3}$$

This is called the transformation shift model in Dabrowska, Doksum and Miura(1989), and may be thought of as an extension of the power transformation model in Box and Cox(1964). Hence we see that the proportional model (2.1), the special LOT model (2.2) and the transformation shift model (2.3) are all equivalent.

Next we consider a necessary condition for a LAT model (1.1) to be re-expressible as a LOT model. Let a LAT model be given, that is, a transformation $D(\,\cdot\,;\theta)$ in (1.1) is given. The derivatives of the functions ν, F, D, Ψ and h with respect to the appropriate arguments are assumed to exist in what follows.

Proposition 1. *A necessary condition for a given LAT model (1.1) to be re-expressible as a LOT model (1.2) is that there exist a nonnegative function* $g(t)$ *on* $[0,1]$ *and a function* $\eta(\theta)$ *such that*

$$J(t;\theta) \triangleq \frac{d(t;\theta)}{\dot{D}(t;\theta)} = g(t)\cdot\eta(\theta)$$

Proof. If (1.1) can be rewritten in the form of (1.2), we must have

$$D(F(x);\theta) = \Psi(h(x) - \nu(\theta)),$$

and so $D(t;\theta) = \Psi(\alpha(t) - \nu(\theta))$, where $\alpha(t) = h(F^{-1}(t))$. Taking derivatives with respect to t and θ respectively and taking their ratio,

$$\frac{d(t;\theta)}{\dot{D}(t;\theta)} = \alpha'(t) \cdot [-\dot{\nu}(\theta)]^{-1}$$

where $\alpha'(t) = (d/dt)\alpha(t)$ and $\dot{\nu}(\theta) = (d/d\theta)\nu(\theta)$. Since $\alpha(t)$ is nondecreasing, $\alpha'(t)$ is nonnegative. ∎

This condition enables us to show that the examples below cannot be re-expressible as a LOT model.

Example 4. Let $D(t;\theta) = \theta t + (1-\theta)t^2$, $\theta \in [0,1]$. Then we have

$$J(t;\theta) = \frac{\theta + 2(1-\theta)t}{t(1-t)},$$

and we see that this model cannot be rewritten as a LOT model. This Lehmann alternative was studied in Lehmann(1952).

Example 5. Set $D(t;\theta) = (e^{\theta t} - 1)/(e^{\theta} - 1)$, $\theta \in [0,\infty)$. Then

$$J(t;\theta) = \frac{\theta e^{\theta t}(e^{\theta} - 1)}{te^{\theta t}(e^{\theta} - 1) - (e^{\theta t} - 1)e^{\theta}}.$$

This was considered in Ferguson(1967).

The above two examples, as alternatives for the two-sample problem, are Lehmann alternatives for which locally most powerful rank test is Wilcoxon.

Example 6. If we set $D(t;\theta) = \sum c_i(\theta)t^i$, then

$$J(t;\theta) = \frac{\sum_i c_i(\theta)t^{i-1}}{\sum \dot{c}_i(\theta)t^i},$$

where $c_i(\theta)$ satisfies $\sum c_i(\theta) = 1$. This model was considered in Miura(1985) and may be regarded as mixtures of extremals. Taking $c_i(\theta)$ from Bin $(1,\theta)$ yields Example 4, and taking $c_i(\theta)$ from positive Poisson distribution, i.e.,

$$c_i(\theta) = \frac{\theta^i}{(e^\theta - 1)i!}, \qquad \theta > 0,$$

we obtain Example 5. Also if we take $c_i(\theta) = (1-\theta)^{i-1}\theta$, i.e, geometric distribution, then by easy calculation we get the proportional odds model in Example 2.

Finally we give a sufficient condition by using the translation equation in the functional equation theory (see Aczél(1966)).

Proposition 2. *Assume that (A.1) and (A.2) hold and put* $\nu(\theta) = \log\theta$, $\theta \in (0,\infty)$. *Then a sufficient condition for a given LAT model to be reexpressible as a LOT model (2.2) is that* $D(t;\theta)$ *satisfies*

(2.4) $$D[D(t;\theta_1);\theta_2] = D(t;\theta_1\theta_2), \quad t \in (0,1), \quad \theta_1,\theta_2 \in (0,\infty).$$

Proof. For any t_0 , put

$$\Psi(x) = \tilde{D}(t_0;x), \quad x \in \mathbb{R}, \qquad \tilde{D}(t;\mu) = D(t;\theta) = D(t;e^{-\mu}),$$

where $\mu = -\log\theta$. Note that $\tilde{D}$ satisfies

$$\tilde{D}[\tilde{D}(t;\mu_1);\mu_2] = D[D(t;e^{-\mu_1});e^{-\mu_2}] = D(t;e^{-(\mu_1+\mu_2)}) = \tilde{D}(t;\mu_1+\mu_2)$$

because of (2.4). It then follows that

$$\Psi(x+\mu) = \tilde{D}(t_0; x+\mu) = \tilde{D}(\tilde{D}(t_0; x); \mu) = \tilde{D}(\Psi(x); \mu)$$

Thus $\tilde{D}(t;\mu) = \Psi(\Psi^{-1}(t)+\mu)$, which is equivalent to $D(t;\theta) = \Psi(\Psi^{-1}(t) - \log\theta)$, because Ψ is strictly increasing and continuous by (A.1). It remains to show that $\lim_{x\to-\infty}\Psi(x) = 0$ and $\lim_{x\to\infty}\Psi(x) = 1$. Let us show the latter statement. Suppose, to the contrary, that $\lim_{x\to\infty}\Psi(x) = u \in (0,1)$. Then there would exist μ_1 and μ_2 such that $\tilde{D}(u;\mu_1) \neq \tilde{D}(u;\mu_2)$, and $\tilde{D}(t;\mu)$ being continuous in t,

$$\begin{aligned} u = \lim_{x\to\infty}\Psi(x+\mu_1) &= \lim \tilde{D}(\Psi(x);\mu_1) = \tilde{D}(u;\mu_1) \\ &\neq \tilde{D}(u;\mu_2) = \lim \tilde{D}(\Psi(x);\mu_2) = \lim \Psi(x+\mu_2) = u, \end{aligned}$$

which is a contradiction. $\lim_{x\to-\infty}\Psi(x) = 0$ may be proved similarly. ∎

Remarks.

(i) This proof shows that Ψ is not uniquely determined by D satisfying (2.4).

(ii) If a LAT model is reexpressible as a LOT model (2.2), then for some strictly increasing h and df Ψ, we have

$$D(F(x);\theta) = \Psi(h(x) - \log\theta).$$

If we can change baseline df F appropriately, we may write $h(x) = \Psi^{-1}(F(x))$, and then we have $D(t;\theta) = \Psi(\Psi^{-1}(t) - \log\theta)$. This clearly satisfies (2.4). Thus the condition (2.4) is also necessary in the sense above. We call a LAT model with the property (2.4) a *multiplicative* LAT model.

From the above argument, we now see that the multiplicative LAT model, the transformation shift model (2.3), the LOT model (2.2) and the proportional model (2.1) are all equivalent. A diagram summarizing the relations between those models are given below.

Figure 1. Relations between models

Also we note that LAT model is the most general among the models discussed above. Thus it is natural to consider the inference based on LAT models. In the next section, we consider the estimation of a parameter in general regression problems for the LAT model.

3. Estimation based on Ranks

In this section we consider the following model: the X_i, $i = 1, \cdots, n$ are independent random variables, each of which has df $G_i(x) = D(F(x); \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i))$ where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ is a regression parameter and $\boldsymbol{z}_i = (z_{i1}, \cdots, z_{ip})'$ is a vector of covariates. We assume that this LAT model satisfies the condition (A.1) and

(A.4) There exists a $\theta_0 \in \Theta$ such that $D(t; \theta_0) = t$ for all $t \in (0, 1)$,

and also assume that the regression function $\lambda(\boldsymbol{\beta}, \boldsymbol{z})$ takes values in Θ and satisfies $\lambda(\mathbf{0}, \boldsymbol{z}) = \theta_0$. We sometimes denote $\theta_i = \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i)$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)'$ for simplicity.

The most popular form of λ is $\lambda(\boldsymbol{\beta}, \boldsymbol{z}) = \exp\{\boldsymbol{\beta}'\boldsymbol{z}\}$, which gives a LOT model $h(X) = \boldsymbol{\beta}'\boldsymbol{z} + \epsilon$ if D is multiplicative.

3.1. Previous methods.

We give extensions of the estimation methods proposed in the literature. The first two methods are based on the rank likelihood. Define $R_i \triangleq \text{rank}(X_i) = \text{rank}(F(X_i))$ and $\mathbf{R} = (R_1, \cdots, R_n)'$. Also let $\mathbf{r} = (r_1, \cdots, r_n)'$ be the vector of observed ranks. Then the rank likelihood is given by

$$L_{\mathbf{r}}(\boldsymbol{\theta}) \triangleq \mathrm{P}[\mathbf{R} = \mathbf{r}].$$

It follows from Hoeffding's formula (Hoeffding(1951)) that

$$L_{\mathbf{r}}(\boldsymbol{\theta}) = \frac{1}{n!}\mathrm{E}\left[\prod_{i=1}^{n} d(V_{(r_i)}; \theta_i)\right], \tag{3.1}$$

where $V_{(1)} < \cdots < V_{(n)}$ are the order statistics of a sample of size n from the uniform distribution on $(0,1)$. In most cases this cannot be evaluated explicitly [a notable exception is the proportional hazards model], so that some kinds of approximation have been proposed.

(i)*Pettitt(1982)'s quadratic approximation.* By Taylor series expansion about $\boldsymbol{\beta} = \mathbf{0}$ up to quadratic term (note that $\lambda(\mathbf{0}, \boldsymbol{z}) = \theta_0$), we have

$$\begin{aligned}\log d(v; \lambda(\boldsymbol{\beta}, \boldsymbol{z})) &\approx \log d(v; \theta_0) + \varphi(v; \theta_0)\dot{\lambda}(\mathbf{0}, \boldsymbol{z})\boldsymbol{\beta} \\ &\qquad + \frac{1}{2}\dot{\varphi}(v; \theta_0)\boldsymbol{\beta}'[\dot{\lambda}(\mathbf{0}, \boldsymbol{z})'\dot{\lambda}(\mathbf{0}, \boldsymbol{z})]\boldsymbol{\beta} + \frac{1}{2}\varphi(v; \theta_0)\boldsymbol{\beta}'\ddot{\lambda}(\mathbf{0}, \boldsymbol{z})\boldsymbol{\beta}, \\ &= \varphi(v; \theta_0)\left[\dot{\lambda}(\mathbf{0}, \boldsymbol{z})\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}'\ddot{\lambda}(\mathbf{0}, \boldsymbol{z})\boldsymbol{\beta}\right] + \frac{1}{2}\dot{\varphi}(v; \theta_0)\boldsymbol{\beta}'[\dot{\lambda}(\mathbf{0}, \boldsymbol{z})'\dot{\lambda}(\mathbf{0}, \boldsymbol{z})]\boldsymbol{\beta},\end{aligned}$$

where $\dot{\lambda}(\boldsymbol{\beta}, \boldsymbol{z}) = (\partial/\partial\boldsymbol{\beta})\lambda(\boldsymbol{\beta}, \boldsymbol{z})$ is a $1 \times p$ vector, $\ddot{\lambda}(\boldsymbol{\beta}, \boldsymbol{z}) = (\partial^2/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}')\lambda(\boldsymbol{\beta}, \boldsymbol{z})$ is a $p \times p$ matrix, and

$$\varphi(v; \theta) = \frac{\dot{d}(v; \theta)}{d(v; \theta)}, \qquad \dot{d}(v; \theta) = \frac{\partial}{\partial\theta}d(v; \theta), \qquad \dot{\varphi}(v; \theta) = \frac{\partial}{\partial\theta}\varphi(v; \theta).$$

Here we assume that all required derivatives exist. Furthermore let

$$\boldsymbol{\xi} = (\varphi(V_{(r_1)};\theta_0),\cdots,\varphi(V_{(r_n)};\theta_0))', \qquad \boldsymbol{\eta} = \text{diag}\{\dot\varphi(V_{(r_1)};\theta_0),\cdots,\dot\varphi(V_{(r_n)};\theta_0)\},$$

and Λ be a $n \times p$ matrix whose i th row is given by $\dot\lambda(\mathbf{0}, \boldsymbol{z}_i)$. Then by (3.1), we can write

$$L_{\mathbf{r}}(\boldsymbol{\beta}) \approx \frac{1}{n!}\text{E}\left[\exp\left\{\boldsymbol{\beta}'\Lambda'\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{\beta}'\left(\sum_{i=1}^{n}\xi_i\ddot\lambda(\mathbf{0},\boldsymbol{z}_i)\right)\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}'\Lambda'\boldsymbol{\eta}\Lambda\boldsymbol{\beta}\right\}\right],$$

ignoring the terms more than quadratic in $\boldsymbol{\beta}$. Assuming that $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are approximately multivariate normal, we find

$$L_{\mathbf{r}}(\boldsymbol{\beta}) \approx \frac{1}{n!}\exp\left[\boldsymbol{\beta}'\Lambda'\boldsymbol{a} + \frac{1}{2}\boldsymbol{\beta}'\left(\Lambda'C\Lambda + \Gamma\right)\boldsymbol{\beta}\right],$$

where $\boldsymbol{a} = \text{E}(\boldsymbol{\xi})$, $A = \text{Cov}(\boldsymbol{\xi})$, $B = \text{E}(\boldsymbol{\eta})$, $C = A + B$ and $\Gamma = \sum_{i=1}^{n} a_i\ddot\lambda(\mathbf{0},\boldsymbol{z}_i)$. This expression suggests the estimate

$$\hat{\boldsymbol{\beta}}_Q = [-\Lambda'C\Lambda + \Gamma]^{-1}\Lambda'\boldsymbol{a}$$

provided that $\Lambda'C\Lambda + \Gamma$ is nonsingular. This is essentially the first step approximate estimate of the Newton-Raphson maximization procedure.

(ii)*Doksum(1987)'s likelihood sampler.* The expectation in (3.1) can be estimated by

$$\hat{L}_M(\boldsymbol{\beta}) = \frac{1}{M}\sum_{k=1}^{M}\prod_{i=1}^{n} d(V_{(r_i)}^k;\theta_i),$$

where $V_{(r_1)}^k < \cdots < V_{(r_n)}^k$ are independent orderd samples generated on the computer ($k = 1,\cdots,M$), each of which is the order statistics of a sample of size n from the uniform distribution on $(0,1)$. Then the procedure is to maximize $\hat{L}_M(\boldsymbol{\beta})$ for $M = 100, 200, \cdots$ and to stop when the change in the resulting estimates $\hat{\boldsymbol{\beta}}_M$, $M = 100, 200, \cdots$ from one M to the next is within prescribed precision.

The estimates $\hat{\boldsymbol{\beta}}_Q$ and $\hat{\boldsymbol{\beta}}_M$ are asymptoticaly normal for $\boldsymbol{\beta}$ in neighborhood of $\boldsymbol{\beta} = \mathbf{0}$, but are not consistent for fixed $\boldsymbol{\beta} \neq \mathbf{0}$. This is natural because the approximation is local (near $\boldsymbol{\beta} = \mathbf{0}$).

(iii)*Pettitt(1987)'s least rank mean square.* Let

$$S(\boldsymbol{\beta}) \triangleq \sum_{i=1}^{n} [r_i - \mathrm{E}(R_i)]^2 ,$$

and define the estimate $\hat{\boldsymbol{\beta}}_E$ as the solution to minimizing $S(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. For our model, we have

$$\mathrm{E}(R_i) = 1 + \sum_{j \neq i} \mathrm{P}[X_i - X_j > 0] = 1 + \sum_{j \neq i} \int_0^1 D(t;\theta_j) dD(t;\theta_i),$$

and if we assume that D is multiplicative, this reduces to

$$1 + \sum_{j \neq i} \int_0^1 t dD\left(t; \frac{\theta_i}{\theta_j}\right) .$$

For the proportional hazards and proportional odds model, we easily get

$$\mathrm{E}(R_i) = 1 + \sum_{j \neq i} \frac{\theta_j}{\theta_i + \theta_j}, \qquad \text{(proportional hazards)},$$
$$\mathrm{E}(R_i) = 1 + \sum_{j \neq i} \frac{\theta_i \theta_j}{(\theta_j - \theta_i)^2} \left[\frac{\theta_i}{\theta_j} - 1 - \log \frac{\theta_i}{\theta_j}\right] \qquad \text{(proportional odds)}.$$

By the Monte Carlo simulation, Pettitt reports that $\hat{\boldsymbol{\beta}}_E$ has some good properties even for $\boldsymbol{\beta}$ not near $\mathbf{0}$, but its theoretical properties are not clear.

Clayton and Cuzick(1985,1986) developed an approximate MLE by applying the EM algorithm (Dempster, Laird and Rubin(1977)). The method involves very complicated computation, and the asymptotic properties of the estimator seem difficult to establish. We also expect that it is possible to use the rank inversion idea, which is basically based on Hodges and Lehmann(1963), along the line in Dabrowska *et al.*(1989) and Miura and Tsukahara(1992) and to get an estimate of $\boldsymbol{\beta}$.

3.2. Rank approximate M estimate.

In this subsection, we use the idea of Dabrowska *et al.*(1989) and Cuzick(1988) and derive a *RAM (Rank Approximate M)* estimate for $\boldsymbol{\beta}$.

Suppose first that F is known and consider the full likelihood $L(\boldsymbol{\beta})$ of the X_i . Let $g_i(x_i)$ be the density of X_i . It follows that $g_i(x_i) = d(F(x_i); \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i))f(x_i)$, $L(\boldsymbol{\beta}) = \prod_{i=1}^n g_i(x_i)$ and $l(\boldsymbol{\beta}) \triangleq \log L(\boldsymbol{\beta}) = \sum_{i=1}^n \log g_i(x_i)$. Hence

$$\dot{l}(\boldsymbol{\beta}) \triangleq \frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{\dot{g}_i(x_i)}{g_i(x_i)}$$

where $\dot{g}_i(x_i) = (\partial/\partial\boldsymbol{\beta})g_i(x_i) = \dot{d}(F(x_i); \lambda(\boldsymbol{\beta}; \boldsymbol{z}_i))\dot{\lambda}(\boldsymbol{\beta}; \boldsymbol{z}_i)f(x_i)$. The likelihood equation is then given by

$$\sum_{i=1}^{n} \dot{\lambda}(\boldsymbol{\beta}, \boldsymbol{z}_i) \frac{\dot{d}(F(x_i); \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i))}{d(F(x_i); \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i))} = \mathbf{0}.$$

Replacing $\dot{d}/d$ by any estimating function ϕ satisfying $\mathrm{E}_\beta[\phi(V_i; \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i))] = 0$ where $V_i \sim D(\ \cdot\ ; \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i))$, an *M*-estimate (Huber(1981)) is defined by the solution to

$$\sum_{i=1}^{n} \dot{\lambda}(\boldsymbol{\beta}, \boldsymbol{z}_i)\phi(F(x_i); \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i)) = \mathbf{0}.$$

In the case of unknown F , we shall replace F by its estimate $\mathbb{F}_n^\beta$ given as follows: define

$$\bar{\mathbb{G}}_n(x) \triangleq \frac{1}{n+1} \sum_{i=1}^{n} I_{[X_i \le x]}, \quad \bar{G}_\beta(x) \triangleq \frac{1}{n} \sum_{i=1}^{n} D(F(x); \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i)),$$

$$\bar{D}_\beta(t) \triangleq \frac{1}{n} \sum_{i=1}^{n} D(t; \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i)).$$

Then we have $\bar{G}_\beta(x) = \bar{D}_\beta(F(x))$. This indicates that F may be estimated by

$$\mathbb{F}_n^\beta(x) \triangleq \bar{D}_\beta^{-1}(\bar{\mathbb{G}}_n(x)).$$

Set $\hat{V}_i \triangleq \mathbb{F}_n^\beta(X_i) = \bar{D}_\beta^{-1}(R_i/(n+1))$. Then a RAM estimate $\hat{\boldsymbol{\beta}}_{RAM}$ is defined by the solution to

$$\sum_{i=1}^{n} \dot{\lambda}(\boldsymbol{\beta}, \boldsymbol{z}_i)\phi(\hat{V}_i; \lambda(\boldsymbol{\beta}, \boldsymbol{z}_i)) = \mathbf{0}.$$

Note that $\bar{\mathbb{G}}_n(X_i) = 1/(n+1)R_i$, so that the estimate may be viewed as an M-estimate based on ranks.

4. Monte Carlo Results.

We consider the proportional hazards model with $n = 25$. The regression function $\lambda(\boldsymbol{\beta}, \boldsymbol{z})$ is $\exp\{\boldsymbol{\beta}'\boldsymbol{z}\}$, which yields the Cox's regression model for survival data. To simplify the situation, we will take the simple regression case without intercept, i.e., $\boldsymbol{\beta}'\boldsymbol{z} = \beta z_i$. Then, to obtain the formula for $\hat{\beta}_Q$ we have $\dot{\lambda}(\mathbf{0}, \boldsymbol{z}_i) = z_i$, $\ddot{\lambda}(\mathbf{0}, \boldsymbol{z}_i) = z_i^2$, so that $\Lambda = (z_1, \cdots, z_n)'$. After some calculations we get

$$a_j = \sum_{k=1}^{r_j} \frac{1}{n-k+1} - 1, \quad \Lambda' C \Lambda + \Gamma = \Lambda'(A - B)\Lambda,$$

where $B = \text{diag}\{a_1+1, \cdots, a_n+1\}$ and the (i,j) element of A is given by $\sum_{k=1}^{r_i \wedge r_j} 1/(n-k+1)^2$. To calculate $\hat{\beta}_M$, we only need to know $d(t;\theta) = \theta^{-1}(1-t)^{1/\theta - 1}$. We fix $M = 100$ for simplicity of the procedure, which, however, may make the performance of the estimate. For $\hat{\beta}_E$, as already mentioned before,

$$\mathrm{E}(R_i) = 1 + \sum_{j \neq i} \frac{e^{\beta z_i}}{e^{\beta z_i} + e^{\beta z_j}} = \frac{1}{2} + e^{\beta z_i} \sum_{j=1}^{n} \frac{1}{e^{\beta z_i} + e^{\beta z_j}}.$$

The number of Monte Carlo runs is 200, $z_i = (i-13)/12$, and $\beta = 0,0.5,1,1.5,2,3,4,5$. We generated uniform random numbers and transformed them by $D^{-1}(u; e^{\beta z_i}) = 1-(1-u)^{e^{\beta z_i}}$ to get the Monte Carlo samples $V_1, \cdots, V_n$, where V_i has df $D(u; e^{\beta z_i})$. Then $F^{-1}(V_i)$ has df $G_i = D(F; e^{\beta z_i})$, but we do not have to specify F because we use only ranks of the observations, i.e. $\text{rank}(V_i) = \text{rank}(F^{-1}(V_i))$ (assuming that F is continuous).

Table 1 shows the Monte Carlo results, presenting the bias and MSE of each estimate. $\hat{\beta}_{MPL}$ is the maximum partial likelihood estimate, which is known to be the asymptotically best nonparametric estimate in this case (Splus function `coxreg` provides this estimate). Also the histograms of each estimate based on 200 simulations for each value of β are displayed at the end. When β is near zero, all estimates have similar nice performance. But for large values of β , $\hat{\beta}_Q$, $\hat{\beta}_{RAM}$ and $\hat{\beta}_M$ have large negative bias. It seems natural for β , $\hat{\beta}_Q$, $\hat{\beta}_M$ because these methods are based on local approximation to $L_r(\beta)$ near $\beta = 0$. On the other hand, although the RAM estimate is not derived from local approximation to the rank likelihood, its behavior is not good for large values of β in this experiment. We are urged to find out the reason for this bad behavior. $\hat{\beta}_E$ looks better than $\hat{\beta}_Q$ and $\hat{\beta}_M$ since it somehow follows up the large values of β . However, its behavior is less stable than the others in terms of dispersion as will be seen by looking at the histograms. $\hat{\beta}_{MPL}$ has fairly good performance for all values of β though the MSE tends to be larger as β increases (this is actually expected according to its asymptotic theory). All Splus codes for the estimates are available on request.

REFERENCES

Aczél, J.(1966). *Lectures on Functional Equations and Their Applications*, Academic Press, New York-London.

Begun, J.M.(1987). Estimates of relative risk, *Metrika,* **34**, 65-82.

Begun, J.M. and Reid, N.(1983). Estimating the relative risk with censored data, *J. Amer. Statist. Assoc.*,**78**, 337-341.

Begun, J.M. and Wellner, J.A.(1983). Asymptotic efficiency of relative risk estimates, *Contributions to Statistics: Essays in Honor of Norman L.Johnson*,(ed. P.K.Sen), North-Holland, Amsterdam.

Bennett, S.(1983). Log-logistic regression models for survival data, *Appl. Statist.*,**32**, 165-171.

Box, G.E.P. and Cox, D.R.(1964). An analysis of transformation, *J. Roy. Statist. Soc. Ser. B*, **26**, 211-252.

Clayton, D. and Cuzick, J.(1985). Multivariate generalizations of the proportional hazards model (with discussion), *J. Roy. Statist. Soc. Ser. A*, **148**, 82-117.

Clayton, D. and Cuzick, J.(1986). The semi-parametric Pareto model for regression analysis of survival times, in *Papers on Semiparametric Models at the ISI Centenary Session* (R.D.Gill and M.N.Voors, eds.), 19-30. Centre for mathematics and Computer Science, Amsterdam.

Cox, D.R.(1972). Regression models and life table, *J. Roy. Statist. Soc. Ser. B*, **34**, 187-220.

Cuzick, J.(1988). Rank regression, *Ann. Statist.*, **16**, 1369-1389.

Dabrowska, D.M. and Doksum, K.A.(1988a). Estimation and testing in a two-sample generalized odds-rate model, *J. Amer. Statist. Assoc.*, **83**, 744-749.

Dabrowska, D.M. and Doksum, K.A.(1988b). Partial likelihood in transformation models with censored data, *Scand. J. Statist.*, **15**, 1-23.

Dabrowska, D.M., Doksum, K.A. and Miura, R.(1989). Rank estimates in a class of semiparametric two-sample models, *Ann. Inst. Statist. Math.*, **41**, 63-79.

Dempster, A.P., Laird, N.M. and Rubin, D.B.(1977). Maximum likelihood from incomplete data via the *EM* algorithm, *J. Roy. Statist. Soc. Ser. B,* **39**, 1-38.

Doksum, K.A.(1987). An extension of partial likelihood methods for proportional hazard models to general transformation models, *Ann. Statist.,* **15**, 325-345.

Ferguson, T.S.(1967). *Mathematical Statistics*, Academic Press, New York.

Hodges, J.L.Jr. and Lehmann, E.L.(1963). Estimates of location based on rank tests, *Ann. Math. Statist.,***34**, 598-611.

Hoeffding, W.(1951). "Optimum" nonparametric tests, *Proc. Second Berkeley Symp. Math. Statist. Probab.*, 83-92, Univ. California Press.

Huber, P.J.(1981). *Robust Statistics*, Wiley, New York.

Lehmann, E.L.(1953). The power of rank tests, *Ann. Math. Statist.,* **24**, 23-43.

McCullagh, P.(1980). Regression models for ordinal data, *J. Roy. Statist. Soc. Ser. B,* **42**, 109-142.

McCullagh, P.(1984). On the elimination of nuisance parameters in the proportional odds model, *J. Roy. Statist. Soc. Ser. B,* **46**, 250-256.

Miura, R.(1985). Hodges-Lehmann type estimates and Lehmann's alternative models : A special lecture presented at the annual meeting of Japanese Mathematical Society, the Division of Statistical Mathematics, April, 1985 (in Japanese).

Miura, R. and Tsukahara, H.(1992). One-sample estimation for generalized Lehmann's alternative models, *Statistica Sinica,* **3**, 83-101.

Pettitt, A.N.(1982). Inference for the linear model using a likelihood based on ranks, *J. Roy. Statist. Soc. Ser. B,* **44**, 234-243.

Pettitt, A.N.(1983). Approximate methods using ranks for regression with censored data, *Biometrika* **70**, 121-132.

Pettitt, A.N.(1984). Proportional odds models for survival data and estimates using ranks, *Appl. Statist.,***33**, 169-175.

Pettitt, A.N.(1987). Estimation for a regression parameter using ranks, *J. Roy. Statist. Soc. Ser. B*, **49**, 58-67.

Tsukahara, H.(1992). A rank estimator in the two-sample transformation model with randomly censored data, *Ann. Inst. Statist. Math.*, **44**, 313-333.

APPENDIX: *Results in Parametric Analysis*

Here we state the results in parametric analysis for the purpose of comparison. Let $X_1, \cdots, X_n$ be iid rv's with df $D(F(x);\theta)$. Its density is obviously $f(x)d(F(x);\theta)$. The log-likelihood $l(\theta)$ is then given by $\sum_{i=1}^n \log f(x_i) + \sum_{i=1}^n \log d(F(x_i);\theta)$, so that we get the efficient score

$$\dot{l}(\theta) = \sum_{i=1}^n \frac{\dot{d}(F(x_i);\theta)}{d(F(x_i);\theta)},$$

and the Fisher information

$$\mathrm{E}[(\dot{l}(\theta))^2] = n\int_0^1 \left[\frac{\dot{d}(t;\theta)}{d(t;\theta)}\right]^2 d(t;\theta)dt = nI(\theta), \quad \text{say}.$$

Typically, the MLE $\hat{\theta}$ is obtained by solving $\dot{l}(\theta) = 0$, and under the usual regularity conditions on $d(t;\theta)$, we have $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, 1/I(\theta))$.

We now give these quantities in the following particular models.

(i)*Proportional hazards model.* For $D(t;\theta) = 1 - (1-t)^{1/\theta}$,

$$\frac{\dot{d}(t;\theta)}{d(t;\theta)} = -\frac{1}{\theta} - \frac{\log(1-t)}{\theta^2}, \qquad I(\theta) = \frac{1}{\theta^2}.$$

For the reparametrization $\mu = \log\theta$, we have $I(\mu) = 1$.

(ii)*Proportional odds model.* For $D(t;\theta) = t/[(1-t)\theta + t]$,

$$\frac{\dot{d}(t;\theta)}{d(t;\theta)} = \frac{1}{\theta} - 2\frac{1-t}{(1-t)\theta + t}, \qquad I(\theta) = \frac{1}{3\theta^2},$$

and $I(\mu) = 1/3$.

(iii)*Proportional* γ *-odds model.* For

$$D(t;\theta) = 1 - \left[\frac{\theta(1-t)^\gamma}{1-(1-\theta)(1-t)^\gamma}\right]^{1/\gamma}, \qquad \gamma > 0,$$

we have

$$\frac{\dot{d}(t;\theta)}{d(t;\theta)} = \frac{1}{\gamma\theta} - \left(\frac{1}{\gamma}+1\right)\frac{(1-t)^\gamma}{1-(1-\theta)(1-t)^\gamma}.$$

$I(\theta)$ can be simplified only for γ such that $1/\gamma$ is an integer, say k . For such γ we get

$$I(\theta) = \frac{k}{\theta^2} - \frac{k(k+1)\theta^k}{(\theta-1)^{k+2}}\sum_{r=0}^{k+1}\binom{k+1}{r}\frac{(-1)^r}{r+1}\left(1-\frac{1}{\theta^{r+1}}\right).$$

(iv)*Transformation shift model.* For $D(t;\theta) = \Psi(\Psi^{-1}(t) - \log\theta)$,

$$\frac{\dot{d}(t;\theta)}{d(t;\theta)} = -\frac{1}{\theta}\frac{\psi'(\Psi^{-1}(t)-\log\theta)}{\psi(\Psi^{-1}(t)-\log\theta)}, \qquad I(\theta) = \frac{1}{\theta^2}\int_{-\infty}^{\infty}\left[\frac{\psi'(x)}{\psi(x)}\right]^2 d\Psi(x).$$

In terms of μ we have

$$I(\mu) = \int_{-\infty}^{\infty}\left[\frac{\psi'(x)}{\psi(x)}\right]^2 d\Psi(x),$$

which is the Fisher information of location family $\psi(\cdot - \mu)$.

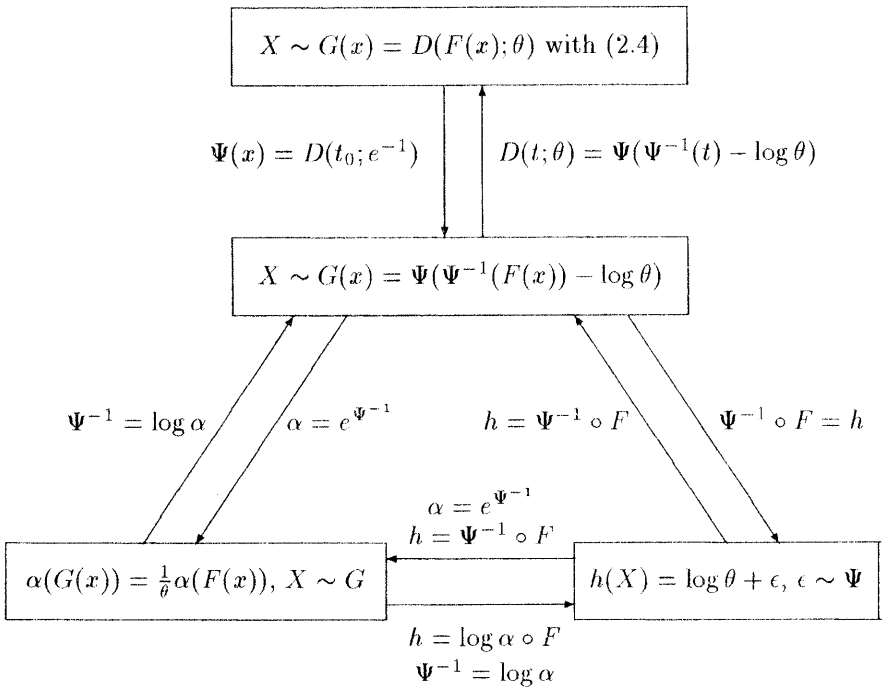
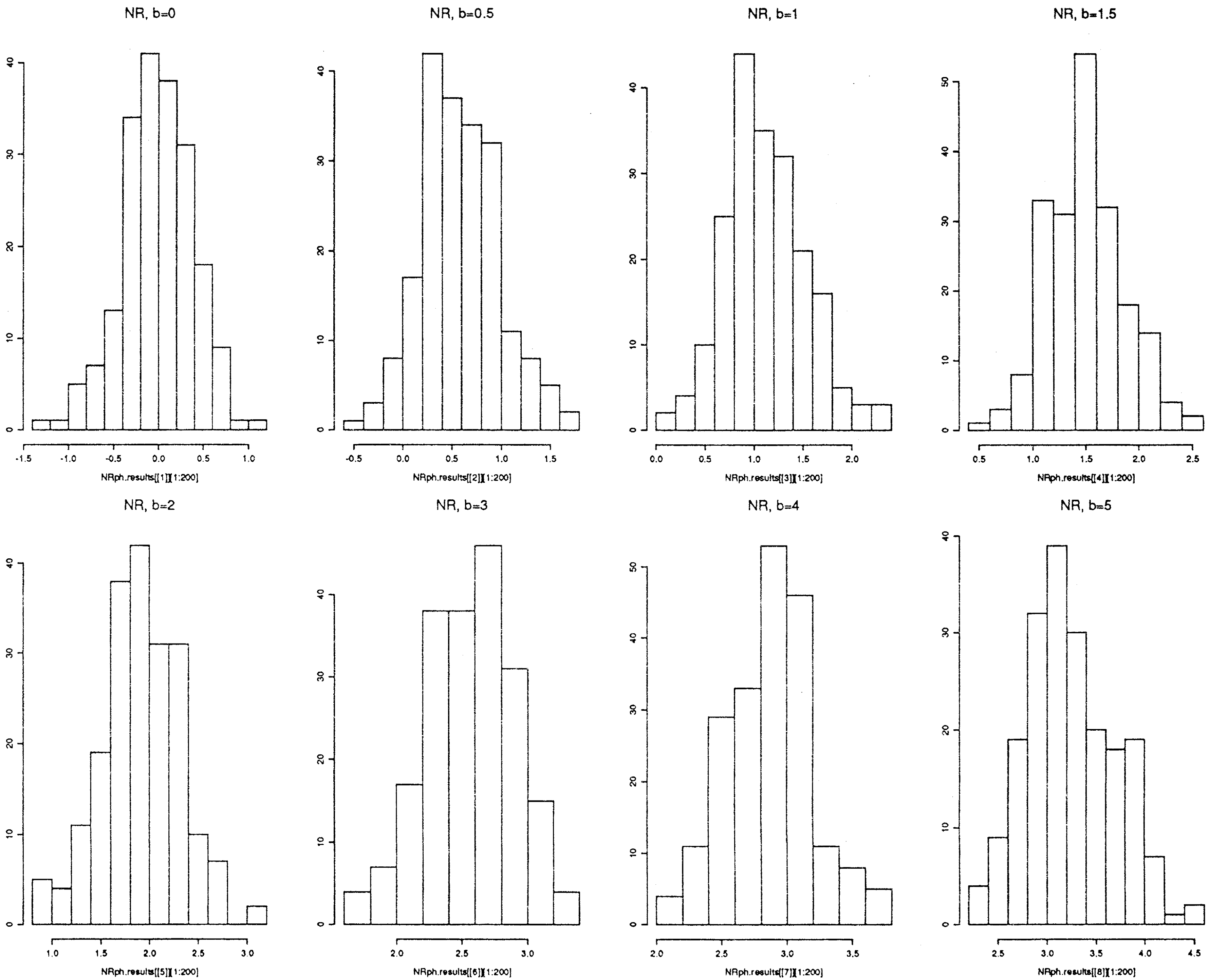


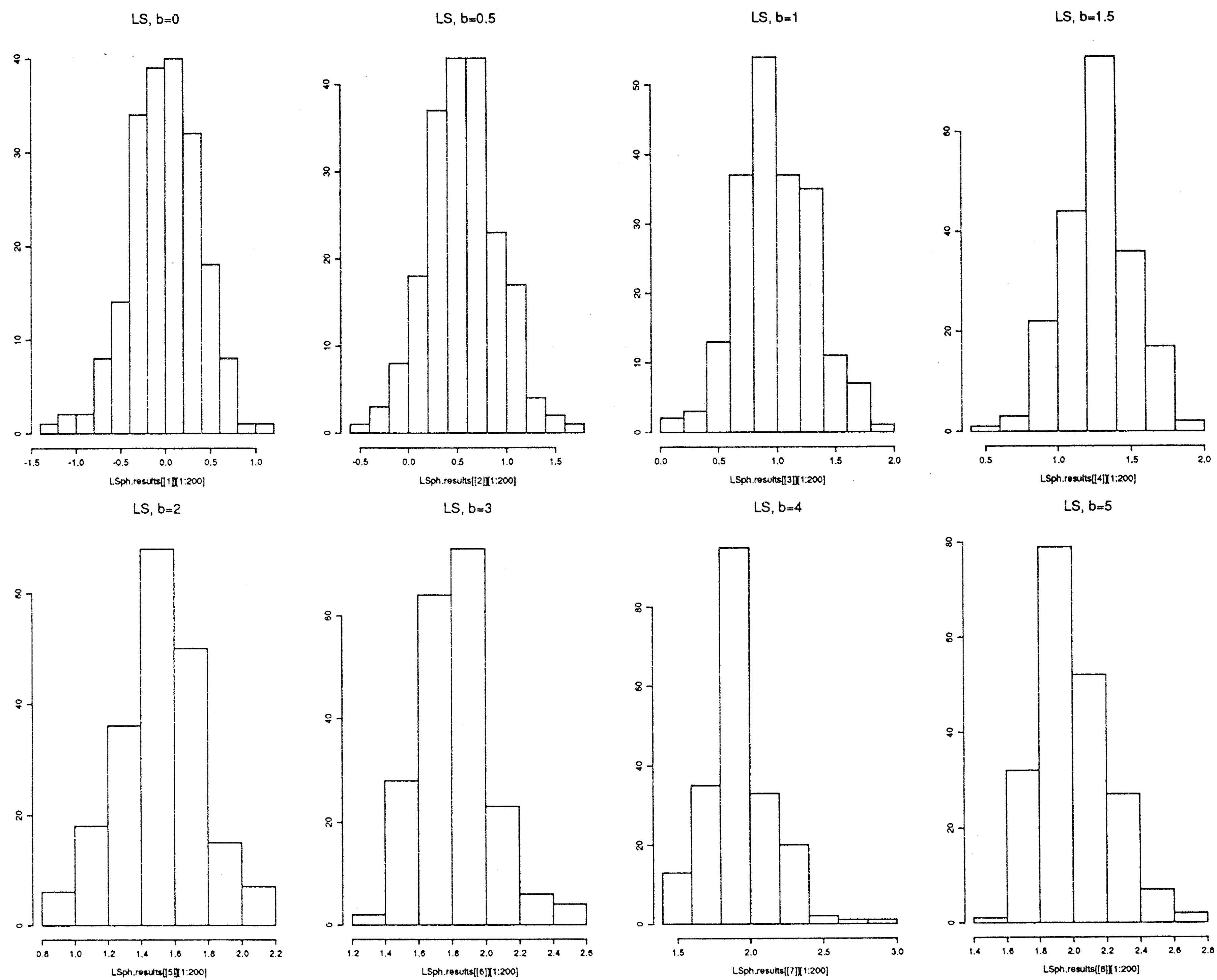
Figure 1: Relations between models.

	β	0	0.5	1	1.5	2	3	4	5
$\hat{\beta}_Q$	Bias	-0.0158	0.0905	0.1255	0.0042	-0.0869	-0.4372	-1.1283	-1.7680
	MSE	0.1522	0.1651	0.1925	0.1281	0.1686	0.3096	1.3816	3.3366
$\hat{\beta}_M$	Bias	-0.0144	0.0652	-0.0012	-0.2233	-0.4955	-1.1807	-2.0742	-2.9908
	MSE	0.1487	0.1360	0.1008	0.1046	0.3069	1.4400	4.3504	8.9892
$\hat{\beta}_E$	Bias	0.0301	0.1123	0.1872	0.1090	0.1543	0.3483	0.3055	0.4075
	MSE	0.1866	0.1941	0.3135	0.2848	0.5760	0.9519	1.6477	2.5861
$\hat{\beta}_{MPL}$	Bias	-0.0155	0.0881	0.1317	0.0526	0.0864	0.2869	0.1504	0.0299
	MSE	0.1495	0.1669	0.2238	0.1889	0.3201	0.7917	0.7919	1.2578
$\hat{\beta}_{RAM}$	Bias	-0.0236	-0.0732	-0.1882	-0.4243	-0.6354	-0.9137	-1.5103	-2.0614
	MSE	0.1201	0.1208	0.2084	0.3756	0.6793	1.2603	2.6259	4.6214

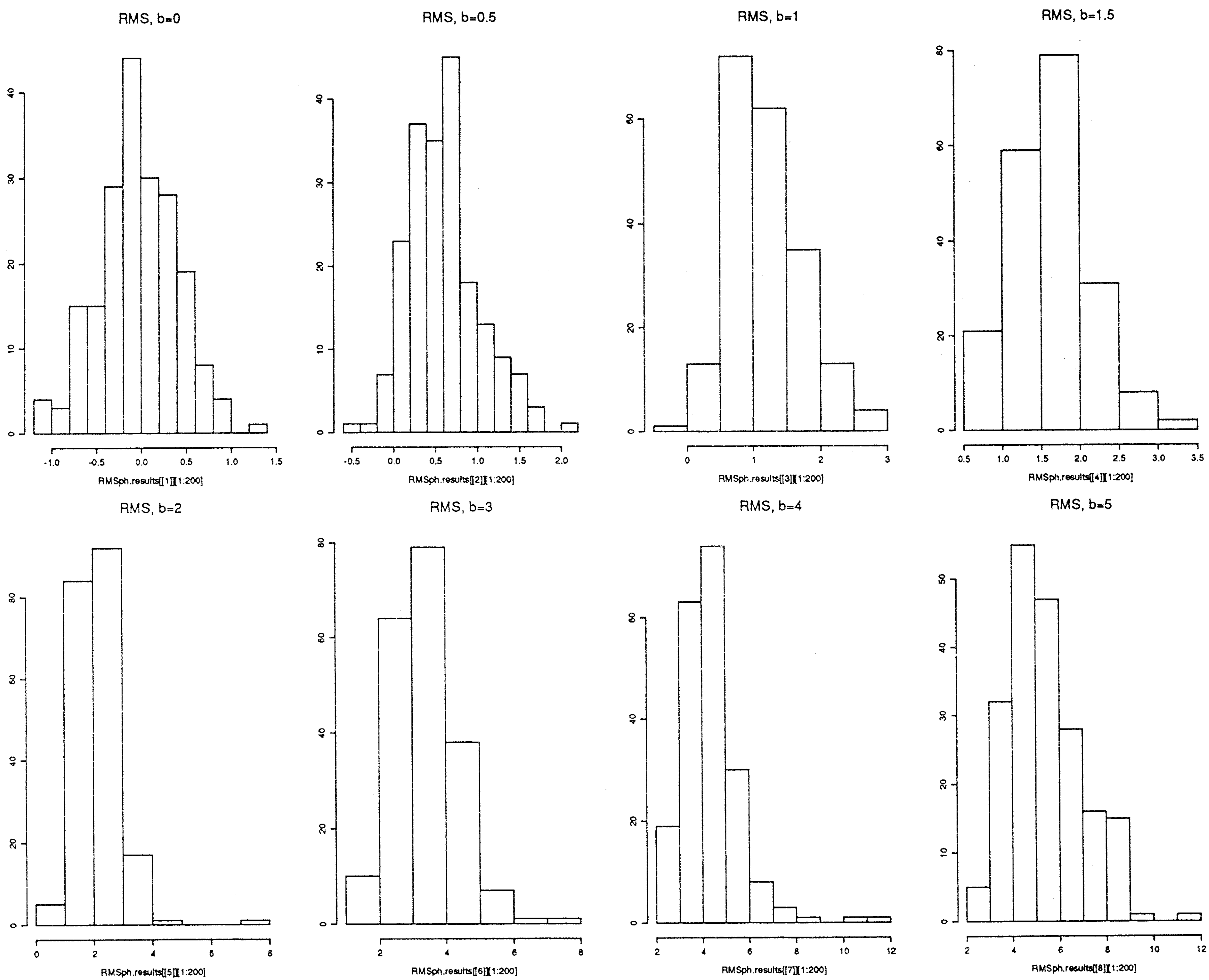
TABLE 1

Monte Carlo results for the proportional hazards model with $n = 25$, $p = 1$, $z_i = (i-13)/12$ and 200 Monte Carlo trials

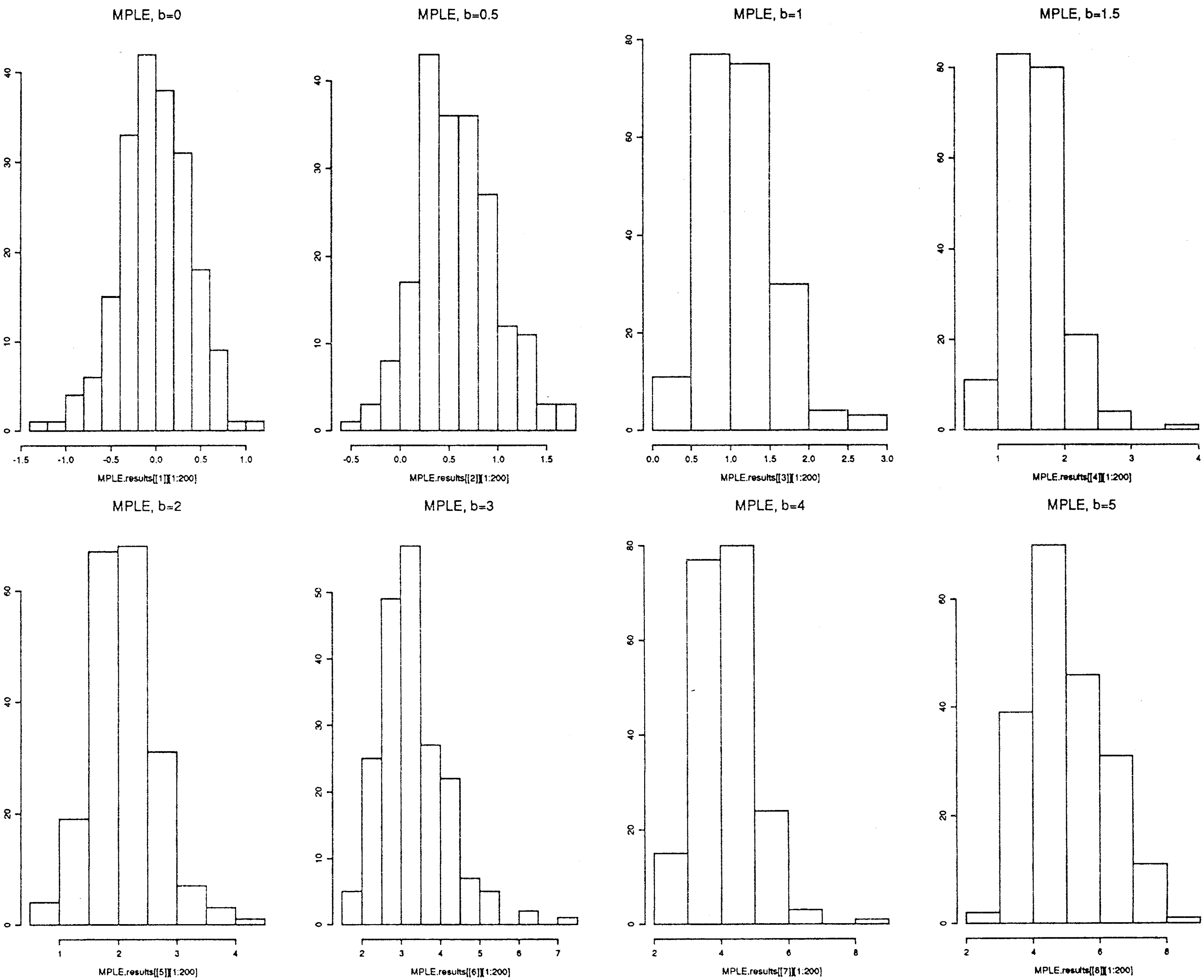

NR, b=0
NRph.results[[1]][1:200]
NR, b=0.5
NRph.results[[2]][1:200]
NR, b=1
NRph.results[[3]][1:200]
NR, b=1.5
NRph.results[[4]][1:200]
NR, b=2
NRph.results[[5]][1:200]
NR, b=3
NRph.results[[6]][1:200]
NR, b=4
NRph.results[[7]][1:200]
NR, b=5
NRph.results[[8]][1:200]



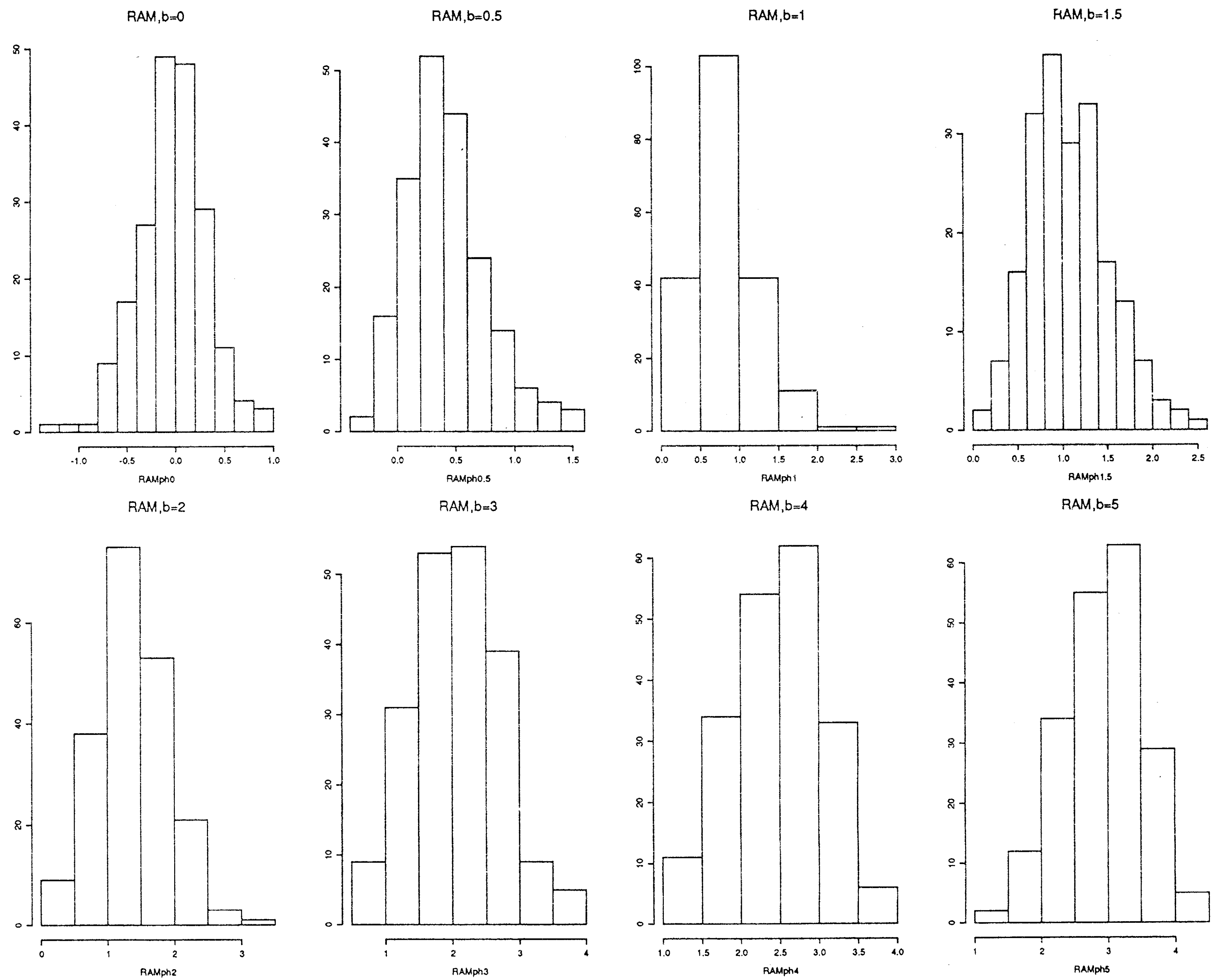
LS, b=0
LSph.results[[1]][1:200]
LS, b=0.5
LSph.results[[2]][1:200]
LS, b=1
LSph.results[[3]][1:200]
LS, b=1.5
LSph.results[[4]][1:200]
LS, b=2
LSph.results[[5]][1:200]
LS, b=3
LSph.results[[6]][1:200]
LS, b=4
LSph.results[[7]][1:200]
LS, b=5
LSph.results[[8]][1:200]



RMS, b=0
RMSph.results[[1]][1:200]
RMS, b=0.5
RMSph.results[[2]][1:200]
RMS, b=1
RMSph.results[[3]][1:200]
RMS, b=1.5
RMSph.results[[4]][1:200]
RMS, b=2
RMSph.results[[5]][1:200]
RMS, b=3
RMSph.results[[6]][1:200]
RMS, b=4
RMSph.results[[7]][1:200]
RMS, b=5
RMSph.results[[8]][1:200]



MPLE, b=0
MPLE.results[[1]][1:200]
MPLE, b=0.5
MPLE.results[[2]][1:200]
MPLE, b=1
MPLE.results[[3]][1:200]
MPLE, b=1.5
MPLE.results[[4]][1:200]
MPLE, b=2
MPLE.results[[5]][1:200]
MPLE, b=3
MPLE.results[[6]][1:200]
MPLE, b=4
MPLE.results[[7]][1:200]
MPLE, b=5
MPLE.results[[8]][1:200]



RAM,b=0
RAMph0
RAM,b=0.5
RAMph0.5
RAM,b=1
RAMph1
RAM,b=1.5
RAMph1.5
RAM,b=2
RAMph2
RAM,b=3
RAMph3
RAM,b=4
RAMph4
RAM,b=5
RAMph5