Hypothesis Testing Based on the Maximum of Two Statistics from Weighted and Unweighted Estimating Equations

Takeshi Emura* and Hisayuki Tsukuma†

Abstract

For testing the regression parameter in multivariate failure time data analysis, the solution to partial likelihood score equation provides useful test statistic. Proper weight insertion in the score equation allows us to use more powerful test statistic, especially when failure times within a cluster are strongly correlated. In some cases two solutions to weighted and unweighted score equations, which are calculated from the same data, are very different. This means two statistics based on the solutions to weighted and unweighted score equations yield different results of the testing. If one has no prior information in advance to prefer one statistic over the other, it is natural to consider the maximum of the two test statistics for testing the regression parameter. In this paper, we discuss the way to approximate the asymptotic distribution of the maximum statistic, and construct hypothesis testing based on the maximum statistic. Simulation studies are carried out to see the validity of the approximation of the maximum statistic and its power properties.

1 Introduction

Let $(T_{i1},\dots,T_{in}),\ i=1,\dots,K$, be K clusters of n-dimensional multivariate failure time data and let $(\boldsymbol{z}_{i1},\dots,\boldsymbol{z}_{in})$ be K clusters of the corresponding p-dimensional covariate vectors. Then, marginal Cox (1972)-type hazard models for the jth member in cluster i, with common regression parameter and baseline hazard function, formulates the hazard function of T_{ij} given $\boldsymbol{z}_{ij}$ as

$$\lambda(t|\boldsymbol{z}_{ij}) = e^{\boldsymbol{z}_{ij}^T\boldsymbol{\beta}}\lambda_0(t), \quad i=1,\dots,K,\ j=1,\dots,n. \tag{1.1}$$

Here, $\boldsymbol{\beta}$ is a p-vector of unknown common regression parameters and $\lambda_0(t)$ is an unknown, unspecified common baseline hazard function. Cai and Prentice (1997) introduced estimating equation for the regression parameter $\boldsymbol{\beta}$, by incorporating weight matrices into partial likelihood score equation. They proved that the solution to the estimating equation, $\hat{\boldsymbol{\beta}}$, is consistent and asymptotically normal.

Let $(C_{i1},\dots,C_{in})$ be censoring time. Assume that, given $\boldsymbol{z}_{ij}$, T_{ij} and C_{ij} are independent for all i,j. Considering the right censoring situation, observed data are $X_{ij}=\min\{T_{ij},C_{ij}\}$, $\delta_{ij}=I(T_{ij}\le C_{ij})$, and $\boldsymbol{z}_{ij}$ for every i,j. Assume that $(T_{i1},\dots,T_{in},C_{i1},\dots,C_{in},\boldsymbol{z}_{i1},\dots,\boldsymbol{z}_{in})$ are independent and identically distributed(i.i.d.)from the same probability model. We use only time independent covariates for brevity, but our methods can easily be extended to the case of time dependent covariates with finite variations. Define $N_{ij}(t)=I(X_{ij}\le t,\delta_{ij}=1)$, $Y_{ij}(t)=I(X_{ij}\ge t)$, then,

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s)e^{\boldsymbol{z}_{ij}^T\boldsymbol{\beta}}\lambda_0(s)ds$$

* Graduate School of Science and Technology, Chiba University, 1-33 Yayoi-cho, Chiba 263-8522, Japan; E-mail: s.s.takeshi@ezweb.ne.jp

† The institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan; E-mail: tsukuma@ism.ac.jp

is a martingale with respect to a marginal filtration (see Spiekerman and Lin (1998)). An estimated martingale replaces this martingale with its empirical counterpart, and thus,

$$\hat{M}_{ij}(\boldsymbol{\beta}, t) = N_{ij}(t) - \int_0^t Y_{ij}(s) e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}} d\hat{\Lambda}_0(\boldsymbol{\beta}, s),$$

where $\hat{\Lambda}_0(\boldsymbol{\beta}, t) = \sum_{i=1}^K \sum_{j=1}^n \int_0^t dN_{ij}(s) / (\sum_{k=1}^K \sum_{l=1}^n Y_{kl}(s) e^{\boldsymbol{z}_{kl}^T \boldsymbol{\beta}})$ and $\boldsymbol{\beta}$ is an arbitrary p-vector. Then, estimating equation of Cai and Prentice (1997) are defined as

$$\sum_{i=1}^K \sum_{j=1}^n \int_0^T \Big\{ \sum_{k=1}^n \boldsymbol{z}_{ik} w_{ikj}(\boldsymbol{\beta}, s) \Big\} d\hat{M}_{ij}(\boldsymbol{\beta}, s) = \mathbf{0}.$$

Here K weight matrices

$$\boldsymbol{W}_i(\boldsymbol{\beta}, t) = \begin{pmatrix} w_{i11}(\boldsymbol{\beta}, t) & \cdots & w_{i1n}(\boldsymbol{\beta}, t) \\ \vdots & \ddots & \vdots \\ w_{in1}(\boldsymbol{\beta}, t) & \cdots & w_{inn}(\boldsymbol{\beta}, t) \end{pmatrix}, \quad i = 1, \dots K, \tag{1.2}$$

are arbitrary functions of $\boldsymbol{\beta}$ and t and $T > 0$ is a fixed constant. In case every $\boldsymbol{W}_i(\boldsymbol{\beta}, t)$ is an identity matrix, (1) is partial likelihood score equation, and we call these unweighted estimating equation throughout this paper. Cai and Prentice (1997) showed in their simulation studies that the inverses of some correlation matrices between martingales allow the solution to the estimating equation to have smaller variance than the solution to the unweighted estimating equation.

One might not know whether weight insertion achieves great efficiency prior to the analysis of the experiment. It is expected that weight insertion always reduces the variance of the regression estimator when the weight matrices (1.2) are properly selected and are completely known. However, the proper weight matrices have generally to be estimated from data, and the deviation of the weight matrix estimators from the true values often causes the efficiency of the regression parameter estimator to reduce slightly. One might calculate two test statistics from both weighted estimating and unweighed estimating equation. Usually the two analyses yield very similar results, but in some cases, the results can be very different. On the nonparametric test of the equality of two survival curves, the quite similar problem of whether one uses the Mantel's logrank test or the generalized Wilcoxon test is discussed earlier by Tarone (1981). He suggests the maximum of the logrank test statistic and the Wilcoxon test statistic when none of prior information to prefer one statistic over the other is given.

In this paper, we consider the problem of the level α hypothesis testing of $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ in favor of the alternative $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ under model (1.1). Let $\hat{\boldsymbol{\beta}}_1$ be the solution to partial likelihood score equation, or unweighted estimating equation,

$$\boldsymbol{U}_1(\boldsymbol{\beta}) = \sum_{i=1}^K \sum_{j=1}^n \int_0^T \boldsymbol{z}_{ij} d\hat{M}_{ij}(\boldsymbol{\beta}, s) = \mathbf{0}, \tag{1.3}$$

and let $\hat{\boldsymbol{\beta}}_2$ be the solution to weighted estimating equation

$$\boldsymbol{U}_2(\boldsymbol{\beta}) = \sum_{i=1}^K \sum_{j=1}^n \int_0^T \Big\{ \sum_{k=1}^n \boldsymbol{z}_{ik} w_{ikj}(\boldsymbol{\beta}, s) \Big\} d\hat{M}_{ij}(\boldsymbol{\beta}, s) = \mathbf{0}. \tag{1.4}$$

For each $l (l = 1, 2)$, asymptotic normality of $\sqrt{K}(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_0)$ $(l = 1, 2)$ and consistency of their variance estimators $\hat{\boldsymbol{\Omega}}_l(\hat{\boldsymbol{\beta}}_l)$, which are described in Cai and Prentice (1997), yields two different Wald type hypothesis testings: One is the test which rejects H_0 whenever

$$Q_1 = K(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0)^T \hat{\boldsymbol{\Omega}}_1(\hat{\boldsymbol{\beta}}_1)^{-1} (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0) > \chi_p^2(\alpha) \tag{1.5}$$

and the other is the test which rejects H_0 whenever

$$Q_2 = K(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_0)^T \hat{\boldsymbol{\Omega}}_2(\hat{\boldsymbol{\beta}}_2)^{-1}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_0) > \chi_p^2(\alpha). \tag{1.6}$$

Here $\chi_p^2(\alpha)$ is a percentage point of chi square distribution with degrees of freedom p. Since one might not be able to select a statistic prior to the analysis of experiment, we will examine the next test which rejects H_0 whenever

$$Q_{max} = \max\{Q_1, Q_2\} > \hat{\chi}_{max}^2(\alpha). \tag{1.7}$$

Here $\hat{\chi}_{max}^2(\alpha)$, which is depend on data, is an estimator of a percentage point we describe later. The use of the third statistic (1.7) is regarded as the direct application of Tarone (1981)'s idea to the marginal Cox-type regression analysis. In Section 2, we examine the asymptotic distribution of Q_{max}, and construct an asymptotically level α test based on Q_{max}. In Section 3, simulation studies are provided under several parameter configurations to see the validity of the approximation of the statistic Q_{max}, and to compare the power properties of the three test statistics. In Section 4, it is concluded from our simulation studies that the empirical power of the test (1.7) is higher than the power of either the test (1.5) or the test (1.6) for most of the alternatives.

2 Theoretical results and test based on Q_{max}

2.1 Notation and theoretical results

We consider $\boldsymbol{\beta}$ an arbitrary element in a compact set $\mathcal{B}(\subset \boldsymbol{R}^p)$, which contains the true value $\boldsymbol{\beta}_0$. Define $\boldsymbol{a}_{1,ij}(\boldsymbol{\beta}, t) = \boldsymbol{z}_{ij}$ and $\boldsymbol{a}_{2,ij}(\boldsymbol{\beta}, t) = \sum_{k=1}^n \boldsymbol{z}_{ik} w_{ikj}(\boldsymbol{\beta}, t)$, and the following notation:

$$\begin{aligned}
S^{(0)}(\boldsymbol{\beta}, t) &= \frac{1}{K}\sum_{i=1}^K \sum_{j=1}^n Y_{ij}(t) e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}}, \\
\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t) &= \frac{1}{K}\sum_{i=1}^K \sum_{j=1}^n Y_{ij}(t) \boldsymbol{z}_{ij} e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}}, \\
\boldsymbol{S}_l^{(2)}(\boldsymbol{\beta}, t) &= \frac{1}{K}\sum_{i=1}^K \sum_{j=1}^n Y_{ij}(t) \boldsymbol{a}_{l,ij}(\boldsymbol{\beta}, t) e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}} \quad (l = 1, 2), \\
\boldsymbol{S}_l^{(3)}(\boldsymbol{\beta}, t) &= \frac{1}{K}\sum_{i=1}^K \sum_{j=1}^n Y_{ij}(t) \boldsymbol{a}_{l,ij}(\boldsymbol{\beta}, t) \boldsymbol{z}_{ij}^T e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}} \quad (l = 1, 2), \\
\boldsymbol{E}_l(\boldsymbol{\beta}, t) &= \frac{\boldsymbol{S}_l^{(2)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} \quad (l = 1, 2), \\
\boldsymbol{V}_l(\boldsymbol{\beta}, t) &= \frac{\boldsymbol{S}_l^{(3)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} - \frac{\boldsymbol{S}_l^{(2)}(\boldsymbol{\beta}, t) \boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t)^T}{S^{(0)}(\boldsymbol{\beta}, t)^2} \quad (l = 1, 2).
\end{aligned}$$

Under the model assumptions of the listed earlier the regularity conditions outlined in Section 3 of Cai and Prentice (1997), there exist non-random functions, on $\mathcal{B} \times [0, T]$, $\boldsymbol{s}_l^{(r)}(\boldsymbol{\beta}, t)$, $(r = 0, 1, 2, 3, 4)$, $\boldsymbol{e}_l(\boldsymbol{\beta}, t)$, and $\boldsymbol{v}_l(\boldsymbol{\beta}, t)$ such that $\sup_{\boldsymbol{\beta}, t} \|\boldsymbol{S}_l^{(r)}(\boldsymbol{\beta}, t) - \boldsymbol{s}_l^{(r)}(\boldsymbol{\beta}, t)\| \to^P \infty$, $\sup_{\boldsymbol{\beta}, t} \|\boldsymbol{E}_l(\boldsymbol{\beta}, t) - \boldsymbol{e}_l(\boldsymbol{\beta}, t)\| \to^P \infty$ and $\sup_{\boldsymbol{\beta}, t} \|\boldsymbol{V}_l(\boldsymbol{\beta}, t) - \boldsymbol{v}_l(\boldsymbol{\beta}, t)\| \to^P \infty$ $(K \to \infty)$ for $(l = 1, 2)$, and the matrix

$$\boldsymbol{A}_l(\boldsymbol{\beta}) = \sum_{j=1}^n \int_0^T \boldsymbol{v}_l(\boldsymbol{\beta}, s) s^{(0)}(\boldsymbol{\beta}, s) \lambda_0(s) ds$$

is assumed to be positive definite and symmetric. Define

$$\boldsymbol{D}_{l,i} = \sum_{j=1}^{n} \int_0^T \Big\{ \boldsymbol{a}_{l,ij}(\boldsymbol{\beta}_0, s) - \boldsymbol{e}_l(\boldsymbol{\beta}_0, s) \Big\} dM_{ij}(s) \quad (l = 1, 2),$$

then K random $2p$-vertors $(\boldsymbol{D}_{1,i}^T, \boldsymbol{D}_{2,i}^T)$ are i.i.d. and the covariance matrices between $\boldsymbol{D}_{l,i}$ and $\boldsymbol{D}_{m,i}$ are

$$\begin{aligned}\boldsymbol{\Sigma}_{lm}(\boldsymbol{\beta}_0) = E\Bigg\{ & \sum_{j=1}^{n} \int_0^T \Big\{ \boldsymbol{a}_{l,1j}(\boldsymbol{\beta}_0, s) - \boldsymbol{e}_l(\boldsymbol{\beta}_0, s) \Big\} dM_{1j}(s) \Bigg\} \\ & \times \Bigg\{ \sum_{j=1}^{n} \int_0^T \Big\{ \boldsymbol{a}_{m,1j}(\boldsymbol{\beta}_0, s) - \boldsymbol{e}_m(\boldsymbol{\beta}_0, s) \Big\} dM_{1j}(s) \Bigg\}^T \quad (l, m = 1, 2).\end{aligned}$$

Asymptotic distribution of $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$ such that $\boldsymbol{U}_1(\hat{\boldsymbol{\beta}}_1) = \boldsymbol{0}$ and $\boldsymbol{U}_2(\hat{\boldsymbol{\beta}}_2) = \boldsymbol{0}$ is summarized in the next proposition.

Proposition 1 *Suppose the model assumptions listed earlier and the regularity conditions A-D of the section 3 in Cai and Prentice (1997, pp 200) hold. Then,*

$$(\sqrt{K}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0)^T, \sqrt{K}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_0)^T) \xrightarrow{d} (\boldsymbol{W}_1^T, \boldsymbol{W}_2^T) \quad (K \to \infty),$$

where both $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, are p-vectors and $(\boldsymbol{W}_1^T, \boldsymbol{W}_2^T)$ is 2p-dimensional normal distribution with $E(\boldsymbol{W}_l) = \boldsymbol{0}$ and $E(\boldsymbol{W}_l \boldsymbol{W}_m^T) = \boldsymbol{A}_l(\boldsymbol{\beta}_0)^{-1} \boldsymbol{\Sigma}_{lm}(\boldsymbol{\beta}_0) \boldsymbol{A}_m(\boldsymbol{\beta}_0)^{-1} \equiv \boldsymbol{\Omega}_{lm}(\boldsymbol{\beta}_0)$ for $l, m = 1, 2$.

Proof. By the same argument in the Appendix of Cai and Prentice (1997, pp 212),

$$\sqrt{K}(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{K}} \sum_{i=1}^{K} \boldsymbol{A}_l(\boldsymbol{\beta}_0)^{-1} \boldsymbol{D}_{l,i} + \boldsymbol{o}^{p \times 1}(1),$$

for $l = 1, 2$, where $\boldsymbol{o}^{p\times 1}(1)$ is a $p \times 1$ vector with $o_p(1)$ elements. To establish multivariate convergence, it is sufficient by the Cramer-Wold device to prove

$$\sum_{l=1}^{2} \boldsymbol{c}_l^T \sqrt{K}(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_0) \xrightarrow{d} \sum_{l=1}^{2} \boldsymbol{c}_l^T \boldsymbol{W}_l$$

for all $\boldsymbol{c}_1 \in \boldsymbol{R}^p$ and $\boldsymbol{c}_2 \in \boldsymbol{R}^p$. The central limit theorem for i.i.d. random vectors establishes

$$\sum_{l=1}^{2} \boldsymbol{c}_l^T \frac{1}{\sqrt{K}} \sum_{i=1}^{K} \boldsymbol{A}_l(\boldsymbol{\beta}_0)^{-1} \boldsymbol{D}_{l,i} \xrightarrow{d} N(0, \sigma),$$

where $\sigma = \sum_{l=1}^{2} \sum_{m=1}^{2} \boldsymbol{c}_l^T \boldsymbol{A}_l(\boldsymbol{\beta}_0)^{-1} \boldsymbol{\Sigma}_{lm}(\boldsymbol{\beta}_0) \boldsymbol{A}_m(\boldsymbol{\beta}_0)^{-1} \boldsymbol{c}_l$. It is easy to see that the distribution of $\sum_{l=1}^{2} \boldsymbol{c}_l^T \boldsymbol{W}_l$ is equal to $N(0, \sigma)$. □

If the weight functions are unknown, one needs to replace $w_{ikj}(\boldsymbol{\beta}, s)$ by its estimator $\hat{w}_{ikj}(\boldsymbol{\beta}, s)$. We suspect that, by adding some regularity conditions such as the condition E of Cai and Prentice (1997), Proposition 1 can be extended to the case where weight functions are estimated.

Note that the covariance matrix of $(\boldsymbol{W}_1^T, \boldsymbol{W}_2^T)$ can be estimated by the following plug-in estimators and we define them for the estimating equation with both deterministic weights and estimated weights.

$$\hat{\boldsymbol{\Omega}}_{ll}(\hat{\boldsymbol{\beta}}_l) = \hat{\boldsymbol{A}}_l(\hat{\boldsymbol{\beta}}_l)^{-1} \hat{\boldsymbol{\Sigma}}_l(\hat{\boldsymbol{\beta}}_l) (\hat{\boldsymbol{A}}_l(\hat{\boldsymbol{\beta}}_l)^{-1})^T \quad (l = 1, 2), \tag{2.1}$$

$$\hat{\mathbf{\Omega}}_{lm}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_m) = \hat{\boldsymbol{A}}_l(\hat{\boldsymbol{\beta}}_l)^{-1}\hat{\mathbf{\Sigma}}_{lm}(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\beta}}_m)(\hat{\boldsymbol{A}}_m(\hat{\boldsymbol{\beta}}_m)^{-1})^T \quad (l \neq m), \tag{2.2}$$

$$\hat{\boldsymbol{A}}_l(\boldsymbol{\beta}) = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n}\int_0^T \hat{\boldsymbol{V}}_l(\boldsymbol{\beta}, s)dN_{ij}(s) \quad (l = 1, 2)$$

$$\hat{\mathbf{\Sigma}}_{11}(\boldsymbol{\beta}) = \frac{1}{K}\sum_{i=1}^{K}\Big\{\sum_{j=1}^{n}\int_0^T \{\boldsymbol{z}_{ij} - \boldsymbol{E}_1(\boldsymbol{\beta}, s)\}d\hat{M}_{ij}(\boldsymbol{\beta}, s)\Big\} \times \Big\{\sum_{j=1}^{n}\int_0^T \{\boldsymbol{z}_{ij} - \boldsymbol{E}_1(\boldsymbol{\beta}, s)\}d\hat{M}_{ij}(\boldsymbol{\beta}, s)\Big\}^T,$$

$$\hat{\mathbf{\Sigma}}_{22}(\boldsymbol{\beta}) = \frac{1}{K}\sum_{i=1}^{K}\Big\{\sum_{j=1}^{n}\int_0^T \{\hat{\boldsymbol{a}}_{2,ij}(\boldsymbol{\beta}, s) - \hat{\boldsymbol{E}}_2(\boldsymbol{\beta}, s)\}d\hat{M}_{ij}(\boldsymbol{\beta}, s)\Big\} \times \Big\{\sum_{j=1}^{n}\int_0^T \{\hat{\boldsymbol{a}}_{2,ij}(\boldsymbol{\beta}, s) - \hat{\boldsymbol{E}}_2(\boldsymbol{\beta}, s)\}d\hat{M}_{ij}(\boldsymbol{\beta}, s)\Big\}^T,$$

$$\hat{\mathbf{\Sigma}}_{12}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \frac{1}{K}\sum_{i=1}^{K}\Big\{\sum_{j=1}^{n}\int_0^T \{\boldsymbol{z}_{ij} - \boldsymbol{E}_1(\boldsymbol{\beta}_1, s)\}d\hat{M}_{ij}(\boldsymbol{\beta}_1, s)\Big\} \times \Big\{\sum_{j=1}^{n}\int_0^T \{\hat{\boldsymbol{a}}_{2,ij}(\boldsymbol{\beta}_2, s) - \hat{\boldsymbol{E}}_2(\boldsymbol{\beta}_2, s)\}d\hat{M}_{ij}(\boldsymbol{\beta}_2, s)\Big\}^T,$$

$$\hat{\mathbf{\Sigma}}_{21}(\boldsymbol{\beta}_2, \boldsymbol{\beta}_1) = \hat{\mathbf{\Sigma}}_{12}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^T.$$

We define $\hat{\boldsymbol{a}}_{2,ij}$, $\hat{\boldsymbol{E}}_2$, and $\hat{\boldsymbol{V}}_2$ as $\boldsymbol{a}_{2,ij}$, $\boldsymbol{E}_2$, and $\boldsymbol{V}_2$ respectively if weights are not estimated, and as $\boldsymbol{a}_{2,ij}$, $\boldsymbol{E}_2$, and $\boldsymbol{V}_2$ with $w_{ikj}(\boldsymbol{\beta}, s)$ replaced by its estimator $\hat{w}_{ikj}(\boldsymbol{\beta}, s)$ if weights are estimated. The estimators (2.1) are suggested by Cai and Prentice (1997), and they point out that these are consistent estimators for $\mathbf{\Omega}_{11}(\boldsymbol{\beta}_0)$ and $\mathbf{\Omega}_{22}(\boldsymbol{\beta}_0)$. We suggest the estimator (2.2), which is a natural extension of (2.1) to $\mathbf{\Omega}_{lm}(\boldsymbol{\beta}_0)$ $(l, m = 1, 2)$, and it may be a consistent estimator.

2.2 The test based on the maximum statistics Q_{max}

Proposition 1 and the estimator of the covariance matrix, (2.1) and (2.2), allow us to approximate the percentage point of Q_{max}. According to Proposition 1, the distribution of $(\sqrt{K}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0)^T, \sqrt{K}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_0)^T)$ is approximated by $(\boldsymbol{W}_1^T, \boldsymbol{W}_2^T)$. The unknown distribution of $(\boldsymbol{W}_1^T, \boldsymbol{W}_2^T)$ is further approximated by the known $2p$-dimensional normal distribution

$$\begin{bmatrix} \hat{\boldsymbol{W}}_1 \\ \hat{\boldsymbol{W}}_2 \end{bmatrix} \sim \mathcal{N}_{2p}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \hat{\mathbf{\Omega}}_{11}(\hat{\boldsymbol{\beta}}_1) & \hat{\mathbf{\Omega}}_{12}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) \\ \hat{\mathbf{\Omega}}_{21}(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_1) & \hat{\mathbf{\Omega}}_{22}(\hat{\boldsymbol{\beta}}_2) \end{bmatrix}\right).$$

A pair of the known random variables $(\hat{\boldsymbol{W}}_1^T\hat{\mathbf{\Omega}}_{11}(\hat{\boldsymbol{\beta}}_1)^{-1}\hat{\boldsymbol{W}}_1, \hat{\boldsymbol{W}}_2^T\hat{\mathbf{\Omega}}_{22}(\hat{\boldsymbol{\beta}}_2)^{-1}\hat{\boldsymbol{W}}_2)$ can approximate the distribution of (Q_1, Q_2) as long as $\hat{\mathbf{\Omega}}_{ll}(\hat{\boldsymbol{\beta}}_l)$ $(l = 1, 2)$ are positive definite. Exactly the same approximation can be obtained by using standardization of the normal distribution of $\hat{\boldsymbol{W}}_l$ $(l = 1, 2)$. Define $\hat{\boldsymbol{V}}_l = \hat{\mathbf{\Omega}}_{ll}(\hat{\boldsymbol{\beta}}_l)^{-1/2}\hat{\boldsymbol{W}}_l$ and $\hat{\mathbf{\Omega}}_{ll}(\hat{\boldsymbol{\beta}}_l)^{-1/2}\hat{\mathbf{\Omega}}_{ll}(\hat{\boldsymbol{\beta}}_l)^{-1/2} = \hat{\mathbf{\Omega}}_{ll}(\hat{\boldsymbol{\beta}}_l)^{-1}$ $(l = 1, 2)$. Then the distribution of $(\hat{\boldsymbol{V}}_1^T, \hat{\boldsymbol{V}}_2^T)$ is $2p$-dimensional normal with mean $\mathbf{0}$ and covariance matrix given by

$$\begin{pmatrix} \boldsymbol{I}_p & \hat{\mathbf{\Gamma}}_{12} \\ \hat{\mathbf{\Gamma}}_{12}^T & \boldsymbol{I}_p \end{pmatrix},$$

where $\boldsymbol{I}_p$ is a $p \times p$ identity matrix and $\hat{\mathbf{\Gamma}}_{12} = \hat{\mathbf{\Omega}}_{11}(\hat{\boldsymbol{\beta}}_1)^{-1/2}\hat{\mathbf{\Omega}}_{12}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)\hat{\mathbf{\Omega}}_{22}(\hat{\boldsymbol{\beta}}_2)^{-1/2}$. Again, the known distribution $(\hat{\boldsymbol{V}}_1^T\hat{\boldsymbol{V}}_1, \hat{\boldsymbol{V}}_2^T\hat{\boldsymbol{V}}_2)$ can approximate the distribution of (Q_1, Q_2). The percentage

point of $Q_{max} = \max\{Q_1, Q_2\}$, say $\hat{\chi}^2_{max}(\alpha)$ in (1.7), is approximated by the percentage point of the distribution of $\max\{\hat{\boldsymbol{W}}_1^T \hat{\boldsymbol{\Omega}}_{11}(\hat{\boldsymbol{\beta}}_1)^{-1}\hat{\boldsymbol{W}}_1, \hat{\boldsymbol{W}}_2^T \hat{\boldsymbol{\Omega}}_{22}(\hat{\boldsymbol{\beta}}_2)^{-1}\hat{\boldsymbol{W}}_2\}$ or $\max\{\hat{\boldsymbol{V}}_1^T\hat{\boldsymbol{V}}_1, \hat{\boldsymbol{V}}_2^T\hat{\boldsymbol{V}}_2\}$, for example, via Monte Carlo simulations.
Similar to the fact mentioned in Fleming and Harrington (1991, pp.280) for weighted logrank tests, the test (1.7) is closely related to the p-values of the tests (1.5) and (1.6). Suppose one computes the Wald statistics Q_1 and Q_2, and corresponding p-values p_1 and p_2. The test which rejects H_0 if $\min\{p_1,\ p_2\} \le p(\alpha)$ such that $P_{H_0}(\min\{p_1, p_2\} \le p(\alpha)) = \alpha$ is conceivable as long as $p(\alpha)$ can be computed. From the definition of p-value, we can see

$$\min\{p_1, p_2\} \le p(\alpha) \iff Q_{max} \ge \chi^2_p(p(\alpha)).$$

and therefore, the value $p(\alpha)$ is approximated by $\hat{p}(\alpha) = 1 - F_{\chi^2_p}(\hat{\chi}^2_{max}(\alpha))$, where $F_{\chi^2_p}$ is the distribution function of χ^2_p. Hence, the test (1.7) is exactly the same as the test which rejects H_0 whenever $\min\{p_1,\ p_2\} \le \hat{p}(\alpha)$.

3 Simulation Studies

We focused on the case where $p = 1$ in (1.1) throughout our simulation studies and hence β is a scalor in which $H_0 : \beta = \beta_0$ versus $H_1 : \beta \ne \beta_0$ are tested. The models we used were 2- and 3-dimensional models of Clayton and Cuzick (1985): The conditional survival function for $(T_1, \ldots, T_n)$ given $(z_1, \ldots, z_n)$ is

$$S(t_1, \ldots, t_n | z_1, \ldots, z_n) = \Big\{\sum_{j=1}^{n} \exp\big(t_j \theta e^{z_j \beta}\big) - (n-1)\Big\}^{-1/\theta} \quad (n = 2, 3).$$

The parameter θ determines the degree of dependence between T_j and $T_k, (j, k = 1, \ldots, n)$, and in simulation studies we used $\theta = 0.25$ (strongly correlated), $\theta = 1.5$ (weakly correlated) and $\theta = 10$ (nearly independent). The covariates $z_1, \ldots, z_n$ were considered to be mutually independent, $\{0-1\}$ valued random variable with $P(z_1 = 1) = \cdots = P(z_n = 1) = 0.5$. The censoring variables $C_{i1}, \ldots, C_{in}$ were independently generated from exponential distributions with hazard function being $\mu_1, \ldots, \mu_n$ respectively. The number of clusters, K, was fixed at 100. The unweighted score equation (1.3) was solved by using Newton-Raphson procedure and $\hat{\beta}_1$ denotes this solution. For weight matrices, we specified

$$\begin{pmatrix} w_{i11} & w_{i12} & \cdots & w_{i1n} \\ w_{i21} & w_{i22} & \cdots & w_{i2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{in1} & w_{in2} & \cdots & w_{inn} \end{pmatrix} = \begin{pmatrix} 1 & \rho_{i12}(z_{i1}, z_{i2}) & \cdots & \rho_{i1n}(z_{i1}, z_{in}) \\ \rho_{i12}(z_{i1}, z_{i2}) & 1 & \cdots & \rho_{i2n}(z_{i2}, z_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{i1n}(z_{i1}, z_{in}) & \rho_{i2n}(z_{i2}, z_{in}) & \cdots & 1 \end{pmatrix}^{-1},$$

where $\rho_{ijk}(z_{ij}, z_{ik}) = corr\{M_{ij}(X_{ij}), M_{ik}(X_{ik}) | z_{ij}, z_{ik}\}$ is a correlation between two unknown random variables given (z_{ij}, z_{ik}) for $i = 1, \ldots, K$, and $j, k = 1, \ldots, n$. This weight matrices are slightly different from those used by Cai and Prentice (1997) in their simulation studies, in that we do not need to know potential censoring times for the failing individuals. Since these weight matrices are unknown, we replaced these by their empirical estimators. The restriction of z_{ij} to be $\{0-1\}$ valued is crucial since the number of parameters in the all weight matrices reduces significantly. In the case of $n = 2$, only four parameters, $\rho_{12}(0, 0)$, $\rho_{12}(0, 1)$, $\rho_{12}(1, 0)$ and $\rho_{12}(1, 1)$ determine the weight matrices, and in the case of $n = 3$, so do only 12 parameters. Each value of $\rho_{ijk}(z_{ij}, z_{ik})$ was estimated by the product moment correlation coefficient of the group of pseudo-data $(\hat{M}_{i'j}(\hat{\beta}_1, X_{i'j}), \hat{M}_{i'k}(\hat{\beta}_1, X_{i'k}))$, $i' \in R(z_{ij}, z_{ik})$, where $R(z_{ij}, z_{ik}) = \{i' | z_{i'j} = z_{ij}, z_{i'k} = z_{ik}\}$. When $\#R(z_{ij}, z_{ik}) = 0$ or 1, we set $\hat{\rho}_{ijk}(z_{ij}, z_{ik}) = 0$.

$\hat{\beta}_2$ denotes the solution to the estimating equation (1.4) with these estimated weights, and this was solved by using Newton-Raphson procedure. Rarely do extremely large values of the estimated weights cause the Newton-Raphson algorithm to stop. We regulated these phenomena by letting the values of $Q_l(l = 1, 2)$ be 0 whenever $-\partial U_l(\beta)/\partial\beta < 0.001$ $(l = 1, 2)$. All simulations were based on 1,000 runs, in which ERR1 and ERR2 indicate the total number of this type of error regulations in both $U_1(\beta)$ and $U_2(\beta)$ respectively. For each run, the statistics Q_1, Q_2 and Q_{max} and the estimator $\hat{\Gamma}_{12}$ were computed. The quantiles of the Q_{max}, that is $\hat{\chi}^2_{max}(0.05)$, were approximated by the quantiles based on the 10,000 random samples from the distribution of $(\hat{V}_1, \hat{V}_2)$ for every run.

First, our aim in this simulation studies is to see how well the distribution of $(\hat{V}_1, \hat{V}_2)$ approximates that of $(\sqrt{K}(\hat{\beta}_1 - \beta_0)/\{\hat{\Omega}_1(\hat{\beta}_1)^{1/2}\}, \sqrt{K}(\hat{\beta}_2 - \beta_0)/\{\hat{\Omega}_2(\hat{\beta}_2)^{1/2}\})$. Since the distribution of $(\hat{V}_1, \hat{V}_2)$ is completely determined by the estimator $\hat{\Gamma}_{12}$, it is worth seeing the performance of $\hat{\Gamma}_{12}$ compared with the actual correlation coefficient of $(\sqrt{K}(\hat{\beta}_1 - \beta_0)/\{\hat{\Omega}_1(\hat{\beta}_1)^{1/2}\}, \sqrt{K}(\hat{\beta}_2 - \beta_0)/\{\hat{\Omega}_2(\hat{\beta}_2)^{1/2}\})$. Mean and standard deviation of $\hat{\Gamma}_{12}$ and product moment correlation coefficient of $(\sqrt{K}(\hat{\beta}_1 - \beta_0)/\{\hat{\Omega}_1(\hat{\beta}_1)^{1/2}\}, \sqrt{K}(\hat{\beta}_2 - \beta_0)/\{\hat{\Omega}_2(\hat{\beta}_2)^{1/2}\})$ are listed as $\bar{\Gamma}_{12}$, SD($\hat{\Gamma}_{12}$), and r respectively in Table 1. P denotes the probability of censoring calculated from the actual sample.

For the test of $\beta_0 = 0$, that is $H_0 : \beta = 0$, the empirical size and power of Q_1, Q_2 and Q_{max} based on these 5% points are given in Tables 2 and 3. Based on the Tables 2 and 3, power functions are drawn in Figures 1-4.

4 Concluding remarks

We have discussed the way to approximate the asymptotic distribution of Q_{max} to construct the hypothesis testing (1.7). As shown in Table 1, the approximation of Q_{max} by $(\hat{V}_1, \hat{V}_2)$ performed well in terms of the correlation coefficient of $(\sqrt{K}(\hat{\beta}_1 - \beta_0)/\{\hat{\Omega}_1(\hat{\beta}_1)^{1/2}\}, \sqrt{K}(\hat{\beta}_2 - \beta_0)/\{\hat{\Omega}_2(\hat{\beta}_2)^{1/2}\})$ and $\hat{\Gamma}_{12}$. Empirical size of Q_{max} listed in Tables 2 and 3 appears to slightly larger than the nominal 5% level, especially in the case of 3-dimensional Clayton model. At almost every alternatives, the empirical power of the test based on Q_{max} did not become the worst ones compared to Q_1 and Q_2. Therefore, we can conclude that the test (1.7) tends to prevent the power reduction caused by selecting the wrong estimating equation.

Acknowledgement

The authors would like to thank Dr. Yoshihiko Konno for helpful comments.

References

[1] Cai, J. and Prentice, R.L. (1997). Regression estimation using multivariate failure time data and a common baseline hazard function model, *Lifetime Data Analysis* **3**, 197–213.

[2] Clayton, D.G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion), *Journal of the Royal Statistical Society* Series A **148**, 82–117.

[3] Fleming, T.R. and Harrington, D.P. (1991). *Counting Process and Survival Analysis* , Wiley: New York.

[4] Spiekerman, C.F. and Lin, D.Y. (1998). Marginal regression models for multivariate failure time data, *Journal of the American Statistical Association* **93**, 1164–1175.

[5] Tarone, R.E. (1981). On the distribution of the maximum of the logrank statistics and the modified Wilcoxon statistics, *Biometrics* **37**, 79–85.

Table 1: Mean and standard deviation of $\hat{\Gamma}_{12}$ and product moment correlation coefficient of $(\sqrt{K}(\hat{\beta}_1 - \beta_0)/\{\hat{\Omega}_1(\hat{\beta}_1)^{1/2}\}, \sqrt{K}(\hat{\beta}_2 - \beta_0)/\{\hat{\Omega}_2(\hat{\beta}_2)^{1/2}\})$

2 dimensional Clayton model

	θ	β	P	$\bar{\Gamma}_{12}$	SD($\hat{\Gamma}_{12}$)	r	ERR1	ERR2
$\mu_1 = \mu_2 = 0.5$	0.25	-1	0.457	0.792	0.06876	0.779	0	0
		0	0.335	0.719	0.07732	0.682	0	0
		1	0.246	0.640	0.09517	0.617	0	0
	1.5	-1	0.457	0.941	0.03920	0.934	0	0
		0	0.334	0.924	0.04760	0.908	0	0
		1	0.245	0.892	0.05763	0.879	0	0
	10	-1	0.456	0.966	0.03047	0.962	0	0
		0	0.334	0.969	0.02786	0.965	0	0
		1	0.245	0.964	0.03144	0.960	0	0
$\mu_1 = \mu_2 = 1.5$	0.25	-1	0.703	0.883	0.05480	0.872	0	0
		0	0.601	0.844	0.05737	0.826	0	0
		1	0.479	0.776	0.06731	0.754	0	0
	1.5	-1	0.704	0.959	0.03291	0.957	0	0
		0	0.602	0.958	0.02913	0.953	0	0
		1	0.479	0.934	0.04295	0.923	0	0
	10	-1	0.704	0.966	0.03241	0.965	0	0
		0	0.602	0.972	0.02467	0.969	0	0
		1	0.479	0.969	0.02839	0.965	0	0

3 dimensional Clayton model

	θ	β	P	$\bar{\Gamma}_{12}$	SD($\hat{\Gamma}_{12}$)	r	ERR1	ERR2
$\mu_1 = \mu_2 = \mu_3 = 0.5$	0.25	-1	0.454	0.694	0.10175	0.717	0	2
		0	0.333	0.597	0.11294	0.564	0	6
		1	0.243	0.545	0.13316	0.538	0	6
	1.5	-1	0.454	0.880	0.06265	0.884	0	0
		0	0.333	0.855	0.06779	0.836	0	0
		1	0.244	0.802	0.09434	0.783	0	2
	10	-1	0.454	0.920	0.05008	0.915	0	0
		0	0.333	0.927	0.04469	0.920	0	0
		1	0.244	0.914	0.05169	0.909	0	0
$\mu_1 = \mu_2 = \mu_3 = 1.5$	0.25	-1	0.703	0.786	0.09564	0.797	0	4
		0	0.601	0.759	0.07355	0.750	0	0
		1	0.478	0.697	0.08451	0.663	0	0
	1.5	-1	0.703	0.908	0.05199	0.910	0	0
		0	0.600	0.910	0.04654	0.919	0	0
		1	0.477	0.876	0.05896	0.872	0	0
	10	-1	0.703	0.916	0.05954	0.921	0	0
		0	0.600	0.930	0.04345	0.929	0	0
		1	0.478	0.925	0.04377	0.923	0	0

Table 2: Empirical power of tests based on Q_1, Q_2, and Q_{max} for nominal 5 % significance level under 2 dimensional Clayton model
(* *indicates the worst empirical power among* Q_1, Q_2, *and* Q_{max})

$\mu_1 = \mu_2 = 0.5$

β (true)	-1	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	**0.0**	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\theta = 0.25$																					
Q_1	0.997*	0.994*	0.985*	0.961*	0.896*	0.779*	0.613*	0.423*	0.213*	0.088*	**0.046**	0.102*	0.230*	0.430*	0.638*	0.827*	0.937*	0.987*	0.999*	1.000	1.000
Q_2	0.999	0.999	0.998	0.996	0.986	0.935	0.835	0.604	0.360	0.138	**0.061**	0.140	0.376	0.700	0.893	0.978	0.998	1.000	1.000	1.000	1.000
Q_{max}	1.000	0.999	0.998	0.993	0.977	0.920	0.793	0.570	0.329	0.128	**0.057**	0.130	0.353	0.655	0.868	0.972	0.997	0.999	1.000	1.000	1.000
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\theta = 1.5$																					
Q_1	0.999	0.998	0.989*	0.966*	0.905*	0.784*	0.612*	0.406*	0.217*	0.089*	**0.041**	0.089*	0.203*	0.409*	0.632*	0.845*	0.956*	0.988*	1.000	1.000	1.000
Q_2	0.998*	0.998	0.991	0.971	0.913	0.806	0.622	0.417	0.238	0.105	**0.056**	0.101	0.259	0.461	0.715	0.892	0.962	0.995	0.999*	1.000	1.000
Q_{max}	0.999	0.998	0.993	0.972	0.915	0.805	0.627	0.430	0.232	0.099	**0.051**	0.097	0.240	0.455	0.690	0.890	0.968	0.995	1.000	1.000	1.000
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\theta = 10.0$																					
Q_1	1.000	0.998	0.986	0.961	0.896	0.779	0.593*	0.381*	0.204*	0.089*	**0.042**	0.088*	0.216*	0.402*	0.630*	0.851	0.952	0.995	0.999*	1.000	1.000
Q_2	1.000	0.995*	0.986	0.953*	0.890*	0.770*	0.597	0.390	0.227	0.097	**0.053**	0.095	0.220	0.416	0.636	0.850*	0.950*	0.992*	1.000	1.000	1.000
Q_{max}	1.000	0.997	0.986	0.960	0.893	0.777	0.608	0.387	0.222	0.094	**0.053**	0.090	0.222	0.416	0.646	0.862	0.958	0.994	1.000	1.000	1.000
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$\mu_1 = \mu_2 = 1.5$

β (true)	-1	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	**0.0**	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\theta = 0.25$																					
Q_1	0.962*	0.930*	0.874*	0.787*	0.681*	0.548*	0.393*	0.254*	0.131*	0.072*	**0.050**	0.082*	0.153*	0.284*	0.466*	0.641*	0.801*	0.925*	0.976*	0.996*	1.000
Q_2	0.981	0.960	0.926	0.861	0.767	0.626	0.474	0.334	0.171	0.093	**0.046**	0.086	0.205	0.380	0.586	0.777	0.923	0.979	0.998	1.000	1.000
Q_{max}	0.979	0.959	0.910	0.849	0.747	0.609	0.449	0.299	0.161	0.077	**0.046**	0.087	0.185	0.357	0.572	0.762	0.907	0.981	0.999	1.000	1.000
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\theta = 1.5$																					
Q_1	0.967	0.934	0.885*	0.801	0.680*	0.536*	0.391*	0.236*	0.139*	0.082*	**0.045**	0.079*	0.159*	0.275*	0.453*	0.654*	0.818*	0.934*	0.976*	0.997	0.998*
Q_2	0.969*	0.931*	0.887	0.794*	0.693	0.547	0.392	0.248	0.161	0.093	**0.054**	0.087	0.174	0.292	0.485	0.676	0.834	0.940	0.987	0.995*	0.999
Q_{max}	0.971	0.934	0.891	0.804	0.684	0.550	0.398	0.249	0.153	0.090	**0.050**	0.083	0.172	0.289	0.480	0.683	0.834	0.943	0.986	0.997	0.999
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\theta = 10.0$																					
Q_1	0.966	0.937	0.887	0.793*	0.675*	0.536	0.379*	0.238*	0.137*	0.074*	**0.057**	0.073*	0.160*	0.294*	0.483	0.661	0.821	0.929	0.986	0.996	0.998
Q_2	0.957*	0.925*	0.877*	0.796	0.678	0.525*	0.387	0.255	0.151	0.084	**0.063**	0.087	0.168	0.301	0.482*	0.654*	0.805*	0.925*	0.979*	0.991*	0.998
Q_{max}	0.967	0.933	0.884	0.797	0.683	0.534	0.381	0.244	0.152	0.078	**0.066**	0.084	0.168	0.306	0.487	0.662	0.812	0.935	0.984	0.994	0.998
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

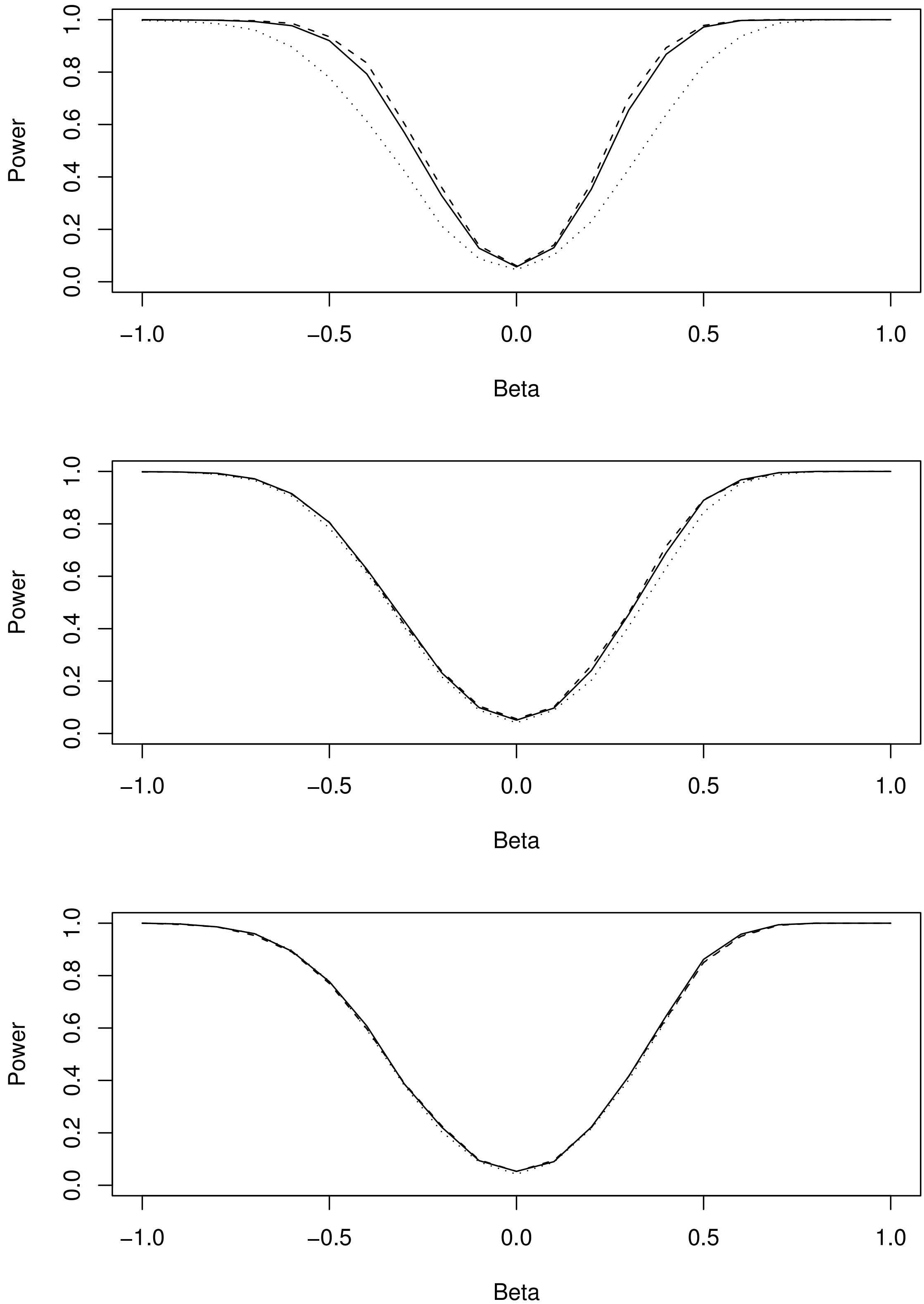

Figure 1: Comparison of powers of Q_1 (dots), Q_2 (dashes), and Q_{max} (plain) for 5% significance level under 2 dimensional Clayton model: Clayton dependence parameters $\theta = 0.25$ (upper figure), $\theta = 1.5$ and $\theta = 10$ (lower figure) with $\mu_1 = \mu_2 = 0.5$

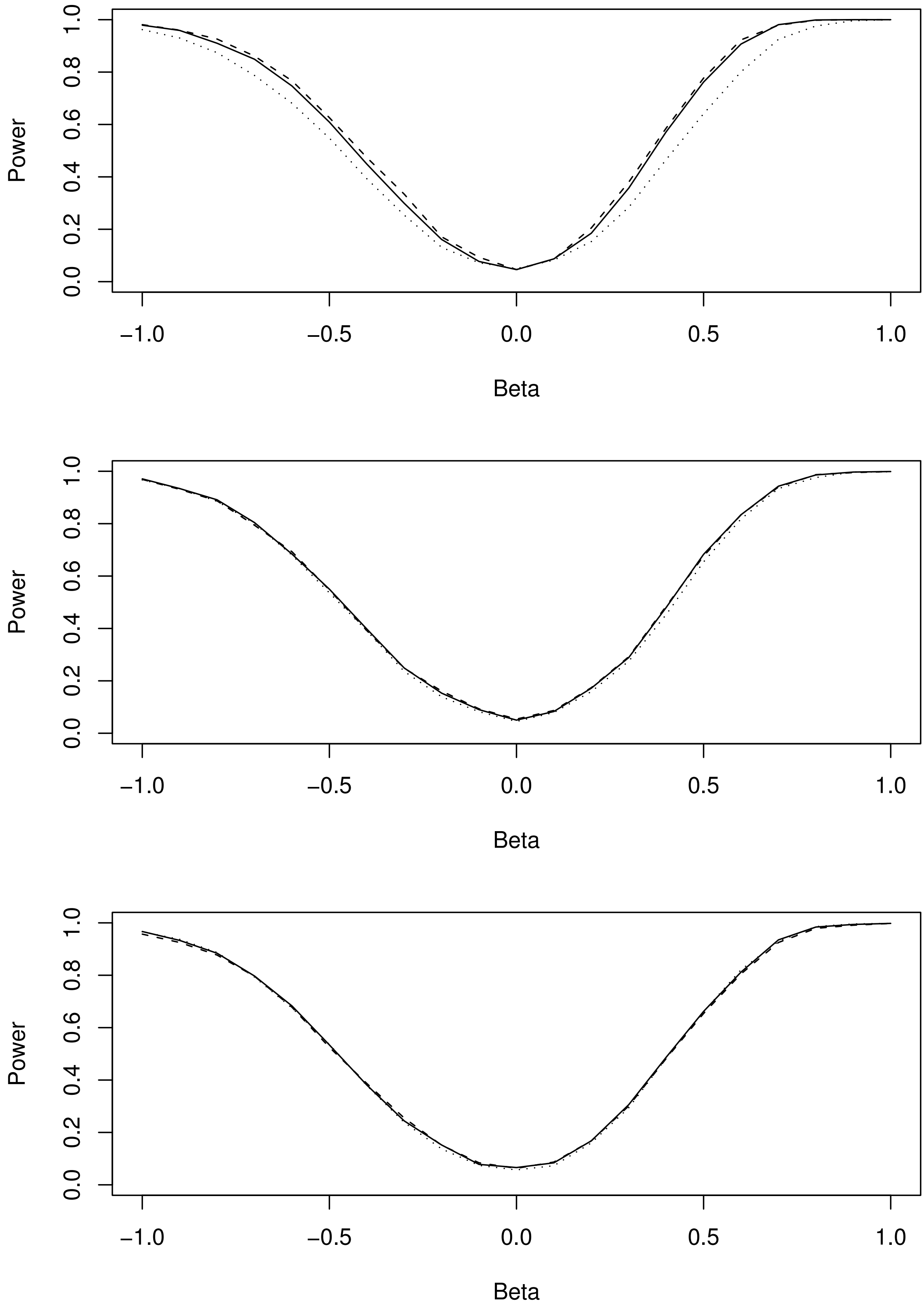

Figure 2: Comparison of powers of Q_1 (dots), Q_2 (dashes), and Q_{max} (plain) for 5% significance level under 2 dimensional Clayton model: Clayton dependence parameters $\theta = 0.25$(upper figure), $\theta = 1.5$ and $\theta = 10$ (lower figure) with $\mu_1 = \mu_2 = 1.5$

Table 3: Empirical power of tests based on Q_1, Q_2, and Q_{max} for nominal 5 % significance level under 3 dimensional Clayton model
(* *indicates the worst empirical power among* Q_1, Q_2, *and* Q_{max})

$\mu_1 = \mu_2 = \mu_3 = 0.5$

β (true)	−1	−0.9	−0.8	−0.7	−0.6	−0.5	−0.4	−0.3	−0.2	−0.1	**0.0**	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\theta = 0.25$																					
Q_1	1.000	0.999	0.999	0.996*	0.986*	0.929*	0.795*	0.589*	0.326*	0.119*	**0.045**	0.106*	0.307*	0.594*	0.842*	0.951*	0.990	0.999	1.000	1.000	1.000
Q_2	0.995*	0.994*	0.998*	0.998	0.995	0.989	0.958	0.842	0.564	0.233	**0.073**	0.221	0.590	0.887	0.974	0.990	0.983*	0.988*	0.985*	0.985*	0.982*
Q_{max}	1.000	1.000	1.000	1.000	0.999	0.993	0.955	0.814	0.526	0.202	**0.064**	0.197	0.547	0.865	0.980	0.997	1.000	1.000	1.000	1.000	1.000
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	2	1	1	0	0	1	3	2	3	4	6	4	8	5	5	8	10	5	9	8	6
$\theta = 1.5$																					
Q_1	1.000	1.000	1.000	0.997*	0.984*	0.928*	0.795*	0.562*	0.306*	0.112*	**0.046**	0.095*	0.296*	0.583*	0.823*	0.951*	0.993*	0.998	1.000	1.000	1.000
Q_2	1.000	1.000	1.000	0.998	0.988	0.934	0.812	0.614	0.364	0.145	**0.071**	0.132	0.382	0.666	0.874	0.974	0.995	0.997*	0.995*	0.997*	0.996*
Q_{max}	1.000	1.000	1.000	0.997*	0.989	0.941	0.819	0.604	0.351	0.135	**0.057**	0.126	0.358	0.661	0.885	0.978	0.997	1.000	1.000	1.000	1.000
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	0	0	0	0	0	0	1	2	0	0	0	0	0	1	0	0	0	1	2	1	2
$\theta = 10.0$																					
Q_1	1.000	1.000	1.000	0.996	0.980	0.927	0.795	0.568	0.295*	0.118*	**0.055**	0.095*	0.283*	0.573*	0.815	0.955	0.992	1.000	1.000	1.000	1.000
Q_2	0.999*	1.000	0.999*	0.993*	0.969*	0.916*	0.783*	0.552*	0.309	0.131	**0.066**	0.126	0.304	0.579	0.810*	0.942*	0.989*	0.999*	1.000	1.000	1.000
Q_{max}	1.000	1.000	1.000	0.997	0.982	0.932	0.796	0.561	0.307	0.131	**0.060**	0.121	0.297	0.586	0.818	0.956	0.990	1.000	1.000	1.000	1.000
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

$\mu_1 = \mu_2 = \mu_3 = 1.5$

β (true)	−1	−0.9	−0.8	−0.7	−0.6	−0.5	−0.4	−0.3	−0.2	−0.1	**0.0**	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\theta = 0.25$																					
Q_1	0.999	0.992*	0.974*	0.933*	0.854*	0.733*	0.560*	0.376*	0.200*	0.085*	**0.038**	0.084*	0.188*	0.407*	0.624*	0.810*	0.948*	0.985*	0.998*	1.000	1.000
Q_2	0.995*	0.994	0.989	0.979	0.945	0.858	0.720	0.517	0.298	0.145	**0.073**	0.122	0.292	0.559	0.802	0.935	0.981	0.997	0.998*	0.997*	0.999*
Q_{max}	1.000	0.997	0.992	0.978	0.937	0.836	0.700	0.468	0.278	0.121	**0.052**	0.095	0.250	0.524	0.766	0.926	0.980	0.998	1.000	1.000	1.000
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	4	2	2	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	1	1	0
$\theta = 1.5$																					
Q_1	0.996*	0.992	0.979	0.941	0.871	0.743*	0.568*	0.370*	0.193*	0.076*	**0.056**	0.084*	0.186*	0.373*	0.626*	0.832*	0.948*	0.983*	0.999	1.000	1.000
Q_2	0.996*	0.989*	0.974*	0.939*	0.867*	0.768	0.595	0.404	0.218	0.110	**0.076**	0.104	0.212	0.419	0.658	0.856	0.950	0.987	0.998*	1.000	1.000
Q_{max}	0.997	0.990	0.975	0.945	0.879	0.756	0.584	0.388	0.198	0.086	**0.066**	0.088	0.205	0.401	0.651	0.856	0.956	0.984	0.999	1.000	1.000
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\theta = 10.0$																					
Q_1	0.997	0.991	0.970	0.932	0.864	0.743	0.573	0.369*	0.173*	0.092*	**0.057**	0.096	0.190*	0.380*	0.626	0.838	0.945	0.987*	0.996	1.000	1.000
Q_2	0.992*	0.985*	0.966*	0.926*	0.840*	0.732*	0.557*	0.380	0.209	0.104	**0.066**	0.095	0.196	0.388	0.620*	0.808*	0.929*	0.988	0.994*	0.998*	1.000
Q_{max}	0.997	0.991	0.969	0.938	0.859	0.749	0.568	0.381	0.195	0.099	**0.062**	0.093*	0.198	0.384	0.632	0.834	0.940	0.988	0.996	1.000	1.000
ERR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ERR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

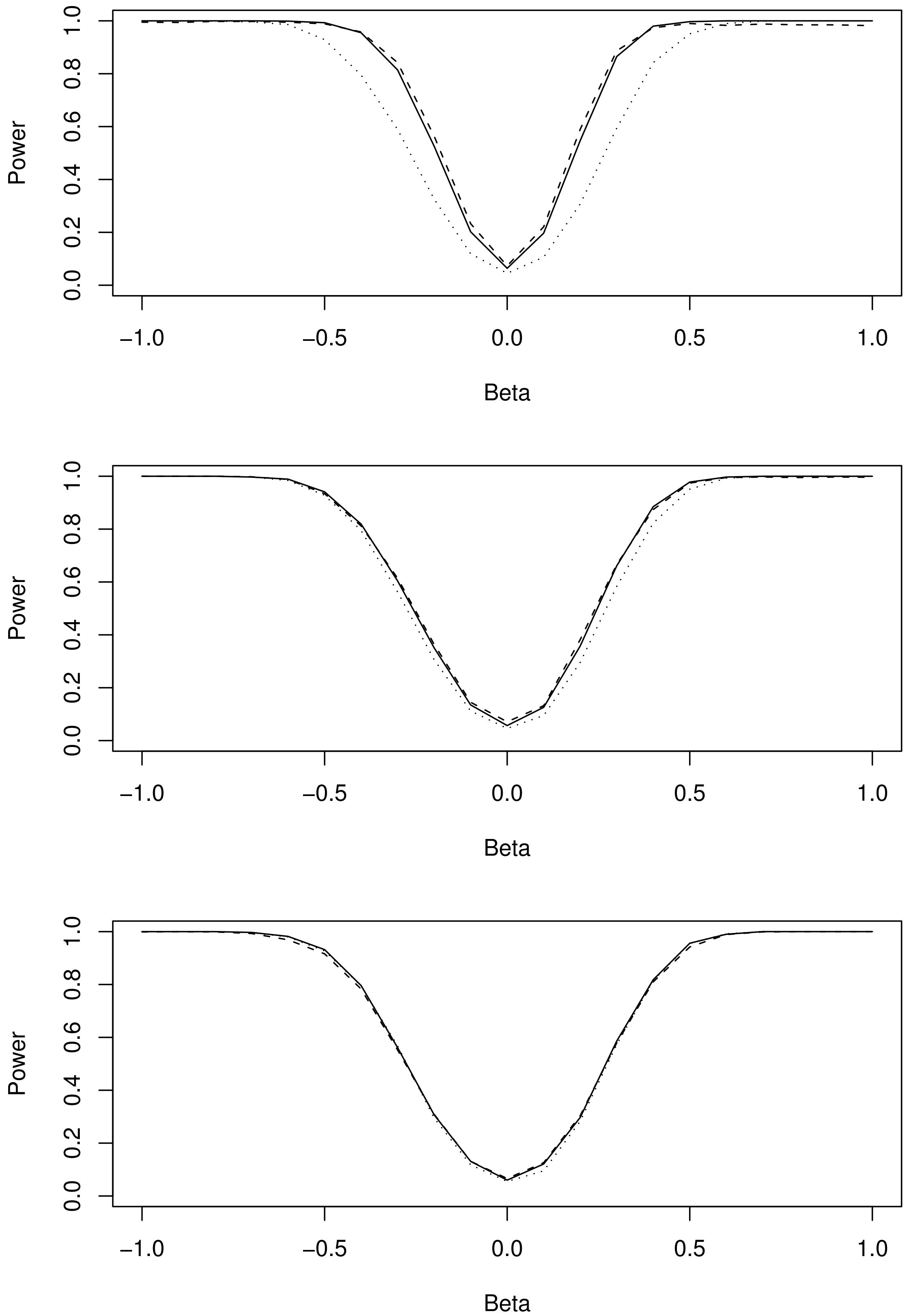

Figure 3: Comparison of powers of Q_1 (dots), Q_2 (dashes), and Q_{max},(plain) for 5% significance level under 3 dimensional Clayton model: Clayton dependence parameters $\theta = 0.25$(upper figure), $\theta = 1.5$ and $\theta = 10$(lower figure) with $\mu_1 = \mu_2 = \mu_3 = 0.5$

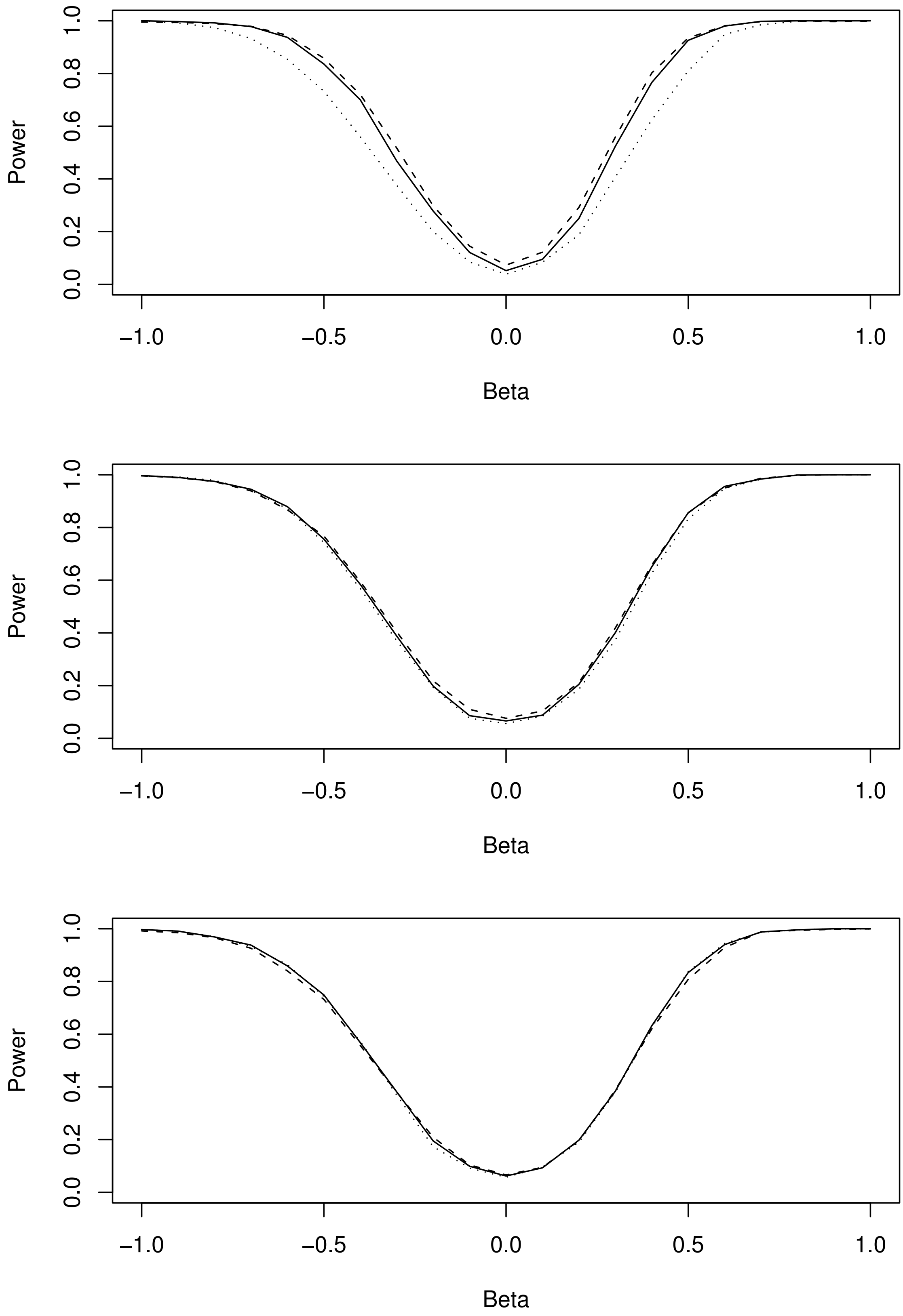

Figure 4: Comparison of powers of Q_1 (dots), Q_2 (dashes), and Q_{max},(plain) for 5% significance level under 3 dimensional Clayton model: Clayton dependence parameters $\theta = 0.25$(upper figure), $\theta = 1.5$ and $\theta = 10$(lower figure) with $\mu_1 = \mu_2 = \mu_3 = 1.5$