

Analysis of Survival Data Using a Partially Linear Single Index Survival Model via Accelerated Failure Time Model

A. Sewalem * §, T. Desmond¹, R. Singh¹ and X. Lu²

*Guelph Food Research Centre, Agriculture and Agri-Food Canada, Guelph, ON, N1G 5C9, Canada, §Canadian Dairy Network, Guelph, ON, Canada, N1K 1E5 ¹ Department of Mathematics and Statistics, University of Guelph, Guelph, ON, N1G 2W1, Canada ² Department of Mathematics and Statistics, University of Calgary, Calgary, AB. T2N 1N4 Canada

ABSTRACT

In many practical situations the linear model is not complex enough to capture the underlying relationship between the response variable and its regressors. This paper explores this association in dairy cattle breeding data using the partially linear single-index model via the accelerated failure time model in addition to the ordinary Weibull model. Calves survival data were used this study. Each calf record contains the following information, survival time till weaning, type of birth (TB), calving ease score (CE), season of birth (SEAS), origin of farm or herd (HRD), number of treatments received (NTR), weight at the time of event taken (BWT), total volume of colostrums (TVOL) and serum total protein (TP) g/dl. Calves that survived up to 120 days were considered as censored. BWT, TVOL and TP were included in the nonparametric vector in the PSLISM model. The results show that the estimates of the parametric component are similar in the two models. However, the estimates of the nonparametric component differ from parametric analysis. This difference may be attributed largely to the nonlinearity of the estimated function indicating the standard linear model does not adequately capture the underlying association between the response and regressors in this study.

INTRODUCTION

In studying the relationship between a response and a set of predictor variables, the mean response variable is often assumed to be a linear regression function of the regressors. In many practical situations, however, the linear model is not complex enough to capture the underlying relationship between the response variable and its associated

covariates. Indeed, some components can be highly nonlinear. A natural generalization of the linear model is to allow only some of the predictors to be modeled linearly, with others being modeled nonlinearly.

For the past decades, to overcome the difficulty caused by the curse of dimensionality, several models have been developed to study high dimensional data by nonparametric or semiparametric regression models. Hardle and Stoker (1989), Powell et al (1989) and Newey and Stoker (1997) investigated single-index models. Further, Carroll et al (1997) and Xia and Stoker (2006) extended the single-index model to the generalized partially linear single-index model. Even so, these models are used to study the relationship between the response and the predictor variables when data are fully observable.

In practice, however, survival data are often subject to censoring. When it occurs, the incompleteness of the observed data may induce a substantial bias in the sample. Several approaches have been developed to overcome the associated difficulties in some specific models, including the partial likelihood method in the Cox proportional hazards model. Several studies were carried out using parametric and semi-parametric censored regression models by assuming a parametric regression form or assumed that the error distribution is parametric (Buckley and James, 1979; Koul et al., 1981; Lai et al., 1995; Powell, 1986; Duncan, 1986; Fernandez, 1986; Horowitz, 1988; Ichimura, 1998; Lewbel, 1998; Buchinsky and Hahn, 1998 and Heuchenne and Van Keilegom, 2007). The Cox (1972) regression model plays a central role in survival analysis in which the conditional hazard of failure at time t given the covariate vector Z takes a semi-parametric form

$$h(t;Z) = h_0(t) \exp(\beta^T Z),$$

where the baseline hazard $h_0(t)$ is a nonparametric function of time t , and the covariate effects are specified in such a way that the parameter vector β represents the log-linear effects on the hazard function. However, the Cox model has its limitations in dealing with more sophisticated covariate effects arising from real data. Several studies have tried to extend the Cox model to include nonparametric or semi-parametric covariate effects on censored failure data; for example, Dabrowska (1987) and Nielsen and Linton (1995), where the hazard function $h(t, z)$ is completely unspecified. Fan et al. (1997) relaxed the fully nonparametric specification to the form $h(t, z) = h_0(t) \lambda(z)$ where both h_0 and λ are nonparametric functions.

Fan and Gijbels (1994) proposed a censored nonparametric regression estimator based on a class of unbiased data transformations using only univariate regressors. Wang and Zheng (1997) and Liang and Zhou (1998), however, extended the univariate regressors to multiple regressors. Further, Singh and Lu (2002) studied censored nonparametric additive regression models based on some special data transformations. On studying the estimation of an unknown multiple regression function, Lu and Burke (2005) proposed a method called censored average derivative estimation (CADE). Lu et al. (2006) examined a class of partially linear single-index proportional hazards models for survival data. Lewbel and Linton (2002) and Chen et al. (2005) considered identification and estimation of a nonparametric location-scale model under fixed censoring. However, in most biological and agricultural fields censoring is random hence the application of these models is limited.

In survival analysis an alternative model to the proportional hazards model or the multiplicative hazards model is the accelerated failure time model (Lawless 2003). Lu and Cheng (2007) investigated a class of partially linear single-index models under random censoring as a class of accelerated failure-time models without the specification of the distribution function of the response variable. To date, in the case of survival analysis in dairy cattle, the covariates are often fitted in the model as linear multiplicative effects although several covariates are believed to have nonlinear effects on the survival of a cow. Therefore, the objective of this study was to explore the association of the response variable and various regressors in dairy cattle breeding data using the partially linear single-index model via the accelerated failure time model using the model proposed by Lu and Cheng (2007).

MATERIALS AND METHODS

2.1 Description of the data set used in this study

Calf mortality is a serious problem in dairy cattle production. Apart from raising labor and veterinary expenses, it mainly increases replacement cost. Therefore, identifying calves that have a good genetic potential for survival is an important aspect of dairy farming. An experiment was set up to study the association between disease incidence and pre-weaning survival time of calves from birth till the first 120 days.

Data were collected from 16 dairy farms from Ontario-enrolled calves in the study from January through December of 2008. In this experiment calf health was monitored and recorded for various calf disease occurrences. Calves were treated at least once for various diseases in the study period. The farms selected for the study were generally well managed farms, enrolled in CanWest dairy herd improvement milk recording program. Size and management styles varied between farms, allowing for an overall picture of the Ontario dairy industry. The data set included survival time of calves in days, number of treatments each calf received and type of disease that occurred, calf weight, height, origin of farm or herd, season of birth (seasons were Jan-March, April-June, July-Sept, and Oct-Dec), calving ease score (scores were 0 = unassisted, 1 = easy pull, 2 = hard pull, 3 = surgery), type of birth (0 = single birth, 1 = multiple birth), the total volume of colostrums (Kg) consumed during the first 24 hours of life and serum total protein (g/dl).

Pre-weaned calves were housed both outdoors and indoors in a variety of barn set ups. Age was used as a criterion for weaning on all farms participating in the study. Some criteria included grain intake and body weight. Enrolment occurred at birth. Producers were also asked to complete a Birth Record for each calf born on the farm. These records listed the date, time and location of the birth, as well as calving ease score (surgery, assisted easy pull was required during the delivery). Colostrums management, including volume and time of colostrums feeding was also recorded, along with any perinatal treatments. Calves were uniquely identified using National Livestock Identification for Dairy tags and Holstein Canada registration numbers (where applicable).

During weekly farm visits, various measurements to determine the overall health status of a calf were taken. Calves were measured at 4 different time points throughout the study; the initial assessment occurred at 1-8 days of age (T1), with the three follow up visits occurring at 14-22 (T2), 35-43 (T3), and 90-120 (T4) days of age. Body weight and height measurements were recorded in the individual calf event record book at each visit. Calves were also assigned a health score at each of the 4 sampling visits. A blood sample was obtained from each calf during the initial assessment to determine the total protein level. Producers were provided with case definitions of diseases and asked to record any treatments provided to calves. For every treatment which a producer administered to a calf, they were asked to record the date, rectal temperature of the calf, type of treatment given, as well as the suspected disease. A health score was also given to every treated calf. A total of 511 calves were included in this study. Each calf record contains the following information, survival time until weaning, type of birth (**TB**), calving ease score

(**CE**), season of birth (**SEAS**), origin of farm or herd (**HRD**), number of treatments received (**NTR**), weight at the time of event taken (**BWT**), total volume of colostrums (**TVOL**) and serum total protein (**TP**) g/dl. Calves that survived up to 120 days were considered as censored.

The primary goal of the study was to assess which of the covariates are useful in predicting mortality of calves from birth to the first 120 days. First the data was using the ordinary Weibull model

$$\lambda(t) = \lambda_0(t) \exp\{\mathbf{x}'_m(t)\boldsymbol{\beta}\}$$

Here $\lambda(t)$ is the hazard of a calf, i.e., the probability of dying at time t given she is alive just before t ; $\lambda_0(t) = \lambda_0 \rho (\lambda t)^{\rho-1}$ is the Weibull baseline hazard function with scale parameter λ and shape parameter ρ and t is the time in days from one calving to death or censoring. $\boldsymbol{\beta}$ contains the parameters affecting the hazard with $\mathbf{x}'_m(t)$ being the corresponding covariate vectors

In addition to the above ordinary Weibull model the following partially linear single-index model was used to analyze the data,

$$Y = \beta_0^T V + \lambda_0(\alpha_0^T X) + \sigma(V, X)\epsilon \quad \text{with } \|\alpha_0\| = 1,$$

where Y is the log survival time, V and X are the associated regressors, q and p vectors, respectively. For ease of understanding we denote the covariates using two different symbols: V and X which comprise the parametric and nonparametric component, respectively. The parametric component is characterized by an unknown q vector with parameter β_0 . The nonparametric component is characterized by λ_0 , an unknown smooth univariate function defined on the real line, and an unknown projection p -vector parameter α_0 . $\sigma(\cdot, \cdot)$ is the conditional variance representing possible heteroscedacity; $\|\cdot\|$ denotes the Euclidean norm. The constraint $\|\alpha_0\| = 1$ on the single-index coefficient parameters is required for parameter identifiability. Assume that (V, X) and ϵ are independent, $E(\epsilon) = 0$ and let C be the random censoring time associated with the log survival time Y , $\text{Var}(\epsilon) = 1$. Assume C is independent of (V, X, Y) . Denote $Z = \min(Y, C)$ and $\delta = I(Y \leq C)$. The observations are $\{(V_i, X_i, Z_i, \delta_i) : i = 1, \dots, n\}$ which are regarded as a random sample from the population (V, X, Z, δ) .

Application of the partially linear single-index approach using the above mentioned model, raises practical issues of which covariates go into the nonparametric

(V) vector and which ones go into the parametric (X) vector. In this study, we utilized the subject matter knowledge related to the individual calf and the underlying physiological mechanism that influences the ability of calf to reach the next stage. In addition preliminary analysis was carried out to investigate if the covariates were associated linearly or nonlinearly with the response variable.

Estimation Procedure

Since the distribution of error in the model is not specified, applying the a full likelihood function in the above model is not possible. Therefore, a quasi-likelihood estimation procedure was implemented using an iterative minimization algorithm (Lu and Cheng, 2007). The term quasi-likelihood here is similar to that of Wedderburn (1974) in that only first and second assumptions are made about the distribution of the response Y.

Let $\theta = (\alpha, \beta)$ be the vector of model parameters and if the data is fully observed, i.e., $Z \equiv Y$ the quasi likelihood estimator of $\theta_0 = (\alpha_0, \beta_0)$ and λ_0 are the minimizers of the following quasi-likelihood function of $\{(V_i, X_i, Z_i, \delta_i) : i = 1, \dots, n\}$,

$$\ell_n(\theta, \lambda) = \sum_{i=1}^n [Y_i - \{\beta^T V_i + \lambda(\alpha^T X_i)\}]^2 \text{ with } \|\alpha\| = 1$$

This model is similar to the generalized linear single-index models as presented by Carroll et al (1997) for complete data. This procedure encounters difficulties in estimation due to censoring and the involvement of the nonparametric function λ . To overcome this difficulty first synthetic data or pseudo responses were produced using:

$$Z_{i\hat{G}} = (1 + \phi) L_{i\hat{G}} - \phi K_{i\hat{G}} \text{ (following the procedure of Lu and Cheng, 2007).}$$

Here $L_{i\hat{G}} = \int_{-\infty}^{\infty} \left(\frac{I[Z_i \geq s]}{(1 - \hat{G}(s-))} - I[s < 0] \right) ds$, $K_{i\hat{G}} = \frac{Z_i \delta_i}{(1 - \hat{G}(Z_i-))}$, ϕ is a tuning parameter which control the weights put on censored and uncensored observations and $I(\cdot)$ is the indicator function. $(1 - \hat{G}(\cdot-))$ is the left continuous version of Kaplan-Meier estimator defined by

$$1 - \hat{G}(t) = \prod_{k=1}^n \left[\frac{n-i}{n-i+1} \right]^{I[Z_{(i)} \leq t, \delta_{(i)}=0]},$$

$Z_{(1)} \leq Z_{(2)} \dots Z_{(n)}$ are the order statistics of Z-sample and δ_i is the associated δ with $Z_{(i)}$, $i = 1, 2, \dots, n$. The observed data $(V_i, X_i, Z_i, \delta_i)$ is replaced by $(V_i, X_i, Z_{i\hat{G}})$. The pseudo responses are such that when G is known, the expected value of $Z_{i\hat{G}}$ equals the expected value of Y, i.e. $E(Z_{i\hat{G}}) = E(Y)$. Thus the censored observations are unbiasedly transformed to pseudo responses, which approximate or impute the unobserved values.

When G is unknown, we may substitute the Kaplan-Meier estimator \hat{G} for G . The transformation is still asymptotically unbiased in this case. This class of transformation was introduced by Fan and Gijbels [8], and Koul, Susarla, and Van Ryzin [16] abbreviated as KSV. Using the transformed data, both parametric (β_0) and nonparametric (λ_0) were estimated by applying the local linear fit to the quasi log likelihood iteratively.

Two well-known qualities of the local linear fit are the reduction of the bias for the estimation of the nonparametric function and the avoidance of boundary effects. Suppose that $\lambda(\cdot)$ is continuously differentiable. Then in a neighborhood of a fixed point u , we can write $\lambda(v) \approx a_0 + a_1 (v - u)$, where $a_0 = \lambda(u)$ and $a_1 = \lambda'(u)$. This is called the local linear fit. α

Let $W(\cdot)$ be a kernel with a given bandwidth b and a given parameter vector θ . One can obtain local estimators $\hat{a}_0 \equiv \hat{a}_0(\mathbf{u}; \mathbf{b}, \theta)$, $\hat{a}_1 \equiv \hat{a}_1(\mathbf{u}; \mathbf{b}, \theta)$ by minimizing the following local quasi log likelihood

$$\ell_n(a_0, a_1) = \sum_{i=1}^n [Z_{i\hat{G}} - \{\beta^T V_i + a_0 + a_1(\alpha^T X_i - u)\}]^2 W_b(\alpha^T X_i - u),$$

where $W_b(\cdot) = b^{-1}W(\cdot/b)$, and u is a fixed real number.

When the true parameter vector θ is unknown, in order to obtain estimators for the model, we need to iteratively update the estimates of the nonparametric component $\lambda_0(\cdot)$ and the parametric components $\theta_0 = (\alpha_0, \beta_0)$. The iterative algorithm consists of the following steps,

Step 1: Treat the pseudo-responses $Z_{i\hat{G}}$ as complete data and apply the estimation procedure for the partially linear single-index models, to obtain initial estimates $\hat{\alpha}$ and $\hat{\beta}$ of α_0 and β_0 respectively, with the restriction $\|\hat{\alpha}\| = 1$ and $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$.

Step 2: Find $\hat{\lambda}(u; b, \hat{\theta}) = \hat{a}_0$ as a function of u by maximizing the local quasi log-likelihood with respect to a_0 and a_1 with fixed $\theta = \hat{\theta}$ and a suitable bandwidth b as described by Lu and Cheng (2007).

Step 3. Update $\hat{\theta}$ by minimizing the following equation with respect to $\theta = (\alpha, \beta)$

$$\sum_{i=1}^n [Z_{i\hat{G}} - \{\beta^T V_i + \hat{\lambda}(\alpha^T X_i; b, \hat{\theta})\}]^2$$

Step 4. Cycle Steps 2 and 3 until $\hat{\theta}$ convergence of $\hat{\theta}$.

Standard errors of parameters were estimated by generating 500 independent bootstrap samples with replacement . For each independent sample drawn the above aforementioned model was fitted and the corresponding parameters were calculated as described.

Results and Discussion

To compare the results obtained from this study particularly the parametric component of the model, the data set was analyzed using the ordinary Weibull model and the results are presented in Table 2.

Table 2. Summary of results from ordinary Weibull model.

		parameter	SE	P
(Intercept)		7.923	0.3808	0.000
BWT	α_1	2.508	0.2013	0.000
TVOL	α_2	-0.114	0.1224	0.035
TP	α_3	0.124	0.0931	0.018
HRD	β_1	-0.166	0.1923	0.003
SEAS	β_2	0.144	0.1168	0.219
TB	β_3	-0.066	0.1006	0.041
CE	β_4	-0.112	0.074	0.016
NTR	β_5	-0.215	0.023	0.002

BWT= body weight, **TVOL**= totla colostrol volume, **TP**= total serum protein level, **HRD**= famr; **SEAS**= season of calving, **TB**= type of birth; **CE**= calving ease score; **NTR**= number of treatments.

The results show that body weight had a significant effect ($P < 0.001$) as did total volume of colostrum, total protein, herd, type of birth and calving ease score ($P < 0.05$). Season of birth had no significant effect on calf survival. Detailed analysis of the data also showed the relative calf mortality was associated with the type of birth. For instance, twin born calves had 1.12 times higher risk of dying than single born calves. Usually twin born calves have lower birth weight and subsequent growth rate than single born calves due to prenatal nutritional competition and other maternal effects. Table 2 also shows significant relationship between calving ease and calf mortality. Calves with hard pull and surgery had a considerably increased relative risk of dying compared to unassisted calving. For instance, calves with hard pull and surgery were 1.27 and 1.92, respectively, more likely to have died compared to unassisted calvings.

Difficult births have a remarkable effect on calf survival and health. When cows have to be assisted or have surgery during birth, there are often lasting effects on the calf. Calves may suffer from anoxia, lack of oxygen and may have damage to joints, bones or

organs. Consequently, the calf feels weak and is slow to stand or nurse the cow. As a result many calves suffer from failure of passive transfer and are more susceptible to disease.

The relationship between the number of treatments received and survival shows that calves who receive treatments twice and three times had a considerably increased relative risk of dying compared to calves that received only one treatment. For instance, calves with two and three treatments were 2.52 and 3.38, respectively, more likely to have died compared to calves with only one treatment group.

Table 3 presents estimates and standard errors obtained by the ordinary Weibull model and partially single linear index model. Table 3 shows that the estimates of the parameters in the parametric component (β) are similar under the ordinary Weibull linear model and the partially linear single-index survival model. However, the estimates of the nonparametric component (single-index) parameter, α , are different.

Table 3. Estimate and standard error of the parameters obtained from Weibull model and the partially single linear index model for the data set

		Ordinary Weibull		PSLISM	
		parameter	SE	parameter	SE
BWT	α_1	2.508	0.2013	0.268	0.105
TVOL	α_2	-0.114	0.1224	0.354	0.162
TP	α_3	0.124	0.0931	0.896	0.247
HRD	β_1	-0.166	0.1923	-0.156	0.125
SEAS	β_2	0.144	0.1168	0.129	0.101
TB	β_3	-0.066	0.1006	-0.078	0.092
CE	β_4	-0.112	0.074	-0.117	0.084
NTR	β_5	-0.215	0.023	-0.151	0.034

BWT= body weight, **TVOL**= total colostrum volume, **TP**= total serum protein level, **HRD**= farm; **SEAS**= season of calving, **TB**= type of birth; **CE**= calving ease score; **NTR**= number of treatments; **PSLISM**= partially single linear index survival model

This difference could be attributed largely to the nonlinearity of the estimated function as observed in Figure 1.

Figure 1 shows that as the index increases the survival of calves increases in nonlinear fashion. At the beginning the effect is dramatic but once it reaches a threshold value it levels off. Looking at the individual components of the single-index such as the body weight reveals that calves with higher body weight tended to have lower risk of dying compared to the calves with lower body weight.

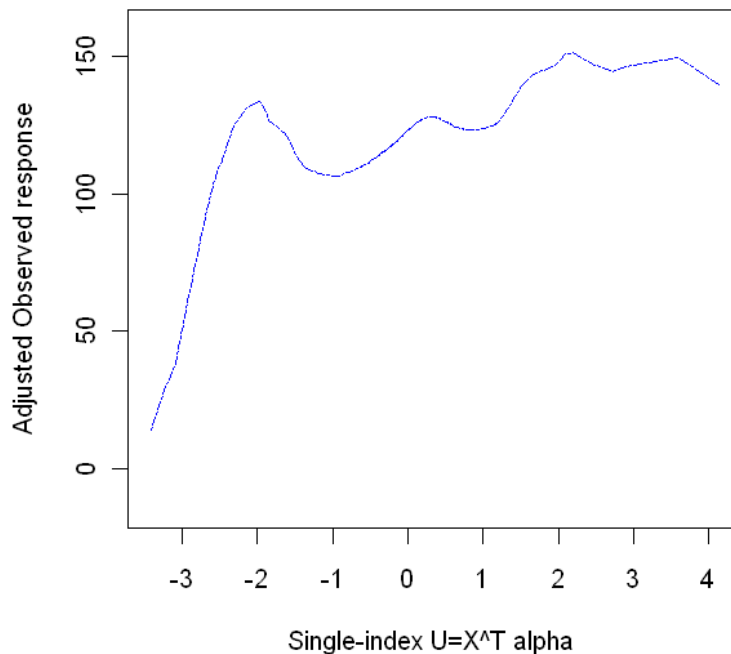


Figure 1. Adjusted observed response against the estimated single-index value

Moreover, it is observed from the results that calves with higher volume of colostrum and serum protein level have higher survival rate than those calves with lower volume. This is because these two components are physiologically important for immune response and ability to resist some disease incidence. The primary factors involved with successful attainment of passive immunity are the volume of colostrum fed, the IgG content of the colostrum and how quickly it was fed. The main role of IgG is to identify and help destroy invading pathogens. Therefore, the consumption and absorption of colostrum early in life is a primary determinant of calf health and growth. In turn, the success of the colostrum feeding is reflected in an increase in serum protein concentrate in the blood. Serum protein, is also important to the calf for growth and development. These three components of the index have somewhat intricate interdependence on each other that influences the calf survival in a complex way which may not be explained by application of the usual standard survival linear model.

In the survival literature, various statistical models were proposed for analyzing censored data in the presence of covariates. Parametric, semiparametric survival models such as the accelerated failure time model, the Cox proportional hazards model and the

additive risk model are commonly used in many clinical trials, biomedical and agricultural studies. These models specify the form of the conditional hazard function of survival time given a set of associated covariates. An alternative approach is to use a direct relationship between survival time and its covariates by means of linear regression. For the last few decades the linear regression techniques for censored survival data were used extensively because of the ease of interpretation of the results.

In the above-mentioned models, however, a specific functional form of relationship between survival time and the associated covariates is given. Despite the fact that all these models have nice theoretical properties, they may not be flexible enough to describe the complexity of biological processes in many real applications. The single-index model as observed in the present study may have greater flexibility than other regression models in terms of analyzing complex data since the link function in the model is assumed arbitrary.

In dairy cattle the breeding goal is to increase lifetime profit per animal and per unit of time. Profit is a function of production and the time that a cow remains in herd (commonly called survival or longevity or herd life). Therefore, survival or longevity of cows is a trait of considerable economic importance since it has a significant impact on profitability. Increased longevity is associated with decreased culling and therefore decreased cost of raising or purchasing replacement females.

Several strategies have been suggested and used to analyze survival data in dairy cattle. These include a simple modeling of a 0-1 variable indicating whether the cow is still alive or dead at any specific time. In this approach, the response variable was considered as a binary trait and analyzed either using a linear or threshold model (VanRaden and Klaaskate, 1993; Jairath et al., 1998; Vollema and Groen, 1998; Boettcher et al., 1999, Sewalem et al 2007). Typically such type of data has a skewed distribution and analysis using traditional linear models may not be appropriate. Survival analysis using a proportional hazard model as suggested by (Smith and Quaas, 1984) is an alternative method for animal breeding survival data. Ducrocq et al. (1988) showed that proportional hazard models could be used for the analysis of length of productive life. Ducrocq and Solkner (1998) developed the Survival Kit typically used by animal breeders for large populations using a Weibull model (Ducrocq, 2002, Sewalem et al., 2004, Sewalem et al., 2005) where several covariates are fitted as a linear effect.

Generally, in dairy cattle production there are several environmental factors (covariates) that influence the survival of cows and those factors need to be accounted in the model in order to get reliable estimates. In this regard Sewalem et al. (2005) studied longevity of Canadian dairy cows using censored linear regression model that included several covariates. In those analyses some of the continuous covariates are grouped and fitted to the model as a class effect. This grouping of covariates may also result in loss of information. In addition, some covariates may have nonlinear effects on the response variable. In this case, the traditional linear models or kernel smoothing methods fail to incorporate both linear and nonlinear covariate effects. On the other hand, when a large amount of covariates have nonlinear effects, the multivariate kernel smooth suffers from the “curse of dimensionality”. The covariate effects in the current model are addressed in a semiparametric fashion, which offers better flexibility in modelling the relationship between the failure time and the covariates than the existing models. Hence, application of the current model is worthy of a full investigation using a larger data set that may include frailty model that accounts the genetic effect of the animal.

As a concluding remark in the present study, we are demonstrating the results for single-index model under random censorship. The results may provide some insights which may be potentially useful in analysis of dairy cattle breeding data in the future which currently suffer with the curse of dimensionality.

REFERENCES

- Allaire, F. R., and Gibson, J. P. (1992). Genetic value of herd life adjusted for milk production, *Journal of Dairy Science*, 75,1349.
- Boettcher, P. J, Jairath, L. K., and Dekkers, J. C. M. (1999). Comparison of methods for genetic evaluation of sires for survival of their daughters in the first three lactations, *Journal of Dairy Science*, 82,1034–1044.
- Buchinsky, M. and Hahn, J. (1998). An alternative estimator for the censored quantile regression model, *Econometrica*, 66, 653–671.
- Buckley, J. and James, I.R. (1979). Linear regression with censored data, *Biometrika*, 66, 429–436.
- Chen, S., Dahl, G.B. , Khan, S. (2005). Nonparametric identification and estimation of a censored location-regression model, *Journal of the American Statistical Association*, 100, 212–221.
- Dabrowska, D. M. (1987). Nonparametric regression with censored survival time data. *Scandinavian Journal of Statistics*, 14, 181-192.

- Ducrocq, V. and Solkner, J. (1998). "The Survival Kit V3.12", a package for large analyses of survival data. Pages 447–450 in Proc. 6th World Congress. on Genetics Applied to Livestock Production, Armidale, Vol. 27.
- Ducrocq, V. Quaas, R. L., Pollak, E. J. and Casella, G. (1988). Length of productive life of dairy cows. 1. Justification of a Weibull model. *Journal of Dairy Science*, 3061–3070.
- Duncan, G.M.(1986). A semi-parametric censored regression estimator, *Journal of Econometrics*, 32, 5–24.
- Fan, J. and Gijbels, I. (1994). Censored regression: local linear approximations and their applications, *Journal of the American Statistical Association*, 89, 560–570.
- Fan, J., Gijbels, I. and King, M. (1997). Local likelihood and local partial likelihood in hazard regression. *The Annals of Statistics*, 25, 1661-1690.
- Fernandez, L. (1986). Non-parametric maximum likelihood estimation of censored regression models, *Journal of Econometrics*, 32, 35–57.
- Heuchenne, C. and Van Keilegom, I. (2007). Location estimation in nonparametric regression with censored data. *Journal of. Multivariate. Analysis*, 98, 1558-1582.
- Härdle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the method of average derivative, *Journal of the. American Statistical Association*, 84, 986–995.
- Horowitz, J.L. (1988). Semiparametric M-estimation of censored linear regression models, *Adv. Econometrics* 7, 45–83.
- Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics*, 58, 71–129.
- Jairath, L., Dekkers, J. C. M., Schaeffer, Liu, Z., Burnside, E. B., and Kolstad, B. (1998). Genetic Evaluation for herd life in Canada. *Journal of Dairy Science*, 81,550-562.
- Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data, *Ann. Statist*, 9, 1276–1288.
- Lai, T.L., Ying, Z. and Zheng, Z. (1995). Asymptotic normality of a class of adaptive statistics with applications to synthetic data methods for censored regression, *Journal of Multivariate Analysis*, 52, 259–279.
- Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*, Wiley, New York.
- Lewbel, A. (1998). Semiparametric latent variable model estimation with endogenous or mismeasured regressors, *Econometrica*, 66, 105–121.
- Lewbel, A., Linton, O. (2002). Nonparametric censored and truncated regression, *Econometrica*, 70, 765–779.

- Liang, H. and Zhou, Y. (1998). Asymptotic normality in a semiparametric partial linear model with right-censored data, *Comm. Statist. Theory & Methods*, 27, 2895–2907.
- Lu, X. and Cheng, T.L. (2007). Randomly censored partially linear single-index models. *Journal of Multivariate Analysis*, 98, 1985-1922.
- Lu, X. and Burke, M.D. (2005). Censored multiple regression by the method of average derivatives, *Journal of Multivariate Analysis*, 95, 182–205.
- Lu, X., Chen, G., Song, X.K. and Singh, R.S. (2006). A class of partially linear single-index survival models, *Canadian Journal of Statistics*, 34, 99–116.
- Nielsen, J.P. and Linton, O. B. (1995). Kernel estimation in a nonparametric marker dependent hazard model. *The Annals of Statistics*, 23, 1735-1748.
- Powell, J.L.(1986). Censored regression quantiles, *Journal of Econometrics*, 32, 143–155.
- Sewalem, A., Miglior, F., Kistemaker, G. J., Sullivan, P., Huapaya G. and Van Doormaall. B. J. (2007). Modification of Genetic Evaluation of Herd life from a 3-trait to 5-trait Model in Canadian dairy cattle. *Journal of Dairy Science*, 90,2025-2028.
- Sewalem, A., Kistemaker, G. J. Ducrocq, V. and Van Doormaal. B. J. (2005). Genetic analysis of herd life in Canadian dairy cattle on a lactation basis using a Weibull proportional hazard model. *Journal of Dairy Science*, 88, 368-375.
- Singh, R.S. and Lu, X. (2002) Censored additive regression models, *Handbook of applied econometrics and statistical inference*, in: A. Ullah, A.T.K. Wan, A. Chaturvedi (Eds.), *Statistics: Textbooks and Monographs*, vol. 165, Dekker, New York, 143–157.
- Smith, S. P. and Quaas. R. L (1984). Productive lifespan of bull progeny groups: failure time analysis. *Journal of Dairy Science*, 67, 2999–3007.
- VanRaden, P. M., and Klaaskate. E. J. H. (1993). Genetic evaluation of length of productive life including predicted longevity of live cows. *Journal of Dairy Science*,76, 2758–2764.
- Vollema, A. R. and Groen. A. F. (1998). A comparison of breeding value predictors for longevity using a linear model and survival analysis. *Journal of Dairy Science*, 81,3315–3320.
- Wang, Q.H. and Zheng, Z.G. (1997). Asymptotic properties for the semiparametric regression model with randomly censored data, *Science China Ser. A*, 40, 945–957.