Hamerle, Moller:

Semiparametric EM-estimation of censored linear regression models for durations

Sonderforschungsbereich 386, Paper 15 (1996)

Online unter: http://epub.ub.uni-muenchen.de/

Semiparametric EM-estimation of censored linear regressionmodels for durations

ALFRED HAMERLE
Department of Statistics
University of Regensburg
93040 Regensburg
Germany

MICHAEL MOLLER
Faculty of Economics and Statistics
University of Konstanz
78434 Konstanz
Germany

This paper investigates the sensitivity of maximum quasi likelihood estimators of the covariate effects in duration models in the presence of misspecification due to neglected heterogeneity or misspecification of the hazard function. We consider linear models for $r(T)$ where T is duration and r is a known, strictly increasing function. This class of models is also referred to as location-scale models. In the absence of censoring, Gould and Lawless (1988) have shown that maximum likelihood estimators of the regression parameters are consistent and asymptotically normally distributed under the assumption that the location-scale structure of the model is of the correct form. In the presence of censoring, however, model misspecification leads to inconsistent estimates of the regression coefficients for most of the censoring mechanisms that are widely used in practice. We propose a semiparametric EM-estimator, following ideas of Ritov (1990), and Buckley and James (1979). This estimator is robust against misspecification and is highly recommended if there is heavy censoring and if there may be specification errors. We present the results of simulation experiments illustrating the performance of the proposed estimator.

KEY WORDS: Censored linear regression models; accelerated failure time models; misspecified models semiparametric EM-estimation; simulation study.

1 Introduction

In recent years the increasing availability of event history or failure time data in biomedical, economic and social science research has led to the widespread application of continuous time hazard rate models. The main purpose of the statistical analysis is to evaluate the association of exposure, treatment and prognostic factors with the distribution of time until the failure or a certain event occurs, and, by analogy with conventional regression analysis, much of the attention in the statistical analysis of failure time data focuses on the effect of covariates on durations. The statistical theory of duration data is described by Kalbfleisch and Prentice (1980), Blossfeld, Hamerle and Mayer (1989), Lancaster (1990) and others. The most frequently used methods for estimating the model parameters begin by specifying the hazard function, i.e. the probability distribution of the duration, up to a finite set of parameters, after which the values of the regression coefficients and any further unknown parameters are estimated by maximum likelihood. Unfortunately, theoretical considerations in biomedical as well as in economic and social science research rarely provide guidance on how the duration distribution should be specified, and there are no readily available methods for identifying the appropriate specifications from data. Therefore, there may be some misspecification, and it is important to find out wether the parameter estimates may or may not be affected by this misspecification. We shall only discuss distributional violations. Two important sources of misspecifications are the functional form of the hazard function and neglected heterogeneity. The latter refers to differences remaining in distribution after controlling for the effects of the included covariates. If such unmeasured heterogeneity is important, then neglecting it is to commit a serious error, and may result in spurious negative duration dependence as well as in inconsistency of parameter estimates (see, e.g., Heckman and Singer, 1984, Vaupel and Yashin, 1985, Blossfeld, Hamerle and Mayer, 1989, and others). Second, parameter estimates may also be sensitive to the special functional form assumed for the distribution of the heterogeneity when unobserved heterogeneity is explicitly included in the model by assuming the presence of unobservable random factors in the hazard function represented by an additional random variable (see, e.g., Manton, Stallard and Vaupel, 1986, Newman and McCulloch, 1984, Trussell and Richards, 1985, Kiefer, 1988). Third, the parameter estimates may be severely affected as well if the individual duration distribution (given the heterogeneity component), or both the individual hazard and the mixing distribution of the heterogeneity are misspecified. In summary, we conclude that in the literature there are different opinions to this topic. In the present paper we focus on maximum quasi likelihood estimation of the regression coefficients in censored linear regression models for durations. This class contains linear models for r(T) where r(T) is a strictly increasing function of the duration T. We only consider the case where r(T) is a known function. This class of models is also referred to as location-scale models. The most widely used special case is r(T) = log T leading to the well-known accelerated failure time models. In the absence of censoring, Gould and Lawless (1988) have shown that maximum quasi likelihood estimators of the regression

parameters are consistent and asymptotically normally distributed under the assumption that the location-scale structure of the model is of the correct form. Hence, reasonable estimates of the regression parameters are obtained regardless wether unobserved heterogeneity is present or not, and regardless wether the distribution of the error term is correctly specified or not. The scale parameter that determines duration dependence cannot be correctly estimated if any misspecification occurs. It is extremely sensitive to both functional form misspecification and omission of important variables or misspecification of included variables. In the presence of censoring, model misspecification leads to inconsistent estimates of the regression coefficients for most of the censoring mechanisms that are widely used in applications. In the present paper we propose a semiparametric EM-estimator, following an idea of Ritov (1990) and Buckley and James (1979) that is robust against misspecification. If there is heavy censoring and if there may be specification errrors, this estimator is highly recommended. For normal error distribution the well-known Buckley-James estimator is obtained, but other error distributions may be used as well leading to different semiparametric estimators. In the last section we present the results of some simulation experiments illustrating the performance of the proposed estimators.

2 Censored linear regression models for durations and semiparametric estimation

A typical censored linear regression model we shall consider in the present paper is of the form

$$y = \min\left(r\left(T\right), r\left(C\right)\right) \tag{1}$$

where

$$r\left(T\right) = \alpha_0 + x^{'}\beta_0 + \sigma\varepsilon \tag{2}$$

y is the observed dependent variable, T is the duration which may be observed completely or is censored otherwise, C is the maximum observable value of T if T is censored from above, r is an increasing function that may or may not be known (here we always assume that r is a known strictly increasing function everywhere on the support of T), x is a vector of explanatory variables and β_0 a correspondig vector of unknown parameters, α_0 is an intercept, $\sigma > 0$ is a scale parameter, and ε is a random error whose probability distribution may or may not be known. It is assumed that ε is distributed independently of x. This class of censored linear regression models is sometimes referred to as 'generalized accelerated failure time models' (Ridder, 1990). Quasi-maximum-likelihood (QML) estimation of the regression parameters is based on the assumed distribution $f(\varepsilon)$ of the error term. In this case the criterion function to be maximized is

$$l_{QML}\left(\alpha, \beta, \sigma\right) = \sum_{i=1}^{n}\left[\delta_i \log f\left(\frac{y_i - \alpha - x_i^{'}\beta}{\sigma}\right) + \left(1 - \delta_i\right)\log\left(1 - F\left(\frac{y_i - \alpha - x_i^{'}\beta}{\sigma}\right)\right)\right]$$

where δ_i is the censoring indicator. On the other hand, the true density and distribution functions $f_0(\varepsilon)$ and $F_0(\varepsilon)$ are unknown. In the presence of misspecification and censoring the estimator of the structural parameters are seriously biased. Some simulation results concerning the relationship between the estimated regression parameter $\hat{\beta}$ based on the assumed QML function and the true parameters β_0 of the density function $f_0(.)$ underlying the data are presented in Moller (1994).

In order to avoid biased estimates the main interest of this paper is the behaviour of a semiparametric EM - estimator (SPEM) when the true survival distribution is unknown. This estimator is defined as follows. Let θ consist of the regression parameters and additional nuisance parameters such as the intercept α and the scale parameter σ. Then, the SPEM estimator $\hat{\theta}$ maximizes by iteration the criterion function

$$l_{SP}(\theta) = \sum_{i=1}^{n} \left[\delta_i \log f(\varepsilon_i, \theta) + (1-\delta_i) \frac{\sum_{j=1}^{n} I_{\{\varepsilon_j > \varepsilon_i\}} \log f(\varepsilon_j, \theta)\, d\hat{F}(\varepsilon_i, \theta_A)}{1 - \hat{F}(\varepsilon_j, \theta_A)} \right] \tag{3}$$

where θ_A denotes the parameter value of the preceding iteration step, $I_{\{.\}}$ is an indicator function, and $d\hat{F}$ are the jumps of the Kaplan-Meier-estimator of the distribution function of the residuals

$$\hat{F}_{KM}(\varepsilon_i) = 1 - \prod_{\varepsilon < \varepsilon_i} \frac{n-i}{n-i+1}$$

Moreover, we define (with $\epsilon \approx 0$)

$\left|\theta_A - \hat{\theta}\right| > \epsilon \rightarrow \theta_A = \hat{\theta}$ and repeat (3)

$\left|\theta_A - \hat{\theta}\right| \leq \epsilon \rightarrow \theta_A = \hat{\theta}$

The semiparametric EM-estimator from (3) is a generalization of the Buckley - James estimator. The Buckley-James estimator is the special case of (3) where $f(.)$ is the density function of the normal distribution, and hence for the score function $s(\varepsilon) = \frac{f'(\varepsilon)}{f(\varepsilon)}$ we have $s(\varepsilon) = \varepsilon$. The SPEM estimation procedure presented here can also be used for other error densities. The estimator can be motivated as follows:

If the nonparametric estimation of the distribution function $\hat{F}_0$ is replaced by its theoretical analogon F_0, the expectation of the SPEM-criterion function, given the covariates x, yields (with the unknown distribution G(.) of the censored survival values C)

$$\begin{aligned} E[l_{SP}(\theta)] &= E_T\left[E_{Y|T}[\delta \log f(t)]\right] + E_C\left[E_{Y|C}\left[(1-\delta) E_{T|T>C,C=t}[\log f(t)]\right]\right] \\ &= \int_{-\infty}^{\infty} (1 - G(t)) \log f(t, \theta) f_0(t)\, dt + \int_{-\infty}^{\infty} \left[\int_t^{\infty} \log f(s, \theta) f_0(s)\, ds\right] g(t)\, dt \end{aligned}$$

Taking integration by parts of the second term, the partial derivatives with respect to θ can be written as

$$\begin{aligned} \frac{\partial}{\partial \theta} E[l_{SP}(\theta)] &= \int_{-\infty}^{\infty} (1 - G(t)) \frac{\partial}{\partial \theta} \log f(t, \theta) f_0(t)\, dt \\ &\quad + \left[G(t) \int_t^{\infty} \frac{\partial}{\partial \theta} \log f(s, \theta) f_0(s)\, ds\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -\frac{\partial}{\partial \theta} \log f(t, \theta) f_0(t) G(t)\, dt \end{aligned}$$

The second term vanishes because its integral and the distribution function G(.) are zero for $t = \infty$ and $t = -\infty$, respectively, and are bounded otherwise.
Hence we have

$$\begin{aligned}\frac{\partial}{\partial\theta}E\left[l_{SP}\left(\theta\right)\right] &= \int_{-\infty}^{\infty}\left(1-G\left(t\right)\right)\frac{\partial}{\partial\theta}\log f\left(t,\theta\right)f_0\left(t\right)dt + \int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\log f\left(t,\theta\right)G\left(t\right)f_0\left(t\right)dt \\ &= \int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\log f\left(t,\theta\right)f_0\left(t\right)dt = E\left[\frac{\partial}{\partial\theta}\log f\left(t,\theta\right)\right]\end{aligned}$$

It turns out that the last term is identical with the result of the maximization condition of QML-functions for linear models without censoring. However, assuming the covariates beeing stochastic, the estimates of these models are consistent for β_0 even in the presence of misspecification (Gould and Lawless 1988), as already mentioned.

Under certain assumptions, Ritov (1990) has shown that the SPEM- estimator of the regression parameters is asymptotically equivalent to the rank regression estimator proposed by Tsiatis (1990) and, therefore, shares its asymptotic properties.

3 An application: Semiparametric EM estimation with Weibull density

In this section we consider an application of the semiparametric estimator. Up to now the single application used in practice is the Buckley - James estimator. This estimator is based on the Lognormal distribution for the duration and is described as a special case of a SPEM-estimator, for example, in Moller (1994). Here we consider another application assuming a Weibull model for the duration. In the next section we present some simulation results. Let us assume that $r(T) = \log(T)$ and that the error term in (2) has an extreme value distribution with survivor function

$$S\left(\varepsilon\right) = \exp\left(-\exp\left(\varepsilon\right)\right)$$

corresponding to a Weibull model for the duration T. For simplicity, the intercept α is absorbed into the parameter vector β assuming that the first component in x is 1. Some calculations concerning consecutive iteration steps of the EM - algorithm are in order. Let β_A, σ_A denote the values of the parameters β and σ obtained in the preceding iteration step. In the actual M-step β_A and σ_A are considered as constants. Moreover, let $\Delta_A = \beta_A - \beta$ and $\varepsilon_{iA} = \frac{\left(y_i - x_i'\beta_A\right)}{\sigma_A}$. The residuals ε_{iA} are used for computing the Kaplan-Meier weights and the expectation. The following relationship is useful

$$E\left[\frac{\left(y_i - x_i'\beta\right)}{\sigma}\,|\beta_A, \varepsilon_A > \varepsilon_{iA}\right] = \frac{\sigma_A}{\sigma}E\left[\frac{\left(y_i - x_i'\beta_A\right)}{\sigma_A}\,|\varepsilon_A > \varepsilon_{iA}\right] + \frac{x_i'\Delta_A}{\sigma}$$

The criterion function depends on the 'old' parameter values as well as on the parameters that maximize the criterion. For the Weibull model the SPEM-criterion function is given by

$$\begin{aligned} l_{SP}(\beta,\sigma) &= \sum_{i=1}^{n} -\log(\sigma) + \delta_i(\varepsilon_i - \exp(\varepsilon_i)) + (1-\delta_i) \\ &\left(\frac{\sigma_a}{\sigma} E[\varepsilon_A \,|\varepsilon_A > \varepsilon_{iA}] + \frac{x_i'\Delta_A}{\sigma} - E\left[\exp\left(\frac{\sigma_a}{\sigma}\varepsilon_A\right)|\varepsilon_A > \varepsilon_{iA}\right]\exp\left(\frac{x_i'\Delta_A}{\sigma}\right)\right) \end{aligned} \tag{4}$$

For the numerical computations, in particular for Newton's method in the M-step, the following derivatives are needed

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \sum_{i=1}^{n} \frac{\delta_i x_i}{\sigma}(\exp(\varepsilon_i) - 1) + (1-\delta_i)\frac{x_i}{\sigma} \\ &\left(E\left[\exp\left(\frac{\sigma_a}{\sigma}\varepsilon_a\right)|\varepsilon_A > \varepsilon_{iA}\right]\exp\left(\frac{x_i'\Delta_A}{\sigma}\right) - 1\right) \\ \frac{\partial l}{\partial \sigma} &= \sum_{i=1}^{n} -\frac{1}{\sigma} - \delta_i\frac{\varepsilon_i}{\sigma}(1-\exp(\varepsilon_i)) - (1-\delta_i)\frac{1}{\sigma} \\ &\left(\frac{\sigma_a}{\sigma}E[\varepsilon_a\,|\varepsilon_A > \varepsilon_{iA}] + \frac{x_i'\Delta_A}{\sigma} - \frac{\sigma_a}{\sigma}E\left[\varepsilon_a\exp\left(\frac{\sigma_a}{\sigma}\varepsilon_a\right)|\varepsilon_A > \varepsilon_{iA}\right]\exp\left(\frac{x_i'\Delta_A}{\sigma}\right)\right. \\ &\left. -E\left[\exp\left(\frac{\sigma_a}{\sigma}\varepsilon_a\right)|\varepsilon_A > \varepsilon_{iA}\right]\exp\left(\frac{x_i'\Delta_A}{\sigma}\right)\frac{x_i'\Delta_A}{\sigma}\right) \\ \frac{\partial^2 l}{\partial\beta\partial\beta'} &= \sum_{i=1}^{n} -\frac{x_i x_i'}{\sigma^2}\left(\delta_i\exp(\varepsilon_i) - (1-\delta_i)E\left[\exp\left(\frac{\sigma_a}{\sigma}\varepsilon_a\right)|\varepsilon_A > \varepsilon_{iA}\right]\exp\left(\frac{x_i'\Delta_A}{\sigma}\right)\right) \end{aligned}$$

The second derivatives are given by

$$\begin{aligned} \frac{\partial^2 l}{\partial\sigma\partial\beta'} &= \sum_{i=1}^{n} \frac{\delta_i}{\sigma}\left(\frac{x_i}{\sigma}(1-\exp(\varepsilon_i)) - \frac{x_i\varepsilon_i}{\sigma}\exp(\varepsilon_i)\right) + (1-\delta_i) \\ &\frac{x_i}{\sigma^2}\left(1 - \frac{\sigma_a}{\sigma}E\left[\varepsilon_a\exp\left(\frac{\sigma_a}{\sigma}\varepsilon_a\right)|\varepsilon_A > \varepsilon_{iA}\right]\exp\left(\frac{x_i'\Delta_A}{\sigma}\right)\right. \\ &\left. -E\left[\exp\left(\frac{\sigma_a}{\sigma}\varepsilon_a\right)|\varepsilon_A > \varepsilon_{iA}\right]\exp\left(\frac{x_i'\Delta_A}{\sigma}\right)\left(1 + \frac{x_i'\Delta_A}{\sigma}\right)\right) \\ \frac{\partial^2 l}{\partial\sigma^2} &= \sum_{i=1}^{n} \frac{1}{\sigma^2} + \frac{\delta_i\varepsilon_i}{\sigma}\left(2\frac{1-\exp(\varepsilon_i)}{\sigma} - \frac{\exp(\varepsilon_i)\,\varepsilon_i}{\sigma}\right) \\ &-\frac{1-\delta_i}{\sigma^2}\left(\frac{\partial l}{\partial\sigma} - \frac{\sigma_a}{\sigma}E[\varepsilon_a\,|\varepsilon_A > \varepsilon_{iA}] - \frac{x_i'\Delta_A}{\sigma}\right. \\ &\left. +\exp\left(\frac{x_i'\Delta_A}{\sigma}\right)\left(\left(\frac{\sigma_a}{\sigma}\right)^2 E\left[\varepsilon_a^2\exp\left(\frac{\sigma_a}{\sigma}\varepsilon_a\right)|\varepsilon_A > \varepsilon_{iA}\right]\right. \right. \\ &\left. \left. +\frac{\sigma_a}{\sigma}E\left[\varepsilon_a\exp\left(\frac{\sigma_a}{\sigma}\varepsilon_a\right)|\varepsilon_A > \varepsilon_{iA}\right]\left(1 + 2\frac{x_i'\Delta_A}{\sigma}\right)\right)\right. \end{aligned}$$

$$+E\left[\exp\left(\frac{\sigma_a}{\sigma}\varepsilon_a\right)|\varepsilon_A > \varepsilon_{iA}\right]\left(\frac{x_i'\Delta_A}{\sigma}+\left(\frac{x_i'\Delta_A}{\sigma}\right)^2\right)\Bigg)\Bigg)$$

4 Simulation results

We report some simulation results of the SPEM estimator based on the exponential-, Weibull- and normal SPEM-criterion function. (Note that the normal SPEM-criterion function yields the Buckley- James-estimator.) We compare the bias and efficiency of the estimators for several different situations. In general, we assume the data generating process

$$\log(T) = \alpha_0 + x'\beta_0 + \sigma_0\varepsilon \tag{5}$$

where the 'true' distribution of T (depending on the distribution of the error term ε) is
a) exponential (i.e. Weibull distribution with scale-parameter $\sigma = 1$)
b) Weibull (with scale-parameter $\sigma = 0.5$)
c) Lognormal ($\sigma = 0.7$)
d) Log-logistic ($\sigma = 0.35$)
e) Gamma ($\sigma = 0.7$), and
f) Weibull or Log-logistic with a Gamma distributed heterogeneity component (Hamerle, Moller 1992)
The sample size of each simulation was $n = 1000$ observations. The matrix of the regressor variables consists of two columns where the first one is drawn from $N(0,2)$ and the second is dichotomous and takes the values 0 or 1 with probability 0.5 for either outcome. The true parameter vector is $\beta_0 = (-0.6, 0.3)'$ and the intercept term is $\alpha_0 = 1$. We used a Type II censoring mechanism with about 50 percent censored observations. Other known censoring mechanisms like a fixed maximum observation limit were also investigated and give similar results. The calculations are done with a SAS/IML program (Moller 1994). In the following table the bias and mean square errors (MSE) of the parameter estimates are given.

Tab. 1 SPEM-estimation with censored observations

true distribution	criterion function		
	Exponential	*Weibull*	*Normal*
Exponential-distribution	−0.44828	−0.24339[a]	−5.92542
	(3.42779)	(0.73366)	(0.60500)
	−0.04957	−0.06722[b]	0.02697
	(0.26878)	(0.32180)	(0.36018)
	0.04821	−0.02627[c]	−0.03884
	(0.76786)	(0.79910)	(0.98382)
	no	−0.63839[d]	2.56875
	estimator	(0.46041)	(0.45683)
Weibull-distribution $\sigma = 0.5$	−1.20287	−0.00061	−2.89069
	(0.66542)	(0.10726)	(0.13069)
	−0.01076	−0.01341	0.02312
	(0.08275)	(0.06589)	(0.11049)
	0.02074	0.00145	−0.02028
	(0.20347)	(0.20089)	(0.27123)
	no	−0.07458	1.38407
	estimator	(0.06030)	(0.07684)
Normal-distribution $\sigma = 0.7$	2.33167	3.27135	−0.07593
	(0.94751)	(0.23561)	(0.16762)
	−0.07318	−0.05153	0.01472
	(0.09578)	(0.09394)	(0.09382)
	0.08421	0.03282	0.01639
	(0.29390)	(0.25506)	(0.25172)
	no	−0.68298	−0.18980
	estimator	(0.23605)	(0.09984)
Logistic-distribution $\sigma = 0.35$	1.89519	2.94299	−0.08847
	(0.69475)	(0.21409)	(0.11558)
	−0.08878	−0.09038	−0.01059
	(0.07301)	(0.07728)	(0.06302)
	0.05541	0.01162	−0.00467
	(0.20833)	(0.25449)	(0.18910)
	no	2.22247	2.58604
	estimator	(0.27940)	(0.10511)

a) Bias α x 10
b) Bias β_1 x 10
c) Bias β_2 x 10
d) Bias σ x 10
(In brackets the MSE of the estimators x 10^2)

Tab. 2 Continuation: SPEM-estimation with censored observations

true distribution	criterion function		
	Exponential	*Weibull*	*Normal*
	−4.64466	−4.26820	−9.46368
	(2.03671)	(0.28016)	(0.55470)
Gamma-	−0.08839	−0.08664	−0.09369
distribution	(0.25707)	(0.20292)	(0.51797)
$\sigma = 0.7$	0.00312	0.05007	0.00995
$a = 1, b = 1$	(0.63836)	(0.56506)	(1.59257)
	no	3.21539	7.76907
	estimator	(0.13698)	(0.46620)
	2.88964	3.74227	−0.30139
	(0.68286)	(0.46927)	(0.24654)
Weibull-	−0.05396	−0.06532	0.01596
Gamma-	(0.12297)	(0.11491)	(0.13925)
mixing	0.04071	0.08598	0.01514
distribution	(0.56741)	(0.39764)	(0.35660)
$\sigma = 0.5$	no	2.41305	3.40627
$a = 1, b = 1$	estimator	(0.40995)	(0.24914)
	4.31866	5.37867	1.78026
	(0.45695)	(0.35323)	(0.18120)
Logistic-	−0.03422	−0.05322	0.02548
Gamma-	(0.08895)	(0.09457)	(0.09682)
mixing	0.01797	0.03011	0.00305
distribution	(0.25470)	(0.31386)	(0.23928)
$\sigma = 0.35$	no	3.24834	3.74058
$a = 1, b = 1$	estimator	(0.32334)	(0.15886)

The results of the simulations show that the bias of the different SPEM-estimators for β is not serious, regardless wether the assumptions about the underlying survival distribution function correspond to the choice of the SPEM-criterion function or not. Of course, this does not hold for the estimation of the nuisance parameters, i.e. the intercept α or the scale parameter σ, in case of misspecification. Estimation of the 'true' intercept ($\alpha_0 = 1$) of a model with Weibull- distributed duration time (with scale-parameter $\sigma_0 = 0.5$) using the original Buckley-James or exponential SPEM-estimator (i.e. a Weibull-SPEM-estimator with fixed 'wrong' scale parameter $\sigma = 1$) yields 'incorrect' results with $\hat{\alpha}_{BJ} = 0.71$ and $\hat{\alpha}_{exp} = 0.88$. Estimates based on these nuisance- parameters like the mean duration time should be used with caution.
More important, the efficiency of $\hat{\beta}$ of the different SPEM-estimators also strongly depends on the assumed distribution. Furthermore, comparing the three SPEM-estimators none of them is strictly superior to the other in all situations. It is apparent that even the 'simple' Exponential SPEM-estimator is superior to the Buckley-James estimator when a Weibull- distribution (with $\sigma_0 = 0.5$) is assumed.

The efficiency gain (1-MSE(EXP)/MSE(BJ)) of the exponential-SPEM estimator to the BJ-estimator in this case is about 22 percent for β_1 and about 26 percent for β_2 ($MSE\hat{\beta_{BJ}} = 0.0011||0.0027$, $MSE\hat{\beta_{Exp}} = 0.000828||0.00203$). On the contrary, when the assumed distribution of the error term is logistic, the MSE of the Buckley-James estimator is smaller than the Weibull- or Exponential-SPEM- MSE ($MSE\hat{\beta_{BJ}} = 0.06302||0.1891$, $MSE\hat{\beta_{Exp}} = 0.07301||0.20833$, $MSE\hat{\beta_{Wei}} = 0.07728||0.25449$), i.e. using Weibull or exponential-SPEM-criterion functions results in efficiency losses between 9 and 26 percent, depending on the kind of regressor and the use of a Weibull or Expontial-SPEM-criterion function.
Furthermore it is remarkable that the flexibility of the Weibull- SPEM-criterion function estimating the scale- parameter may be a disadvantage compared with the Exponential- SPEM-estimator with 'fixed scale-parameter'. Assuming logistic distributed errors, the exponential SPEM-criterion is more efficient than the Weibull-SPEM-function in estimating the regression coefficients. This indicates that the choice of σ with a Weibull-SPEM-function should be used to minimize the variance of the regression- parameter $\hat{\beta}$ and should not be the object of maximizing the criterion function.
In conclusion, it should be noticed that the whidespread use of the Buckley-James estimator, i.e. the SPEM- estimator assuming normal errors, in duration analysis is questionable with regard to possible efficiency advantages of other SPEM - estimators, depending on the assumptions about the true survival distribution. Here, further research about the relation of the true survival distribution and the impact on the efficiency of the different SPEM-estimators is desirable.

References

Blossfeld, H.P., Hamerle, A., Mayer, K.U. (1989) Event history analysis. Hillsdale: Lawrence Erlbaum

Buckley, J., James, I. (1979) Linear regression with censored data. Biometrica 66, 429-436

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

Gould, A., Lawless, J.F. (1988) Consistency and efficiency of regression coefficient estimates in location-scale models. Biometrika 75, 535-540

Hamerle, A., Moller, M. (1992) On the sensitivity of covariate effect estimates to misspecification in parametric event history models, in: Statistical Modelling, Herausgeber: van der Heiden, P.G.M., Jansen, W., Francis, B., Seeber, G.U.H., 131- 147

Heckman, J.J., Singer, B. (1984) A method of minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52, 271-320

Kalbfleisch, J.D., Prentice, R.L. (1980) The statistical analysis of failure time data, New York.

Kaplan, E.L., Meier, P. (1958) Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481

Kiefer, N. (1988) Econometric duration data and hazard functions. J. Economic Lit. 26, 646-679

Lancaster, T. (1990) The econometric analysis of transition data. Cambridge University Press

Manton, K.G., Stallard, E., Vaupel, J.W. (1986) Alternative models for the heterogeneity of mortality risks among the aged. J. Amer. Statist. Assoc. 81, 635-644

Moller M. (1994) Quasi-Maximum-Likelihood- und semiparametrische Schtzungen in linearen Transformationsmodellen der Verweildaueranalyse. Hartung-Gorre Verlag Konstanz 1994

Newman, J.L., Mc Culloch, C.E. (1984) A hazard rate approach to the timing of births. Econometrica 52, 939-961

Ridder, G. (1990) The non-parametric identification of generalized accelerated failure-time models. Review of economic studies 57, 167-182

Ritov, Y. (1990) Estimating in a linear regression model with censored data. Ann. Stat. 18, 303-328

Trussel, J., Richards, T. (1985) Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure, (pp 242-276) in N.B. Tuma (Ed.), Sociological Methodology , San Francisco: Jossey Bass

Tsiatis, A.A. (1990) Estimating regression parameters using linear rank tests for censored data. Ann. Stat. 18, 354-372

Vaupel, J.W., Yashin, A.I. (1985) The deviant dynamics of death in heterogenous populations, 179-211 in N.B. Tuma (ed.) Sociological Methodology, San Francisco: Jossey Bass SAS User's Guide]: Basics (1985). SAS Institute, Cary, N.C.

SAS/IML User's Guide (1985). SAS Institute, Cary, N.C.