Testing in Aalen's linear regression model. A simulation study.

Jon Ketil Grønnesby

Abstract

The performance of tests in Aalen's linear regression model is studied using asymptotic power calculations and stochastic simulation. Aalen's original least squares test is compared to two modifications: a weighted least squares test with correct weights and a test where the variance is reestimated under the nullhypothesis. The test with the variance is reestimated provides the highest power of the tests for the setting of this paper, the gain is substantial for covariates following a skewed distribution like the exponential. It is further shown that Aalen's choice for weight function with reestimated variance is optimal in the one-parameter case against proportional alternatives.
Keywords: Aalen's linear regression model; stochastic simulation; survival analysis; weighted least squares; ordinary least squares.

1 Introduction

Aalen's (1980,1989,1993) linear regression model has proved to be useful both as a diagnostic tool for the more commonly used Cox (1972,1975) model and as an alternative model on its own. The Cox and Aalen models differ fundamentally in philosophy. The Cox model has a nonparametric baseline, but the effect of the covariates is modelled parametrically. The Aalen model on the other hand is completely nonparametric in the sense that *functions*, not parameters, are fitted. This means that the Cox model can use the whole time span in the estimation of the parameters, while the Aalen model only uses local information. Thus the Aalen model is more process oriented.

The linear model falls naturally into the theory of martingales and stochastic integrals as described by Aalen (1978), Fleming and Harrington (1991) or Andersen *et al.* (1993). Estimators and test statistics with asymptotic results are easily obtained without iteration. However the model has not been used much in practice, and the theory and diagnostic methods for the linear model are still being developed; see e.g. Henderson and Oman (1993) and Aalen (1989,1993) for recent contributions.

This paper treats tests and this admits a quick evaluation of the methods by using the power function. Tests in the Aalen model was introduced already in the original paper (Aalen, 1980). We will use one of the test statistics proposed by Aalen (1989) and two small modification of this test. One of these modifications generalises the logrank test. The performance will also be compared to some local asymptotic power results, calculated by considering sequences of local alternatives.

Our principal method of investigation will be stochastic simulation. The main simulation consists of 200 realisations of full factorial experiments, where the factors are features known or suspected to influence the power, but also factors which by analogy to the linear normal case are not suspected to influence the power (see Section 4 for details).

Section 2 briefly outlines the model while Section 3 gives expressions for the local asymptotic power and discusses optimality. Section 4 describes the simulation experiment. In Section 5 the simulation results are presented. In that Section we also try to quantify how much we gain by the modified test statistics. Section 6 contains a brief discussion.

2 The model

The model and the statistical methods are here only briefly summarized. For details see Aalen (1980,1989,1993) or Andersen *et al.* (1993). For simplicity we will only consider right censoring and fixed covariates, but the setup can easily be extended to allow both general censoring and time-dependent covariates. We observe counting processes $N_i(t), i = 1, ..., n$; counting observed event times. The intensity process of $N_i(t)$ takes the form $\lambda_i(t) = \mathbf{z}_i^T \beta(t) Y_i(t)$. Here $\mathbf{z}_i = (1, z_{i1}, ..., z_{ip})^T$ is a vector of covariates for individual i, $Y_i(t)$ is an "at risk" indicator process for this individual, and $\beta(t) = (\beta_0(t), ..., \beta_p(t))^T$ is a vector of regression functions. The parameter functions $\beta_j(t)$ are allowed to vary with time; this is different from the Cox model.

Define the "design matrix" $\mathbf{Y}$ with element $Y_{ij} = z_{ij} Y_i$ and let $\mathbf{N} = (N_1, ..., N_n)^T$. An estimator for $\mathbf{B}(t) = \int_0^t \beta(s)ds$ is $\hat{\mathbf{B}}(t) = \int_0^t J(s)\mathbf{Y}^-(s)d\mathbf{N}(s)$, where $\mathbf{Y}^-$ is a generalized inverse and $J(s) = I(\mathbf{Y}(s)$ has rank $p+1)$. This estimator is a multivariate generalization of the Nelson-Aalen estimator. One natural choice of $\mathbf{Y}^-$ is ordinary least squares inverse, $\mathbf{Y}^-(t) = (\mathbf{Y}^T(t)\mathbf{Y}(t))^{-1}\mathbf{Y}^T(t)$, proposed by Aalen (1980). However, except for the one-sample case, the variances are not equal for all N_i, and theoretically better results should be obtained using weighted least squares inverse : $\mathbf{Y}^-(t) = (\mathbf{Y}^T(t)\mathbf{W}(t)\mathbf{Y}(t))^{-1}\mathbf{Y}^T(t)\mathbf{W}(t)$, see Huffer and McKeague (1991). This resembles the linear normal case, see e.g. Weisberg (1985), but remark that here $\mathbf{W}(t)$ is a diagonal matrix with i-th diagonal element $w_i(t)$, inversely proportional to $\lambda_i(t)dt$, the conditional variance of $dN_i(t)$. Thus the weights are functions of the unknown parameters and in practice they will have to be estimated. Inference in the Aalen model is only meaningful until the stopping time τ, the first time where $\mathbf{Y}(s)$ loses full rank.

In this paper we test the regression function $\beta_p(t)$. The p parameter functions $\beta_0(t), ..., \beta_{p-1}(t)$ are considered as nuisance parameter functions. We shall consider the hypothesis $\mathbf{H}_0 : \beta_p(t) = 0$, for all t. The tests shown below are designed to be powerful against monotone alternatives, e.g. $\mathbf{H}_1 : \beta_p(t) > 0$ for all t. A natural family of test statistics is based on $X_p(\tau) = \int_0^\tau L_p(s)d\hat{B}_p(s)$ for some weight function L_p. In this paper we shall mainly consider the weight function $L_p(t) = ((\mathbf{Y}^T(t)\mathbf{Y}(t))^{-1}_{(p,p)})^{-1}$, the reciprocal of element (p,p) of $(\mathbf{Y}^T(t)\mathbf{Y}(t))^{-1}$, suggested by Aalen (1989). By convention $L_p(s)$ is set to 0 whenever $J(s) = 0$.

Define $\mathbf{B}^*(t) = \int_0^t J(s)\beta(s)ds$ and $X_p^*(t) = \int_0^t L_p(s)dB_p^*(s)$. Under mild regularity conditions, see Andersen *et al.* (1993, chapter V.2), $U_p(\tau) = \frac{X_p(\tau) - X_p^*(\tau)}{\widehat{Var}(X_p - X_p^*)(\tau)^{\frac{1}{2}}}$ is asymptotically standard normal distributed. Here $\widehat{Var}(X_p - X_p^*)(\tau)$ is some estimator for the variance of $(X_p - X_p^*)(\tau)$. A variance estimator may be derived in two ways. To this end note first that the predictable variation process for the test statistic, evaluated at τ, is

$$\langle X_p - X_p^* \rangle(\tau) = \int_0^\tau J(t)L_p^2(t)d\langle \hat{\mathbf{B}} - \mathbf{B}^* \rangle_{(p,p)}(t) = \int_0^\tau J(t)L_p^2(t)(\mathbf{Y}_p^-(t))(\text{diag}\lambda(t)dt)(\mathbf{Y}_p^-(t))^T \quad (1)$$

where diag$\lambda(t)$ is the $n \times n$ diagonal matrix with ith diagonal element $\lambda_i(t)$, and $\mathbf{Y}_p^-(t)$ is the last row of the generalized inverse $\mathbf{Y}^-(t)$. To estimate the variation process under $\mathbf{H}_1$, the vector

$\lambda(t)dt$ in (1) is replaced by

$$d\mathbf{N}(t), \tag{2}$$

yielding the optional variation process $[X_p - X_p^*](\tau)$. This is the original proposal of Aalen (1980). We will suggest a slightly different variance estimator obtained by reestimating under $\mathbf{H}_0$. For this estimator the vector, $\lambda(t)dt$ in (1) is replaced by

$$\mathbf{Y}^{(0)}(t)d\hat{\mathbf{B}}^{(0)}(t). \tag{3}$$

Here $\mathbf{Y}^{(0)}(t)$ denotes the design matrix with the last row of $\mathbf{Y}(t)$ deleted, and $\hat{\mathbf{B}}^{(0)}(t)$ is the estimate when $\mathbf{Y}^{(0)}(t)$ is the design matrix. Notice that with (2) the variance estimate is greatly simplified, at each failure time it only depends on the covariates of the individual failing at that time. For $p = 1$, (3) is

$$\mathbf{Y}^{(0)}(t)\frac{\sum dN_i(t)}{\sum Y_i(t)}, \tag{4}$$

where $\mathbf{Y}^{(0)}(t) = (Y_1(t), ..., Y_n(t))^T$, yielding the increment in the Nelson-Aalen estimator for each individual at risk. If then the covariate is dichotomous, and we use (3) in the variance estimator and Aalen's choice for weight function, we have the logrank-test for 2 samples. Using (2) in the variance estimator, the test statistic is slightly different from the logrank-test, see Andersen *et al.* (1993, Example VII.4.1).

3 Local asymptotic power and optimality

Let the setup be as in the previous section. Analogous to Andersen *et al.* (1993, Chapter V.2), we will define a sequence of local alternatives $\mathbf{H}_1^{(n)} : \beta^{(n)}(t) = (\beta_0^{(n)}(t), ..., \beta_{p-1}^{(n)}(t), \beta_p^{(n)}(t))^T = (\beta_0(s), ..., \beta_{p-1}(t), \beta_p(t)/\sqrt{n})^T$.

Define as in Andersen *et al.* (1993): $R_j(t) = \sum_i Y_{ij}(t)$; $R_{jk}(t) = \sum_i Y_{ij}(t)Y_{ik}(t)$;
$R_{jkl}(t) = \sum_i Y_{ij}(t)Y_{ik}(t)Y_{il}(t)$ and assume there exist r_j, r_{jk}, r_{jkl} the limiting expressions of $\frac{1}{n}R_j, \frac{1}{n}R_{jk}, \frac{1}{n}R_{jkl}$ (the limit must be uniform in t). Write $\mathbf{r}(t)$ for the $(p+1)\times(p+1)$ matrix with $r_{jk}(t)$ as element jk. $\mathbf{r}(t)$ is assumed to be of full rank for all t. The inverse then exists and its element jk will be written $\mathbf{r}_{jk}^{-1}(t)$. Further assume that $L_p(t)/d_n$ converges to $l_p(t)$ uniformly in t in probability for a suitable sequence $\{d_n\}$. Now: $\frac{\sqrt{n}}{d_n}X_p(\tau) = \frac{\sqrt{n}}{d_n}\int_0^\tau L_p(s)d(\hat{B}_p - B_p^{(n)*})(s) + \frac{\sqrt{n}}{d_n}\int_0^\tau L_p(s)dB_p^{(n)*}(s)$. The latter term converges in probability to $\int_0^\infty l_p(s)\beta_p(s)ds$ under the sequence of local alternatives $\mathbf{H}_1^{(n)}$. The former term is a martingale, with predictable variation process:

$$\int_0^\tau \frac{L_p^2(u)}{d_n^2} n\langle \hat{B}_p - B_p^{(n)*}\rangle(u) \xrightarrow{\mathcal{P}} \int_0^\infty l_p^2(u) \sum_{g,l,m=0:g\neq p}^{p} r_{glm}(u)\mathbf{r}_{pl}^{-1}(u)\mathbf{r}_{pm}^{-1}(u)\beta_g(u)du \tag{5}$$

(see Andersen *et al.*, 1993, Chapter VII.4.2). Thus, under the local alternatives, $\sqrt{n}X_p(\tau)/d_n$ converges to a normal distribution with mean $\int_0^\infty l_p(s)\beta_p(s)ds$ and variance equal to (5). The variance is the same as under $\mathbf{H}_0$. The standardized statistic $U_p(\tau)$ is then, for both choices of variance estimator, asymptotically Gaussian distributed with unit variance and expectation:

$$\gamma_p = \frac{\int_0^\infty l_p(s)\beta_p(s)ds}{\{\int_0^\infty l_p^2(u)\sum_{g,l,m=0:g\neq p}^{p} r_{glm}(u)\mathbf{r}_{pl}^{-1}(u)\mathbf{r}_{pm}^{-1}(u)\beta_g(u)du\}^{\frac{1}{2}}}. \tag{6}$$

The optimal choice of l_p is the one which maximises the (limiting) expection term of the standardized variable, i.e. maximises (6). For $p = 1$, using a Cauchy-Schwarz argument, $\gamma_1^2 \leq \int_0^\infty \frac{\beta_1^2(u)}{\mathbf{r}_{11}^{-1}(u)\beta_0(u)} du$, with equality if and only if $l_1(s)$ proportional to $\frac{\beta_1(s)}{\beta_0(s)\mathbf{r}_{11}^{-1}(u)}$. Thereby Aalen's choice for weight function (see Section 2) is asymptotically optimal when $\beta_1(s)$ and $\beta_0(s)$ are proportional for all s, in particular for the simulations in this paper (β_0 and β_1 are constant). For general p, we cannot find a general optimal weight without further restrictions on the covariate distribution.

We can use (6) to estimate approximate power for the test, by replacing β_p by $\beta_p/\sqrt{n}$, see Andersen *et al.* (1993, page 376). Notice, however, that we then calculate the variance under $\mathbf{H}_0$, (6) therefore tends to overestimate the power. Better approximations can be obtained by including the β_p-term in the denominator of (6), see the Appendix for details. See also Section 5.3 for a comparison betweeen observed and asymptotic power.

4 Simulations

To assess the impact of the factors, we have performed a series of 200 full factorial experiments. This means that the experiment runs through all possible factor combinations. The factors are described in Table 1 and in Section 4.1. Each test has been dichotomized to whether it is significant or not on two sided asymptotically 5 % level (i.e. we reject if $|U_p(\tau)| > 1.96$), and logistic regression has been used to evaluate the effect of the factors. The simulations have been performed on a PC using the programming system GAUSS (Aptech Inc).

In addition to the full factorial design described above, we have performed some selected simulations to evaluate the difference between the methods as a function of β_p for selected factor combinations (see Figure 1 in Section 5.2). A comparison between obtained and asymptotic power follows in Figure 2 in Section 5.3.

4.1 Choice of factors

Table 1 summerizes the factors studied in this paper. They are divided into three groups: factors describing different methods of estimation, factors to evaluate the robustness of the test statistic and factors known to be important from the power function, see the previous Section and the Appendix.

Table 1: Factors used in the simulation

Group	Factor	Levels
Method factor		
	method	WLS, OLS_0, OLS_1
Power function		
	β_0	1, 2
	β_p	0, 2.5, 5
	c (censoring rate)	$.5(\beta_0 + .5\beta_p), 2(\beta_0 + .5\beta_p)$
	n (Sample size)	50, 100
	Covariate distribution	Dichotomous, Exponential
Robustness		
	p (Nb of nuisance functions)	2, 4
	Nuisance distribution	Dichotomous, Exponential
	Correlation	0, 0.5

(for explanation of the factors, see the text)

Method factor As described above, we will evaluate the impact both of weighted least squares (called WLS) and of reestimating the variance under $\mathbf{H}_0$, (called OLS_0: see equation (3)). The methods will be compared to Aalen's original ordinary least squares method (called OLS_1, see equation (2)). WLS usually involves a two-step procedure where the weights are estimated by kernel smoothing. We have simplified the problem in these simulations by using "superweights" (i.e. the correct weights) and thereby indicating the maximum potential for the WLS. In applications the gain will of course be smaller due to the fact that the weights must be estimated. Notice that for WLS under $\mathbf{H}_0$, the weights should also be calculated under $\mathbf{H}_0$. In these simulations, with superweights and the nuisance parameters being set to 0, the weights are all equal, i.e. WLS_0=WLS_1. The two factors method and weighing have therefore been pulled together into one factor (called method), coded into 3 levels.

Factors from the power function The baseline function ($\beta_0(s)$), censoring function ($c(s)$) and the covariate function ($\beta_p(s)$ are constants in the simulations. Notice that the censoring rate is calculated as 0.5 or 2 times $\beta_0+(EZ_{ip})\beta_p$ to ensure that the number of failure is about half for high censoring compared to low censoring. Here EZ_{ip} is the expected covariate value. Furthermore, the sample size (n) is included as factors.

The covariate distributions were standardized to have the same expectation (.5) and variance (.25). We have chosen to have one distribution with high probability of extreme covariate values (exponential), and one distribution with stable covariates (dichotomous). One should therefore expect a reduced power if these extreme covariate values from the exponential influence the estimation heavily, by analogy to ordinary linear regression (see Cook and Weisberg (1982) or Weisberg (1985)).On the other hand, for dichotomous covariates we are expected to lose some information because the design matrix will in mean lose full rank with more individuals at risk.

Robustness factors McKeague and Sasieni (1994) assert that the full Aalen model is "too nonparametric" to provide good estimates for many covariate/nuisance functions. By varying the number of nuisance parameters, we can evaluate the impact of estimating unnecessary functions. The nuisance factors $(Z_{i1}^0, Z_{i2}\ldots, , Z_{i,p-1})^T$ are drawn either from the exponential or from the dichotomous distribution. We then *correlate* the covariate with the nuisance factor Z_{i1}^0 by transforming $(Z_{ip}, Z_{i1})^T = \mathbf{A}(Z_{ip}^0, Z_{i1}^0)^T$ such that $\text{Corr}(Z_{ip}, Z_{i1}) \in \{0, .5\}$ and the first row of $\mathbf{A} = (1, 0)$ (i.e Z_{ip} has the same distribution as Z_{ip}^0). The correlations $\text{Corr}(Z_{ij}, Z_{ik}) = 0$ for all other combinations j, k. Now $(Z_{i1}, ..., Z_{ip})^T$ is the covariate vector we are working on. The robustness factors are not included in the (approximated) power function (6). In ordinary linear regression inference is quite robust to these factors, at least when the nuisance parameters are not too extremely distributed. The idea is to see if the same applies here.

5 Simulation results

5.1 Logistic regression analysis

We can not expect the same effect of the factors for different values of the parameter β_p. Separate logistic regression analyses have therefore been performed for the three values of β_p. The results are reported in Tables 2-4. It is worth noting that this is not a standard statistical analysis where the aim is to assess whether the factors are "significant". We can have any factor (and any interaction term) "significant" at any level by increasing the number of simulations sufficiently. We shall therefore rather focus on the real impact of the factors. The logistic regression has been done with indicator coding of the independent variables with the lowest value as reference. This means, for instance, that the fraction of rejections if all the independent variables have the lowest value is $\exp(-3.07)/(1 + \exp(-3.07)) = .044$ if $\beta_p \equiv 0$ (see the estimate of the constant term in Table 2).

Table 2. Log odds ratios obtained in the multiple logistic regression for $\beta_p = 0$

Factor	Value	Estimate	SE
Method	WLS	0	.
	OLS_1	-.010	.040
	OLS_0	.029	.039
β_0	1	0	.
	2	-.001	.032
Censoring	Low	0	.
	High	.065	.032
Covariate distribution	Dichotomous	0	.
	Exponential	.199	.032
Sample size (n)	50	0	.
	100	.031	.032
Nb of nuisance parameters (p)	2	0	.
	4	-.025	.032
Nuisance parameter distr	Dichotomous	0	.
	Exponential	.038	.032
Correlation	0	0	.
	.5	.044	.032
Constant		-3.07	.052

When $\beta_p = 0$ (Table 2), WLS (with superweights) and OLS_1 reduce to the same test, whilst OLS_0 is slightly different. The nominal and actual levels agree quite well for all three methods. Remark the impact of the factor covariate distribution. Exponentially distributed covariates give a higher achieved level (6.40 %) than dichotomous covariates (4.62 %), indicating that the asymptotics is inferiour for the exponential covariates. (See Section 5.2 for a comment on the skewness of the test statistic).

Table 3. Log odds ratios obtained in the multiple logistic regression for $\beta_p = 2.5$

Factor	Value	Estimate	SE
Method	WLS	0	.
	OLS_1	-.281	.021
	OLS_0	.377	.021
β_0	1	0	.
	2	-1.343	.018
Censoring	Low	0	.
	High	-1.274	.018
Covariate distribution	Dichotomous	0	.
	Exponential	-1.056	.018
Sample size (n)	50	0	.
	100	1.472	.018
Nb of nuisance parameters (p)	2	0	.
	4	-.096	.017
Nuisance parameter distr	Dichotomous	0	.
	Exponential	.004	.017
Correlation	0	0	.
	.5	.029	.017
Constant		1.608	.028

Table 4. Log odds ratios obtained in the multiple logistic regression for $\beta_p = 5$

Factor	Value	Estimate	SE
Method	WLS	0	.
	OLS_1	-.440	.025
	OLS_0	.383	.025
β_0	1	0	.
	2	-1.176	.022
Censoring	Low	0	.
	High	-1.601	.023
Covariate distribution	Dichotomous	0	.
	Exponential	-1.543	.022
Sample size (n)	50	0	.
	100	1.929	.024
Nb of nuisance parameters (p)	2	0	.
	4	-.124	.021
Nuisance parameter distr	Dichotomous	0	.
	Exponential	.009	.021
Correlation	0	0	.
	.5	.006	.020
Constant		3.171	.037

Tables 3 and 4 give similar results, and will therefore be discussed together. The factors censoring and sample size determine the number of events, and it is no surprise that these influence the power heavily. A high value of the factor β_0 is "washing out" the covariate effect, and therefore has an equally strong influence. Notice that the reduction in power for exponential covariates is of the same order as these other factors. This is due to the fact that with exponential covariates the high risk individuals will in mean fail earlier, and for the remaining individuals, the difference in covariate values will be lower. The big difference in poser is confirmed by asymptotic results, see Figure 2 and the Appendix.

We see that the OLS_0 give higher power than WLS, both for $\beta_p = 2.5$ and $\beta_p = 5$. The gain, however, differs for the two covariate distributions. For illustration, consider $\beta_p = 2.5$, stratified on the covariate. For dichotomous covariates the estimates for OLS_1 and OLS_0 are -.051 and .100 respectively, for exponential covariates -.484 and .597. This illustrates that the three methods are quite similar for dichotomous covariates. For exponential covariates, however, a quite substantial improvement in power is obtained by OLS_0 versus WLS, and by WLS versus OLS_1. A more detailed assessment of the real impact of the factor 'method' is done in Section 5.2.

The robustness factors do not influence the power much, and except 'p' they are even within the random variation. This is comparable to the experience with ordinary linear regression. Therefore we may use the case without nuisance parameters as an adequate approximation for the case with nuisance parameters in the asymptotic power calculations (see the Appendix).

5.2 What is the impact of the factor 'method'?

In Figure 1 we compare the estimated power functions for the three methods described in this paper. 5000 simulations have been performed for each value of β_p and each of the the methods. In the simulations the covariate distribution and the censoring rate are varied, and the other factors are left constant as: $\beta_0 = 1$; $n = 50$; no correlation; $p = 2$; the distribution of the confounders is exponential. With $n = 50$, one could expect about 17 and 33 failures for high and low censoring respectively. The asymptotic power may therefore be inadequate, at least for high censoring.

The simulations confirm that for dichotomous covariates, the difference between the methods is negligable, even with high censoring. For the combination high censoring and exponential covariates, however, the difference between the methods is astonishingly high, even for quite low values of β_p. This is partly due to the heavy left skewness of these test statistics for exponential covariates. The skewness also explains why the lowest power is obtained for $\beta_p > 0$ for the methods OLS_1 and WLS. Note also that OLS_0 is right skewed for exponential covariates, but the skewness here is much less pronounced.

Further analysis of the test statistics (not displayed) shows that there is a substantial difference in the empirical standard deviation (SD) of the three tests outside $\mathbf{H}_0$. Both OLS_1 and WLS have empirical SD decreasing with β_p (as low as .55 for OLS_1 and $\beta_p = 5$); whilst OLS_0 have empirical SD consistently just above 1. The low SD for OLS_1 and WLS may be due to dependence in the estimation of the nominator and denominator of the standardized test U_p (see Section 3).

5.3 Simulated and asymptotic power functions

In Figure 2 the actual and nominal power for OLS_1 has been plotted as functions of β_p for the same factor setup as in Figure 1. As nominal power, we have considered both the local asymptotic power (called 'local' below) and the formula including the β_p-term in the denominator (called 'asymptotic' below). For both. we have used the simplified versions in formulas (8,9), which omit the nuisance parameters except the baseline, even if the simulations include one additional nuisance parameter. This is because we found in Section 5.1 that a few more nuisance parameters does not affect the power much. The power (8,9) has been calculated numerically.

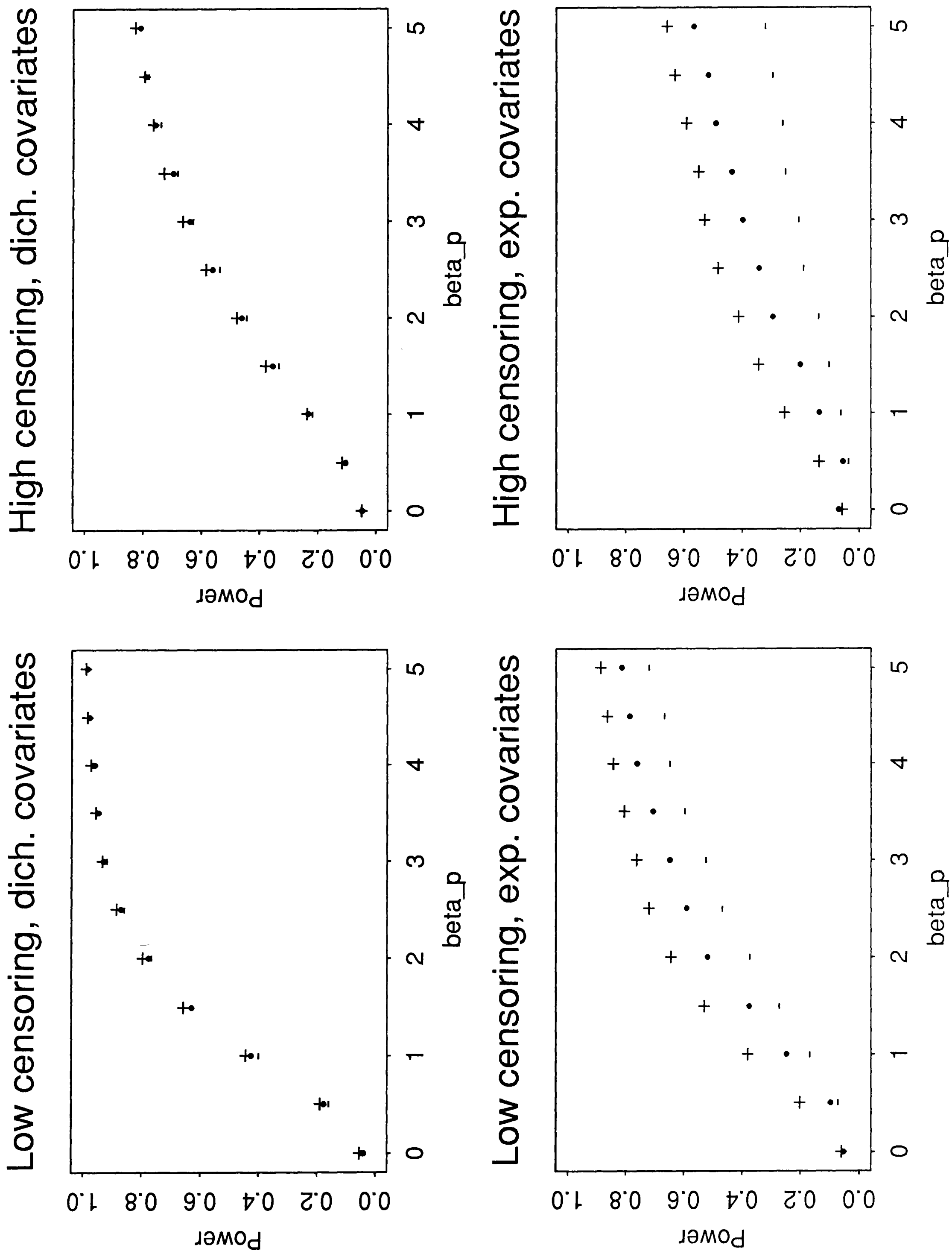

Figure 1: Power functions for the three test statistics $OLS_0(+)$ OLS_1 (-) and WLS (.)

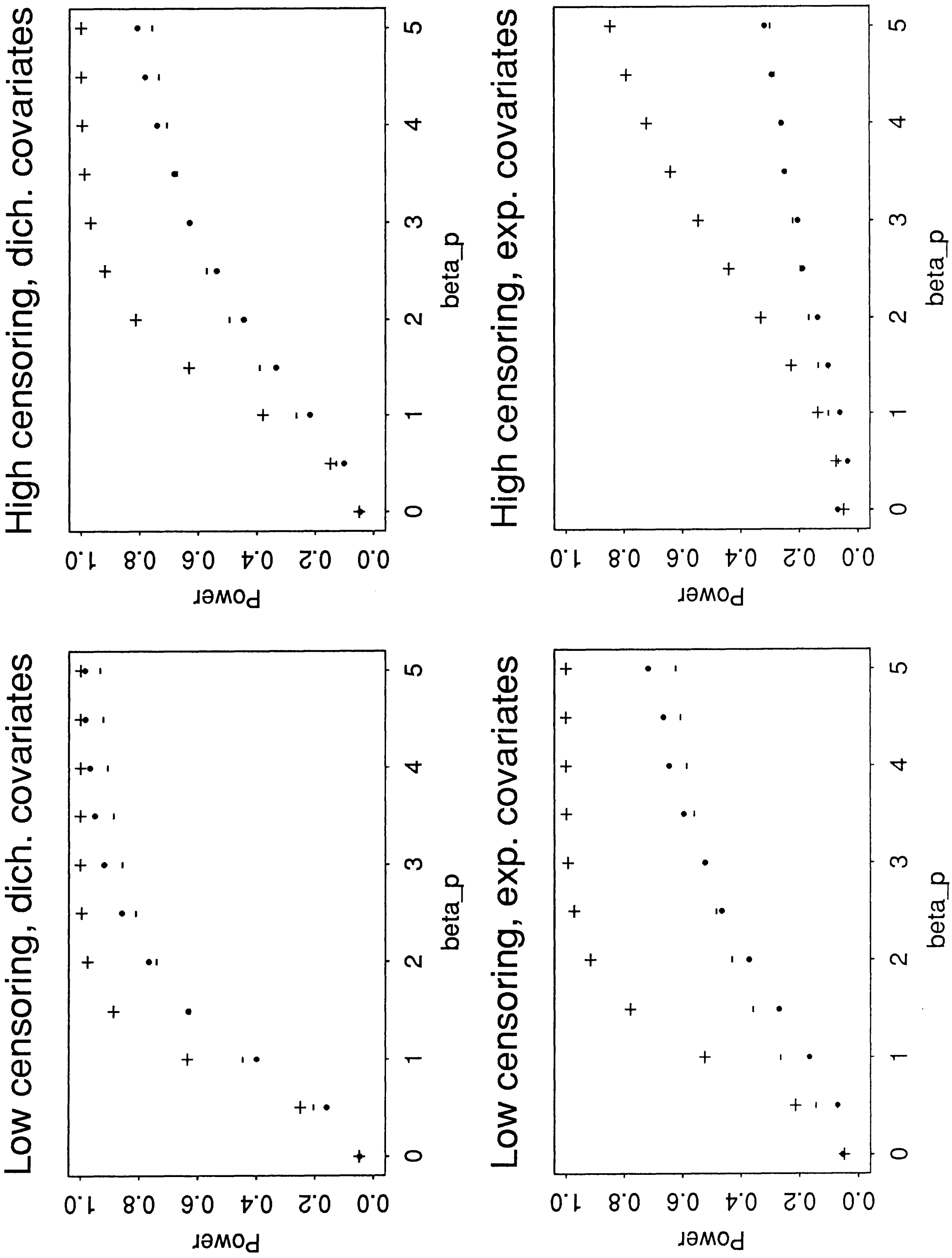

Figure 2: Power function for the OLS_1 (.) compared to local (+) and asymptotic (-)

Here, the local power is generally substantially overestimating the power, because the variance is underestimated when the β_p-term is neglected. The asymptotic formulas (i.e. 8,9), on the other hand, approximate the observed power well. Remark that the asymptotic power is slightly underestimated for big values and slightly underestimated for small values of β_p.

6 Discussion

This paper suggests that Aalen's original OLS_1-statistic can be improved both by using WLS methods and by modifying the variance estimator, yielding OLS_0. The simulations indicate that, at least for some situations, there is more to gain by changing the variance estimator than weighing. The potential gain of weighing is theoretically most pronounced when the interindividual difference in hazard is big.

In practice, the weights have to be estimated, and that the gain displayed here is overoptimistic. Huffer and McKeague (1991) suggest that weights should be determind by nonparametric kernel estimation. As an alternative one could estimate more robust weights by fitting the semiparametric or partly parametric models (see Lin and Ying (1994) and McKeague and Sasieni (1994)) in a first step, and then use these weights in the second step.

We have in this paper only considered individual processes (i.e. $Y_i(t) \in \{0,1\}$). In many practical situations the covariates are categorical (or grouped) and it can be useful to consider the aggregated data such that N_i counts the number of events and Y_i the number at risk in group i. This will speed up the computation, and will also enable analyses which would not be possible with the traditional setup because of memory problems. Here we should use Y_i as weighing factor. Remark in particular that some simplifications in the expressions for local and asymptotic power from Section 3 no longer are possible.

Appendix: Asymptotic power function for $p = 1$

The asymptotic power function is complicated for general p. However the number of nuisance parametes does not seem to be very influencial for the power, at least when $\beta_1 = \ldots = \beta_{p-1} = 0$ (see Section 5.1). We will therefore do calculations for $p = 1$ and use this as an approximation (upper limit) for general p. We will only consider Aalen's choice for weight function, i.e. $L_p(t) = R_{pp}(s)^{-1} = ((\mathbf{Y}^T(t)\mathbf{Y}(t))^{-1}_{(p,p)})^{-1}(t)$. For $p = 1$ the limit is : $l_1(t) = \frac{\Delta(t)}{r_{00}(t)}$, where $\Delta(s) = r_{11}(s)r_{00}(s) - r_{10}^2(s)$.

Using (6) and including the β_1-term (see Section 3) we get:

$$\gamma_1 = \frac{\sqrt{n}\int \beta_1(s)\frac{\Delta(s)}{r_{00}(s)}ds}{\{\int \frac{\Delta^2(s)}{r_{00}^2(s)}(\frac{r_{00}(s)}{\Delta(s)}\beta_0(s) + \frac{1}{\Delta^2(s)}(r_{111}(s)r_{00}^2(s) + r_{10}^3(s) - 2r_{11}(s)r_{10}(s)r_{00}(s))\beta_1(s))ds\}^{\frac{1}{2}}}. \tag{7}$$

This expression depends heavily on the covariate distribution. Simplifications of (7) can be made when $\beta_0(s), \beta_1(s)$ and $c(s)$ are constant, yielding:

- Dichotomous covariates:

$$\gamma_1 = \frac{\frac{\beta_1\sqrt{n}}{2}\int_0^\infty \frac{e^{-(\beta_0+c)s}}{1+e^{\beta_1 s}}ds}{\{\frac{1}{2}\beta_0\int_0^\infty \frac{e^{-(\beta_0+c)s)}}{(1+e^{\beta_1 s})^2}ds + \frac{1}{2}\beta_1\int_0^\infty \frac{e^{-(\beta_0+c)s}}{(1+e^{\beta_1 s})(1+e^{-\beta_1 s})}ds\}^{\frac{1}{2}}}. \tag{8}$$

- Exponential covariates:

$$\gamma_1 = \frac{2\beta_1\sqrt{n}\int_0^\infty \frac{e^{-(\beta_0+c)s}}{(\beta_1 s+2)^3}ds}{\{\beta_0\int_0^\infty \frac{2e^{-(\beta_0+c)s}}{(\beta_1 s+2)^3}ds + \beta_1\int_0^\infty \frac{6e^{-(\beta_0+c)s}}{(\beta_1 s+2)^4}ds\}^{\frac{1}{2}}}. \tag{9}$$

The expressions (8- 9) have been calculated numerically in Figure 2 .
In separate analyses (not displayed) one sees that the predictable variation tend to overestimate the variance under $\mathbf{H}_1$.

Acknowledgement I want to thank my advisers Odd O Aalen and Ørnulf Borgan for many helpful discussions. Their patience and encouragement has been invaluable in the preparation of this paper.

References

Aalen O.O. (1978). Non-parametric inference for a family of counting processes. *Ann Statist* 6,701–726.

Aalen O.O. (1980). A model for nonparametric regression analysis for counting processes. *Springer Lect Notes in Statist.* **2**, 1–25.

Aalen O.O. (1989). A linear model for the analysis of life times. *Statist in Med* **8**,907-925.

Aalen O.O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statist in Med* **12**,1569–1588.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes.* Springer Verlag, New York.

Cook R.D., Weisberg S. (1982). *Residuals and Influence in regression.* Chapman and Hall.

Cox D.R. (1972). Regression models and life tables (with discussion). *J Roy Statist Soc B* **34**,187–220.

Cox D.R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

Fleming T.R., Harrington D.P.(1991). *Counting processes and survival analysis.* Wiley, New York.

Henderson R., Oman P. (1993). Influence in linear hazard models. *Scand J Statist* **20**, 195–212.

Henderson R., Milner A. (1991). Aalen plots under proportional hazards. *Appl Statist* **40**, 401–409

Hosmer D.W., Lemeshow S. (1989). *Applied logistic regression.* Wiley, New York.

Huffer F.W., McKeague I.W. (1991). Weighted least squares estimation for Aalen's additive risk model. *J Amer Statist Assoc* **86**, 114–129 .

McKeague I.W., Sasieni P.D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501–14.

Lin D.Y., Ying Z. (1994) Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–72.

Weisberg S. (1985). *Applied linear regression.* Wiley, New York.