Working Paper 94-31
Statistics and Econometrics Series 12
September 1994

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

SEMIPARAMETRIC LINEAR REGRESSION WITH CENSORED DATA AND STOCHASTIC REGRESSORS

Juan Mora*

Abstract

We propose three new estimation procedures in the linear regression model with randomly-right censored data when the distribution function of the error term is unspecified, regressors are stochastic and the distribution function of the censoring variable is not necessarily the same for all observations ("unequal censoring"). The proposed procedures are derived combining techniques which produce accurate estimates with "equal censoring" with kernel-conditional Kaplan-Meier estimates. The performance of six estimation procedures (the three proposed methods and three alternative ones) is compared by means of some Monte Carlo experiments.

Key Words
Censoring; Linear regression; Kaplan-Meier estimator; Kernel estimator.

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid.
This paper is based on research funded by Spanish Dirección General de Investigación Científica y Técnica (DGCIYT), reference number PB92-0247. I am grateful to Miguel A. Delgado for his comments and suggestions.

1. INTRODUCTION

Consider the linear regression model

$$E[T|X] = X'\beta \quad a.s. \tag{1.1}$$

where (T,X) is an $\mathbb{R}\times\mathbb{R}^p$-valued random variable such that $E|T|<\infty$ and β is an $\mathbb{R}^p$-vector of unknown parameters. Suppose that we do not observe the variable T but instead we observe

$$Z = min(C,T) \text{ and } \delta = I(T<C), \tag{1.2}$$

where C is an $\mathbb{R}$-valued random variable and $I(A)$ denotes the indicator function of event A. This is referred to as the linear regression model with randomly-right censored data and stochastic regressors. T and C are usually termed, respectively, the survival time and the censoring variable. This chapter deals with estimation of β based on a random sample $\{(Z_i,\delta_i,X_i), 1\leq i\leq n\}$ when the distribution function of the error term $\varepsilon \equiv T - E[T|X]$ is of unknown functional form.

The linear regression model with randomly-right censored data appeared as an alternative to the proportional hazards model introduced by Cox (1972)[2]. In practice, the linear model (1.1) has been used to analyse censored data in the context of survival times in medical trials; T denotes the survival time (usually in logarithms) of a patient and X is a vector of individual characteristics. Censorship appears because patients often survive beyond the end of the trial or are dropped from the study for other reasons; see Kalbfleisch and Prentice (1980) for examples. In econometrics, this model is

[2] The continuous version of the proportional hazards model specifies $f^T(t|x)/(1-F^T(t|x))^{-1} = \lambda(t)exp\{x'\beta\}$, where $f^T(.|x)$ and $F^T(.|x)$ denote the underlying conditional density and distribution function of $T|X=x$, respectively. Estimation procedures in this model and applications may be found, for example, in Kalbfleisch and Prentice (1980).

of interest, in many situations, when analysing duration of unemployment spells (see, for example, Heckman and Singer 1984) or the timing and spacing of births (see, for example, Heckman and Walker 1990).

During the past 18 years, different estimation procedures have been suggested in this model when no assumption on the distribution function of the error term is made. Most of these procedures are based on the well-known Kaplan-Meier (KM) estimator of the distribution function (Kaplan and Meier 1958). Miller (1976) and Buckley and James (1979) proposed iterative estimators for the simple linear regression model when the regressor is non-random. The former may be also used in multiple regression with random regressors but the latter, which also assumes that the censoring variable is non-random, depends crucially on these assumptions. Koul et al. (1981) and Leurgans (1987) proposed procedures which do not require any iteration scheme. Both estimators may be used with fixed or random regressors but it is necessary to assume equal censoring for all observations, that is, the distribution function G_i of $C|X=x_i$ is the same for all i ($G_i = G$, $1 \le i \le n$). Chatterjee and McLeish (1986) discussed a method termed the linear attribute method. Gonzalez-Manteiga and Cadarso-Suarez (1991, 1994) proposed procedures based on prior nonparametric estimation of the regression function for random and non-random regressors, respectively.

The objective of this chapter is to propose and compare various estimation procedures when regressors are stochastic and when G_i is not necessarily the same for all observations, that is, unequal censoring. Specifically, we analyse six estimation procedures. Three of them are new, at the best of our knowledge. These three procedures result from combining methods which are known to produce accurate estimates with equal censoring (Buckley and James 1979, Koul et al. 1981 and Leurgans 1987) with kernel nonparametric estimates of G_i. The three other procedures which we consider in this chapter have already appeared in the literature (Miller 1976, Chatterjee and McLeish 1986 and Gonzalez-Manteiga and Cadarso-Suarez 1991) and may be used in this context of random regressors and unequal censoring with no modification (the first one and the second one were not specifically designed for this stochastic-regressors model, but may be straightforwardly adapted to it).

In Section 2 we first describe briefly the well-known Kaplan-Meier and kernel conditional estimators. The methodological contribution of this chapter is contained in Sections 2.2 to 2.4, where we describe the three new estimation procedures. For completeness, we also present the three other estimators to be compared. In Section 3 we carry out an extensive simulation study in order to examine the performance of all described estimators. In Section 4 conclusions on the usefulness of the proposed procedures are drawn. Proofs are confined to an appendix.

2. ESTIMATION PROCEDURES

2.1. Kaplan-Meier estimator and other related estimators

The key component of the three procedures we propose is the kernel-conditional KM estimator (see Beran 1981 or Dabrowska 1987, 1989), which combines KM weights and kernel nonparametric weights to yield a censored-data-set based estimate of the conditional distribution function. First of all, let us describe briefly the KM and the kernel-conditional KM estimates[3].

Given a random sample $\{(Z_i,\delta_i),\ 1\le i\le n\}$, where $Z_i = min(T_i,C_i)$ and $\delta_i = I(T_i<C_i)$, denote $F^T(t)$ and $G(t)$ the distribution functions of T and C, respectively, $H_1(t) \equiv P(Z>t,\delta=1)$ and $H_2(t) \equiv P(Z>t)$ (these are usually referred to as subsurvival functions). It is assumed that

$$T \text{ and } C \text{ are independent random variables, and} \tag{2.1}$$

$$\forall\ t\in\mathbb{R},\ 1 - F^T(t) > 0 \text{ and } 1 - G(t) > 0. \tag{2.2}$$

They are both standard assumptions. (2.1) is an identifiability condition, whereas (2.2) is necessary to obtain equation (2.4) below. The latter assumption is not very restrictive in practice, because T usually denotes survival time (often in logarithms) of an individual. The cumulative hazard

[3] The following description is adapted from Kalbfleisch and Prentice (1980) and Dabrowska (1989).

function associated with $F^T(.)$ is then defined as

$$\Lambda(t) \equiv \int_{\infty}^{t} (1 - F^T(s-))^{-1} dF^T(s), \tag{2.3}$$

where, for any real function $U: \mathbb{R} \longrightarrow \mathbb{R}$,

$$U(s-) \equiv \lim_{h \to 0^-} U(s+h).$$

It is possible to relate $F^T(.)$ and the subsurvival functions $H_1(.)$ and $H_2(.)$, since the following relations hold:

$$F^T(t) = 1 - \prod_{s \le t} (1 - d\Lambda(s)), \tag{2.4}$$

$$\Lambda(t) = - \int_{\infty}^{t} (H_2(s-))^{-1} dH_1(s), \tag{2.5}$$

Notice that, by (2.1), $H_2(t) = (1-F^T(t))(1-G(t))$, which is greater than 0 by (2.2). As Z and δ are observable, it is possible to estimate the subsurvival functions $H_1(.)$ and $H_2(.)$ by their sample counterparts,

$$\hat{H}_1(t) = n^{-1} \sum_j I(Z_j > t,\ \delta_j = 1),\ \hat{H}_2(t) = n^{-1} \sum_j I(Z_j > t),$$

where, hereafter, all summations run from 1 to n unless otherwise specified and $I(A)$ denotes the indicator function of event A. Now, replacing $H_1(.)$, $H_2(.)$ by $\hat{H}_1(.)$, $\hat{H}_2(.)$ in (2.5) we obtain an estimate $\hat{\Lambda}$ of the cumulative hazard function, which is referred to as Aalen-Nelson estimate (Aalen 1978, Nelson 1972); and replacing Λ by $\hat{\Lambda}$ in (2.4) we obtain the Kaplan-Meier (KM) estimate of $F^T(.)$, which will be denoted as $\hat{F}^T_{KM}(.)$ (Kaplan and Meier 1958). When there are no ties among the observations of Z, the KM estimate may be expressed as

$$\hat{F}^T_{KM}(t) = 1 - \prod_{j=1}^{n} \left[\frac{\sum_s I(Z_s > Z_j)}{\sum_s I(Z_s \ge Z_j)} \right]^{\gamma_j(t)},$$

where $\gamma_j(t) \equiv I(Z_j \le t,\ \delta_j = 1)$ and, hereafter, we arbitrarily define $0/0$ to be 0

and 0^0 to be 1.

$\hat{F}^T_{KM}(t)$ is a non-decreasing right-continuous function which takes values on $[0,1]$. Furthermore, let us denote $Z_{(n)} \equiv max(Z_1,...,Z_n)$. Then, the KM estimate satisfies that

$$\hat{F}^T_{KM}(t)=1 \Leftrightarrow t\geq Z_{(n)} \text{ and } \delta_j=1 \ \forall\ j \text{ such that } Z_j=Z_{(n)}; \tag{2.6}$$

hence, if there is a censored observation j such that $Z_j=Z_{(n)}$ then $1-\hat{F}^T_{KM}(t) > 0$ for all t.

Susarla and Van Ryzin (1980) introduced the following variant of the KM estimator,

$$\hat{F}^T_{SV}(t) \equiv 1-\prod_{j=1}^{n}\left[\frac{1+\sum_s I(Z_s>Z_j)}{1+\sum_s I(Z_s\geq Z_j)}\right]^{\gamma_j(t)}.$$

They proved that this estimator has the same asymptotic properties as $\hat{F}^T_{KM}(.)$. It was introduced because it satisfies that $1-\hat{F}^T_{SV}(t) > 0$ for all t, a property which allows us to consider $log(1-\hat{F}_{SV}(t))$ (see Section 2.3 below). Note that, when there are no ties among the observations of Z, $\hat{F}^T_{SV}(.)$ is equal to the KM estimate which we would obtain if we had $n+1$ observations, consisting of the original sample plus an observation (Z_{n+1},δ_{n+1}) such that $Z_{n+1}\geq Z_{(n)}$ and $\delta_{n+1} = 0$.

Let us now consider the case when there are regressors in the model. Suppose that our random sample consists of $\{(Z_i,\delta_i,X_i),\ 1\leq i\leq n\}$, where Z_i and δ_i are as before. It is now assumed that

$$T|X=x \text{ and } C|X=x \text{ are independent random variables almost surely,} \tag{2.7}$$

$$\forall x\in\mathbb{R}^p \text{ and } \forall\ t\in\mathbb{R},\ 1 - F^T(t|x) > 0 \text{ and } 1 - G(t|x) > 0. \tag{2.8}$$

where $F^T(.|x)$ and $G(.|x)$ denote now the conditional distribution functions of $T|X=x$ and $G|X=x$, respectively. If we denote $H_1(.|x)$, $H_2(.|x)$ and $\Lambda(.|x)$ the conditional subsurvival functions and cumulative hazard function, respectively (these are defined in a similar way to $H_1(.)$, $H_2(.)$ and $\Lambda(.)$), then similar equations to (2.3), (2.4) and (2.5) also hold. In order to obtain a similar estimate to the KM estimate, we now estimate $H_1(.|x)$, $H_2(.|x)$ using nonparametric kernel weights. Thus, for a given $x\in\mathbb{R}^p$, let us denote

$$B_{ns}(x) \equiv K((X_s-x)/h)/(\textstyle\sum_j K((X_j-x)/h)) \tag{2.9}$$

for a certain kernel function $K:\mathbb{R}^p \longrightarrow \mathbb{R}$, and a sequence $h \equiv h_n$ of smoothing values. We define now

$$\hat{H}_1(t|x) \equiv n^{-1}\textstyle\sum_j I(Z_j>t,\ \delta_j=1)B_{nj}(x),\ \hat{H}_2(t|x) \equiv n^{-1}\textstyle\sum_j I(Z_j>t)B_{nj}(x).$$

Then, the kernel-conditional KM estimate of the distribution function $F^T(t|x)$ of $T|X=x$ is

$$\hat{F}^T_{KC}(t|x) \equiv 1 - \textstyle\prod_{s\le t}(1 - d\hat{\Lambda}(s|x)),$$

where, now,

$$\hat{\Lambda}(t|x) \equiv - \int_{\infty}^{t}(\hat{H}_2(s-|x))^{-1}d\hat{H}_1(s|x).$$

The estimates $\hat{F}^T(.|x)$ and $\hat{\Lambda}(.|x)$ have been studied, among others, by Beran (1981) and Dabrowska (1987, 1989). As before, when there are no ties among the observations of Z, we may rewrite $\hat{F}^T_{KC}(t|x)$ as

$$\hat{F}^T_{KC}(t|x) = 1-\prod_{j=1}^{n}\left[\frac{\sum_s I(Z_s>Z_j)B_{ns}(x)}{\sum_s I(Z_s\ge Z_j)B_{ns}(x)}\right]^{\gamma_j(t)}. \tag{2.10}$$

We will assume that the kernel function K and the sequence of smoothing values h_n satisfy that

$$K(0)>0,\ K(u)=0\ \forall\ u\notin(-1,1)^p,\ \int K(u)du=1,\ \int u_jK(u)du=0,\ 1\le j\le p, \tag{2.11}$$

$$h_n \longrightarrow 0,\ nh_n^p \longrightarrow \infty,\ nh_n^{p+4} \longrightarrow 0,\ \text{as } n \longrightarrow \infty. \tag{2.12}$$

Assumptions (2.11) and (2.12) are introduced in order to make sure that $\hat{F}^T_{KC}(.|x)$ satisfies the weak and strong uniform consistency properties derived in Dabrowska (1989). If we let $h_n = Mn^{-\alpha}$ for some $\alpha>0$, $M>0$, then (2.12) holds if and only if $\alpha \in (1/(p+4)^{-1},1/p)$, that is, the smoothing value must converge to 0 faster than the optimal smoothing value h^{opt}_n in nonparametric estimation (which satisfies $h^{opt}_n = Mn^{-1/(p+4)}$).

As before, we will also consider the following variant of the kernel-conditional KM estimator,

$$\hat{F}^T_{KS}(t|x) \equiv 1-\prod_{j=1}^{n}\left[\frac{K(0)+\sum_s I(Z_s>Z_j)K((X_s-x)/h)}{K(0)+\sum_s I(Z_s\geq Z_j)K((X_s-x)/h)}\right]^{\gamma_j(t)} .$$

As $K(0)>0$, this estimate satisfies that $1-\hat{F}_{KS}(t|x)>0$. On the other hand, when there are no ties among the observations of Z, $\hat{F}^T_{KS}(t|x)$ coincides with the kernel-conditional KM estimate which we would obtain if we had $n+1$ observations: the original sample plus an observation $(Z_{n+1},\delta_{n+1},X_{n+1})$ such that $Z_{n+1}\geq Z_{(n)}$, $\delta_{n+1} = 0$ and $X_{n+1} = x$.

We derive now three procedures to estimate β in (1.1). Our procedures adapt those introduced by Buckley and James (1979), Koul et al. (1981) and Leurgans (1987).

2.2. Estimators based on Buckley and James procedure

2.2.1. Buckley and James procedure in the equal censoring model.

Buckley and James (1979) assume that $\{(x_i,c_i),\ 1\leq i\leq n\}$ are fixed variables[4]. Thus, equation (1.1) becomes

$$T_i = x_i'\beta + \varepsilon_i \qquad 1\leq i\leq n.$$

---

[4] Throughout this chapter we use capital letters to denote random variables and small letters to denote fixed non-random variables.

They also assume that $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed (i.i.d.) random variables with distribution function F^{ε}, and exploit the following linear relationship,

$$E[\delta_i Z_i + (1-\delta_i)H_i] = x_i'\beta, \tag{2.13}$$

where, if $\delta_i = 0$ then $H_i \equiv E[T_i | T_i > c_i] = x_i'\beta + E[\varepsilon_i | \varepsilon_i > c_i - x_i'\beta]$, and if $\delta_i = 1$ then H_i may be arbitrarily defined. Note that if $\delta_i = 0$ then $P(T_i > c_i) > 0$ and the expectation in H_i is well-defined. The idea behind the Buckley-James estimator is to replace, when $\delta_i = 0$, the unknown value $E[\varepsilon_i | \varepsilon_i > c_i - x_i'\beta]$ by a KM estimator. Specifically, let $\hat{\varepsilon}_j \equiv Z_j - x_j\hat{\beta}_0$ be estimated residuals obtained from an initial estimate $\hat{\beta}_0$ of β. It is possible to construct with them a KM estimate $\hat{F}^{\varepsilon}_{KM}(\hat{\beta}_0)$ of the distribution function $F^{\varepsilon}(.)$. We can estimate H_i by

$$\hat{H}_i \equiv x_i'\hat{\beta}_0 + \frac{\sum_j \hat{\varepsilon}_j I(\hat{\varepsilon}_j > c_i - x_i'\hat{\beta}_0) w_j(\hat{\beta}_0)}{\sum_j I(\hat{\varepsilon}_j > c_i - x_i'\hat{\beta}_0) w_j(\hat{\beta}_0)}, \tag{2.14}$$

where $w_j(\hat{\beta}_0)$ denotes the size of the jump in $\hat{\varepsilon}_j$ of the KM estimate $\hat{F}^{\varepsilon}_{KM}(\hat{\beta}_0)$. Now it is possible to obtain $\hat{Z}_i = \delta_i Z_i + (1-\delta_i)\hat{H}_i$. Equation (2.13) suggests that we could obtain a good estimate of β applying the least squares (LS) procedure to the data set $\{(\hat{Z}_i, x_i), i=1,\ldots,n\}$. This is precisely the Buckley-James estimator, $\hat{\beta}^{BJ} = \{\sum_i x_i x_i'\}^{-1}\sum_i x_i \hat{Z}_i$. Of course, iteration is possible and it may improve the performance of the estimate. Buckley and James (1979) suggest to use the LS estimate for all observations as initial value $\hat{\beta}_0$.

Buckley and James (1979) do not establish the asymptotic properties of their estimator. James and Smith (1984) studied its weak consistency assuming, among other conditions, that regressors and censoring variables are all non-random. Ritov (1990) and Lai and Ying (1991) proposed modified Buckley-James estimators and established their asymptotic properties using stochastic integral representations of their modified estimators. We do not follow their approach here. Instead, we will transform relation (2.13) to permit random censorship and discuss how we can use the resulting equalities to obtain estimates of β.

2.2.2. Buckley and James procedure in the unequal censoring model.

Given $x \in \mathbb{R}^p$, let T_x and C_x denote the conditional random variables $T|X=x$ and $C|X=x$, respectively and $F^T(.|x)$, $G(.|x)$ their distribution functions. There are two useful expressions which can be looked upon as generalisations of (2.13). On the one hand, given $x \in \mathbb{R}^p$ such that $P(C_x \le T_x) > 0$, denote

$$J(x) \equiv E[T_x | C_x \le T_x],$$

and $J(x)$ may be arbitrarily defined if x is such that $P(C_x \le T_x) = 0$ (for example $J(x) = 0$ if $P(C_x \le T_x) = 0$). Under certain conditions it can be shown (see Proposition 2 in the appendix) that

$$E[\delta Z + (1-\delta)J(X) | X=x] = x'\beta. \tag{2.15}$$

This is the most obvious way to generalise (2.13). We must now estimate $J(X_i)$ for those i such that $C_i \le T_i$. First of all, following Buckley and James procedure, if we also assume that the error term in (1.1) satisfies

$$\varepsilon \equiv T - X'\beta \text{ is independent of the regressors set } X. \tag{2.16}$$

We prove in the appendix (Proposition 3) that if x is such that $P(C_x \le T_x) > 0$, then

$$J(x) = x'\beta + \frac{\int sG(s+x'\beta|x)dF^{\varepsilon}(s)}{P(C_x \le T_x)} = x'\beta + \frac{\int sG(s+x'\beta|x)dF^{\varepsilon}(s)}{\int G(s+x'\beta|x)dF^{\varepsilon}(s)}. \tag{2.17}$$

With an initial estimate $\hat{\beta}_0$, we can construct $F^{\varepsilon}_{KM}(\hat{\beta}_0)$ as before. Additionally, we can reverse the roles of C and T and estimate $G(u|x)$ using a kernel-conditional KM estimate $\hat{G}_{KC}(u|x)$ as defined above. We can then consider

$$\mathcal{J}^{(1)}(x) \equiv x'\hat{\beta}_0 + \frac{\sum_j \hat{\varepsilon}_j \hat{G}_{KC}(\hat{\varepsilon}_j + x'\hat{\beta}_0 | x) w_j(\hat{\beta}_0)}{\sum_j \hat{G}_{KC}(\hat{\varepsilon}_j + x'\hat{\beta}_0 | x) w_j(\hat{\beta}_0)}, \tag{2.18}$$

where $w_j(\hat{\beta}_0)$ is as defined in (2.14). Observe that this estimate depends on an initial value $\hat{\beta}_0$ and on a smoothing value h. As before, it is possible to obtain the transformed data set $\{(\hat{Z}_i^{(1)},X_i),\ i=1,\dots,n\}$, where $\hat{Z}_i^{(1)}= \delta_i Z_i + (1-\delta_i)\hat{J}^{(1)}(X_i)$, and construct the LS estimate

$$\hat{\beta}_A^{BJ} \equiv (\textstyle\sum_i X_i X_i')^{-1}\sum_i X_i \hat{Z}_i^{(1)}. \tag{2.19}$$

Again, iteration is possible. We will use the LS estimate for the whole sample as initial value $\hat{\beta}_0$.

We can estimate $J(x)$ in an entirely different way without using any initial estimate of β. We prove in the appendix (Proposition 1) that if x is such that $P(C_x \le T_x)>0$, then $J(x)$ can be also written as

$$J(x) = \frac{\int uG(u|x)dF^T(u|x)}{\int G(u|x)dF^T(u|x)}. \tag{2.20}$$

We can estimate directly $G(u|x)$ and $F^T(u|x)$ from the original data set with kernel-conditional Kaplan-Meier estimates $\hat{F}^T_{KC}(.|x)$ and $\hat{G}^T_{KC}(.|x)$ and introduce them into (2.20) to obtain

$$\hat{J}^{(2)}(x) = \frac{\int u\hat{G}_{KC}(u|x)d\hat{F}^T_{KC}(u|x)}{\int \hat{G}_{KC}(u|x)d\hat{F}^T_{KC}(u|x)} = \frac{\sum_j Z_j \hat{G}_{KC}(Z_j|x)w_j^*}{\sum_j \hat{G}_{KC}(Z_j|x)w_j^*}, \tag{2.21}$$

where now w_j^* denotes the size of the jump of $\hat{F}^T_{KC}(.|x)$ in Z_j. As before, we can now apply the LS procedure to the transformed data set $\{(\hat{Z}_i^{(2)},X_i),\ i=1,\dots,n\}$, where $\hat{Z}_i^{(2)}= \delta_i Z_i + (1-\delta_i)\hat{J}^{(2)}(X_i)$ and obtain $\hat{\beta}_B^{BJ}$ whose expression is like the right-hand expression in (2.19), replacing $\hat{Z}_i^{(1)}$ by $\hat{Z}_i^{(2)}$.

Another equality which is a generalisation of (2.13) in our context may be obtained as follows: for $(t,x)\in\mathbb{R}\times\mathbb{R}^p$ such that $P(t\le T_x)>0$, denote $L(t,x) \equiv E[T_x|t\le T_x]$ and, as before, $L(t,x)$ may be arbitrarily defined if $P(t\le T_x)=0$. Then, we prove in the appendix (Proposition 5) that under certain conditions

$$E[\delta Z + (1-\delta)L(Z,X)|X=x] = x'\beta. \qquad (2.22)$$

If we want to use this expression in order to get an estimator of β, we must estimate the unknown values $L(Z_i,X_i)$ for those i such that $C_i \le T_i$. Note that (Proposition 4 in the appendix) if (t,x) is such that $P(t \le T_x)>0$, then

$$L(t,x) = \frac{\int uI(t\le u)dF^T(u|x)}{P(t\le T_x)} = \frac{\int uI(t\le u)dF^T(u|x)}{\int I(t\le u)dF^T(u|x)}. \qquad (2.23)$$

As before, we can replace $F^T(.|x)$ by its kernel-conditional KM estimate $\hat{F}^T_{KC}(.|x)$ and obtain

$$\hat{L}(t,x) = \frac{\sum_j Z_j I(t\le Z_j) w^*_j}{\sum_j I(t\le Z_j) w^*_j},$$

where w^*_j is as defined in (2.21). Again, we can define $\hat{Z}^{(3)}_i = \delta_i Z_i + (1-\delta_i)\hat{L}(Z_i|X_i)$ and obtain $\hat{\beta}^{BJ}_C$ applying LS to the transformed data set.

To sum up, we have obtained three different estimates for β by adapting Buckley and James procedure to our model. The first one depends on an initial estimate $\hat{\beta}_0$ and a smoothing value h; the second and the third ones depend on a smoothing value h but not on any initial estimate. In Section 3 we will examine and compare the performance of these estimates.

2.3. Estimator based on Koul, Susarla and Van Ryzin's procedure

Koul et al. (1981) assume that regressors are non-random and the distribution function of the censoring variable C_i is the same for all observations. They exploit the relation

$$E[\delta_i Z_i (1-G(Z_i))^{-1}] = x_i'\beta, \qquad (2.24)$$

where G denotes the distribution function of C_i, which is assumed to satisfy $1-G(t) > 0$ for all t. Then, they replace the unknown quantity $1-G(Z_i)$ by

$1-\hat{G}_{SV}(Z_i)$, where $\hat{G}_{SV}(t)$ is as defined above reversing the roles of C and T (that is to say, considering as uncensored those observations satisfying $C_i \le T_i$). A trimming function, trimming out large Z values, is also introduced. This trimming is necessary because the explosive behaviour of the asymptotic variance of $\hat{G}(t)$ for large t might worsen dramatically the performance of the estimate. Koul et al. (1981) prove that, under certain regularity conditions, this estimate is consistent and asymptotically normal. However, in simulation studies and empirical applications it seems to perform very poorly (see Miller and Halpern 1982, Leurgans 1987, Heller and Simonoff 1990 and Gonzalez-Manteiga and Cadarso-Suarez 1991).

In our model, (2.24) may be transformed to yield (see Proposition 6 in the appendix),

$$E[\delta Z(1-G(Z|X))^{-1}|X=x] = x'\beta \qquad (2.25)$$

where now $G(.|x)$ denotes the distribution function of $C|X=x$, which is assumed to satisfy (2.8). In view of (2.25), we are motivated to define $\hat{Z}_i^K = \delta_i Z_i(1-\hat{G}_{KS}(Z_i|X_i))^{-1}$, where $\hat{G}_{KS}(.|x)$ is as defined above reversing, again, the roles of C and T. As mentioned above, we use $\hat{G}_{KS}(.|x)$ rather than $\hat{G}_{KC}(.|x)$ because the former satisfies $1-\hat{G}_{KS}(t|x) > 0$ for all t. The corresponding LS estimate is $\hat{\beta}^K = (\sum_i X_i X_i' I_i)^{-1}\sum_i X_i \hat{Z}_i^K I_i$ where $I_i = I(Z_i \le M)$ for a sequence $M \equiv M_n$ of trimming values. The asymptotic behaviour of this estimate is discussed in Section 3.

2.4. Estimator based on Synthetic Data

Leurgans (1987) introduced the use of synthetic data in the estimation of linear regression models with random censoring. We discuss here how this procedure is implemented in our model.

The synthetic data procedure arises by generalising a well-known property of classical least-squares estimation. In a linear regression model like (1.1), given $n \in \mathbb{N}$ and $b \in \mathbb{R}^k$, let U_n be a discrete uniform random variable with support $\{1, ..., n\}$, independent of (T_j, X_j), $1 \le j \le n$. Define

$$\xi_{nb} \equiv \sum_j I(U_n=j)(T_j-X_j'b).$$

Thus, ξ_{nb} is a random variable which is equal to $T_j-X_j'b$ with probability n^{-1} $(1\le j\le n)$. Denote $\Omega_n \equiv \{X_1=x_1, \ldots, X_n=x_n\}$. The distribution function of $\xi_{nb}|\Omega_n$ is $H_{nb}(t|\Omega_n) = n^{-1}\sum_j F_j^T(t+b'x_j) = n^{-1}\sum_j F_j^{\varepsilon}(t+(b-\beta)'x_j)$, where $F_j^T(.)$ and $F_j^{\varepsilon}(.)$ denote the distribution functions of $T_j|X_j=x_j$ and $\varepsilon_j|X_j=x_j$, respectively. If we assume that the error term in (1.1) satisfies that

$$E[\varepsilon^2|X] = \sigma^2 \in (0,\infty) \text{ a.s.}, \tag{2.26}$$

then $E[\xi_{nb}^2|\Omega_n] = \sigma^2 + (b-\beta)'(n^{-1}\sum_j x_j x_j')(b-\beta)$, and

$$\beta = \underset{b}{argmin}\ E[\xi_{nb}^2|\Omega_n] = \underset{b}{argmin} \int t^2 dH_{nb}(t|\Omega_n). \tag{2.27}$$

Now, if we replace in $H_{nb}(t|\Omega_n)$ the unknown quantities $F_j^T(t+b'x_j)$ by their naive estimates $I(T_j\le t+b'X_j)$ we obtain $\tilde{H}_{nb}(t|\Omega_n)$. If in (2.27) we replace $H_{nb}(.|\Omega_n)$ by $\tilde{H}_{nb}(.|\Omega_n)$ we obtain $\tilde{\beta} = \underset{b}{argmin}\ n^{-1}\sum_j(T_j-b'X_j)^2$, which is precisely the OLS estimator. In the presence of random censorship, the natural way to generalise this property is as follows: as before, let $G(.|x)$ denote the distribution function of $C|X=x$; suppose that (2.7) and (2.16) hold. If we denote $G_j(.) \equiv G(.|x_j)$, then

$$\begin{aligned} H_{nb}(t|\Omega_n) &= 1 - n^{-1}\sum_j P(T_j>t+b'X_j|\Omega_n) = \\ &= 1 - n^{-1}\sum_j P(Z_j>t+b'X_j|\Omega_n)/(1-G_j(t+b'x_j)). \end{aligned} \tag{2.28}$$

With our random sample $\{(Z_i,\delta_i,X_i),\ 1\le i\le n\}$, a naive estimate for $P(Z_j>t+b'X_j|\Omega_n)$ is $I(Z_j>t+b'x_j)$, and we can estimate $G_j(.)$ using a kernel-conditional KM estimate $\hat{G}_{KC}(.|x_j) \equiv \hat{G}_j(.)$ (considering as uncensored those observations for which we observe C_i, i.e., those observations such that $\delta_i = 0$). Replacing $P(Z_j>t+b'X_j|\Omega_n)$ and $G_j(t+b'x_j)$ by these estimates in the last expression of (2.28) we obtain

$$\hat{H}_{nb}(t|\Omega_n) = 1 - n^{-1}\sum_j \hat{H}_{nj}(t+b'x_j),$$

where $\hat{H}_{nj}(u) \equiv I(Z_j>u)/(1-\hat{G}_j(u))$. Observe that $\hat{H}_{nj}(u)$ is a well-defined function: $1-\hat{G}_j(u) = 0 \Rightarrow u \geq Z_{(n)} \Rightarrow I(Z_j>u) = 0 \Rightarrow \hat{H}_{nj}(u) = 0$. The natural way to generalise OLS estimation is to define

$$\hat{\beta}^{SD} \equiv \underset{b}{argmin} \int t^2 d\hat{H}_{nb}(t|\Omega_n) = \underset{b}{argmin} -n^{-1}\sum_j\int(u-b'X_j)^2 d\hat{H}_{nj}(u). \tag{2.29}$$

Let us analyse $\hat{H}_{nj}$. This is a step-function whose discontinuity points are $\{Z_j\} \cup \mathcal{F}_j$, where $\mathcal{F}_j \equiv \{Z_i : \delta_i=0, Z_i<Z_j\}$. $\hat{H}_{nj}(u)$ is 1 until its first discontinuity point, non-decreasing in $(-\infty,Z_j)$ and 0 in $[Z_j,\infty)$. For $j\neq i$, denote $w_j(i)$ the size of the jump up of $\hat{H}_{nj}$ in Z_i ($w_j(i)$ will be zero if $i\notin\mathcal{F}_j$) and W_j the size of the jump down of $\hat{H}_{nj}$ in Z_j. Then, (2.29) may be rewritten as

$$\hat{\beta}^{SD} \equiv \underset{b}{argmin}\ n^{-1}\sum_j\{W_j(Z_j-b'X_j)^2 - \sum_{i\neq j}w_j(i)(Z_i-b'X_j)^2\}. \tag{2.30}$$

When there are no ties among the observations of Z, it is possible to obtain a simple expression for (2.30). Taking into account that (2.10) holds, algebraic manipulation in (2.30) shows (see appendix) that

$$\hat{\beta}^{SD} = \{\sum_j X_{(j)}X'_{(j)}\}^{-1}\sum_j X_{(j)}\hat{Z}^{SD}_{(j)}, \tag{2.31}$$

where $X_{(j)}$ denotes the observation which corresponds to $Z_{(j)}$, which is the jth order statistic of the Z's and the "synthetic data" $\hat{Z}^{SD}_{(j)}$ are defined as

$$\hat{Z}^{SD}_{(j)} = \sum_{i=1}^{j}\tilde{W}_j(i)(Z_{(i)}-Z_{(i-1)}), \tag{2.32}$$

where $Z_{(0)} \equiv 0$, $\tilde{W}_j(1) = 1$ and for $1<i\leq j\leq n$

$$\tilde{W}_j(i) \equiv \prod_{k=1}^{i-1}\left[1 + \frac{B_{n(k)}(X_{(j)})}{\sum_{s=k+1}^{n}B_{n(s)}(X_{(j)})}\right]^{I(\delta_k=0)}, \tag{2.33}$$

and now $B_{n(s)}(x)=K((X_{(s)}-x)/h)/\sum_j K((X_{(j)}-x)/h)$.

The expression (2.32) which we have obtained is more complicated than the expression deduced by Leurgans (1987) in the context of non-random regressors and equal censoring for all observations. However, the relation between the original data and the synthetic data is the same: if no times are censored, $\tilde{W}_j(i)$ is identically equal to 1 and the synthetic data equal the original data; when there are censored observations, the gaps between consecutive observed times are magnified (see Leurgans 1987).

It has not been established yet under what conditions $\hat{\beta}^{SD}$ is a consistent estimate of β. Leurgans (1987) proved that her estimator (whose expression is simpler than (2.32) because their assumptions are more restrictive) is consistent and asymptotically normal when it is used to analyse the difference between two means, but its consistency was not discussed in general linear regression models. In Section 3 we will examine the performance of $\hat{\beta}^{SD}$ in various special models.

2.5. Other estimators

In this section we present three other estimation procedures which can be implemented in our model. They have been already introduced by Miller (1976), Chatterjee and McLeish (1986) and Gonzalez-Manteiga and Cadarso-Suarez (1991). They will be referred to as "Miller estimator", "Linear Attribute Method (LAM) Estimator" and "Gonzalez-Manteiga and Cadarso-Suarez (GC) Estimator".

2.5.1. Miller estimator.

It is well-known that in the linear regression model (1.1)

$$\beta = \underset{b}{argmin} \int u^2 dF^b(u) \tag{2.34}$$

where, for $b \in \mathbb{R}^p$, F^b denotes the distribution function of the random variable $T-X'b$. In the absence of censorship, the OLS estimator is obtained by replacing in (2.34) the unknown distribution function F^b by the empirical distribution function of $T-X'b$ constructed from the sample. When censorship is present, F^b may be replaced by its KM estimator; thus we would obtain what Miller (1976) termed "Kaplan-Meier LS estimator",

$$\hat{\beta}^{KM-OLS} = \underset{b}{argmin} \sum_j w_j(b)(Z_j - X_j'b)^2,$$

where $w_i(b)$ denotes the size of the jump in $Z_i - X_i'b$ of the KM estimate $\hat{F}^b_{KM}$. This procedure is not useful in practice because the resulting optimization problem is so involved that numerical computation of the estimate becomes extremely difficult. Miller suggested a procedure to circumvent this difficulty when the error terms are supposed to be i.i.d. He suggested to obtain estimated residuals with an initial estimate $\hat{\beta}_0$ of β; then, it is possible to construct with them a KM estimate $\hat{F}^{\varepsilon}_{KM}(\hat{\beta}_0)$ of the distribution function of the error term $F^{\varepsilon}(.)$. Now we can consider

$$\hat{\beta}^{MI} = \underset{b}{argmin} \sum_j w_j(\hat{\beta}_0)(Z_j - X_j'b)^2,$$

where $w_j(\hat{\beta}_0)$ is as defined in (2.14). This estimator is usually referred to as Miller estimator; it is much easier to compute than $\hat{\beta}^{KM-OLS}$. Miller suggested to use the OLS estimate for the uncensored data as initial value $\hat{\beta}_0$.

Obviously, iteration is possible and it may improve the performance of the estimate. Though Miller proposed this method for the fixed regressors model, his estimator is also applicable to our stochastic regressors context: observe that, in order to justify the expression for $\hat{\beta}^{MI}$ it is not necessary to assume that X or C are fixed non-random variables or that the distribution functions of $C|X=x_i$ are all the same.

Due to the complexity of the optimization problem which has to be solved to obtain $\hat{\beta}^{MI}$, conditions under which this estimate is consistent have not been established yet. In simulations, it seems to perform adequately in fixed regressors models in which the censoring pattern is generated in a realistic way (see Miller and Halpern 1982, or Heller and Simonoff 1990). However, it is possible to produce examples in which $\hat{\beta}^{MI}$ is not consistent (see Miller 1976).

2.5.2. Linear Attribute Method (LAM) estimator.

This method was first proposed by Chatterjee and McLeish (1986); it is also iterative. Suppose we have an initial estimate $\hat{\beta}_0$ (they suggest to use the OLS estimate for the whole sample as an initial value). Chatterjee and McLeish proposed to estimate β applying OLS to the transformed data set $\{(Z_i^C, X_i), 1 \le i \le n\}$, where $Z_i^C = \delta_i Z_i + (1-\delta_i)max(Z_i, X_i'\hat{\beta}_0)$. This estimate, which we will denote $\hat{\beta}^{LAM}$, seems reasonable. When $\delta_i = 0$, we do not know the value of T_i, but we do know that $T_i > Z_i$; therefore, we will replace the unknown value T_i by the fitted value $X_i'\hat{\beta}_0$ if the latter is greater than Z_i; otherwise we will simply replace T_i by Z_i. Again, iteration is possible.

Chatterjee and McLeish (1986) do not study the asymptotic properties of this estimate when the distribution function of the error term is unspecified. Heller and Simonoff (1990) propose a "modified linear attribute method (MLAM) estimator", but they do not give a proper justification for their modification, so we will not consider it here. Neither did they discuss the consistency of the LAM or MLAM estimators. We will examine the performance of the LAM estimator in Section 3.

2.5.3. Gonzalez-Manteiga and Cadarso-Suarez (GC) estimator.

Gonzalez-Manteiga and Cadarso-Suarez (1991) proposed an estimator of β based on prior nonparametric estimation of the regression function. We may obtain the expression for the GC estimator by generalising the procedure proposed by Koul et al. (1981) in a different way to the one we discussed in Section 2.3, although Gonzalez-Manteiga and Cadarso-Suarez (1991) did not originally deduce their estimator in this way.

Let us denote $\alpha(\delta,Z,X) = \delta Z(1-G(Z|X))^{-1}$ and $m(x) = E[\alpha(\delta,Z,X)|X=x]$. From (2.25) we know that $m(x) = x'\beta$; hence, $E[(m(X)-X'\beta)^2] = 0$ and, under regularity conditions,

$$\beta = \underset{b}{argmin}\ E[(m(X)-X'b)^2]. \tag{2.35}$$

The right-hand expression in (2.35) cannot be directly estimated by its sample analog because $m(.)$ is unknown. This problem may be solved in two steps: first, observe that $m(.)$ is a conditional expectation, so it could be estimated using any standard nonparametric procedure if $\alpha(.,.,.)$ were known; then, $\alpha(.,.,.)$ is not known because we ignore $G(.|x)$, but we may replace this quantity by a kernel-conditional KM estimate $\hat{G}_{KC}(.|x)$. Thus, let us define $\hat{\alpha}(\delta,Z,X) = \delta Z(1-\hat{G}_{KC}(Z|X))^{-1})$ and

$$\hat{m}(X) = \sum_s \hat{\alpha}(\delta_s,Z_s,X)B_{ns}(X), \tag{2.36}$$

where $B_{ns}(.)$ is as defined in (2.9)[5]. In view of (2.35) we are motivated to consider

$$\hat{\beta}^{GC} = \underset{b}{argmin}\ n^{-1}\sum_j(\hat{m}(X_j)-X_j'b)^2,$$

which is to say, $\hat{\beta}^{GC}$ is the estimate we obtain applying the OLS procedure to the transformed data set $\{(\hat{m}(X_j),X_j),\ 1\le j\le n\}$.

Gonzalez-Manteiga and Cadarso-Suarez (1991) give no proof of the consistency or asymptotic normality of $\hat{\beta}^{GC}$. However, in Gonzalez-Manteiga and Cadarso-Suarez (1994) a similar estimator to $\hat{\beta}^{GC}$ is introduced in the context of non-random regressors and its asymptotic properties are established. It is expected that $\hat{\beta}^{GC}$ behaves in a similar way as the estimator which is discussed in Gonzalez-Manteiga and Cadarso-Suarez (1994).

3. SIMULATION STUDY

We have generated observations from six different models and computed the various estimators which have been described in Section 2. On generating models we have tried to cover as wide a variety of situations as possible. Specifically, models have been selected in such a way that it is possible to analyse the influence on all estimators of the following characteristics: the distribution function of the censoring variable; the coefficient of determination R^2; and the degree of censorship (DC), that is, the proportion of censored observations in the sample.

In Models 1-4, $(\varepsilon_1,\varepsilon_2)' \sim N((0,0)',\sigma^2 I_2)$, $X \sim N(0,1)$ and X is independent of $(\varepsilon_1,\varepsilon_2)'$. Additionally, $T = \beta_1 + \beta_2 X + \varepsilon_1$, $C = \gamma + X_2 + \varepsilon_2$ where $\beta_1 = 2$, $\beta_2 =$

---

[5]Observe that the smoothing value used in (2.36) is not necessarily the same as the smoothing value used to compute $\hat{G}_{KC}(Z|X)$.

1 in all models and,

Model 1: $\gamma = 2$; $\sigma^2 = 1$; (DC = 50%; $R^2 = 0.5$).

Model 2: $\gamma = 4$; $\sigma^2 = 1$; (DC = 8%; $R^2 = 0.5$).

Model 3: $\gamma = 2.5$; $\sigma^2 = 0.25$; (DC = 24%; $R^2 = 0.8$).

Model 4: $\gamma = 4$; $\sigma^2 = 4$; (DC = 24%; $R^2 = 0.2$).

In Model 5, $(\varepsilon_1,\varepsilon_2)'$, X and T are as before with $\sigma^2 = 1$ and C is independent of the former and uniformly distributed in $(0,6)$. Finally, Model 6 is a multiple regression model. In this case we have $T= \beta_1+ \beta_2X_1+ \beta_3X_2+ \varepsilon_1$, with $\beta_1= 2$, $\beta_2= \beta_3= 1$, $C= 2+ X_1+ X_2+ \varepsilon_2$ and $(X_1,X_2,\varepsilon_1,\varepsilon_2)' \sim N(0,I_4)$. Thus,

Model 5: $\sigma^2 = 1$; (DC = 24%; $R^2 = 0.5$).

Model 6: $\sigma^2 = 1$; (DC = 50%; $R^2 = 2/3$).

We wish to study both the finite-sample properties and the asymptotic behaviour of all estimators. Hence, we have considered two different sample sizes: $n=50$ and $n=400$. There were 1000 replications for each simulation run when $n=50$; when $n=400$ there were 500 replications except for $\hat\beta^{BJ}_A$, $\hat\beta^{BJ}_B$, $\hat\beta^{BJ}_C$ and $\hat\beta^{GC}$ -for these estimates we only run 100 replications.

In Section 2 we have discussed the following estimators: $\hat\beta^{BJ}_A$, $\hat\beta^{BJ}_B$, $\hat\beta^{BJ}_C$, $\hat\beta^{K}$, $\hat\beta^{SD}$, $\hat\beta^{MI}$, $\hat\beta^{LAM}$ and $\hat\beta^{GC}$. Three of these estimators ($\hat\beta^{BJ}_A$, $\hat\beta^{LAM}$ and $\hat\beta^{MI}$) require an initial estimate: $\hat\beta^{BJ}_A$ and $\hat\beta^{LAM}$ were computed using as $\hat\beta_0$ the OLS estimate for the whole sample; $\hat\beta^{MI}$ was computed in two different ways: in $\hat\beta^{MI-A}$ we used as $\hat\beta_0$ the OLS estimate for the uncensored sample and in $\hat\beta^{MI-B}$ we used as initial estimate $\hat\beta^{K}$, computed using the trimming value M and the smoothing value h_1 which are specified below. All nonparametric estimates have been computed using the univariate Epanechnikov kernel (see Section 2.3.4) as kernel function (in Model 6 the kernel function we used was the product of two univariate Epanechnikov kernels). Five of the estimators ($\hat\beta^{BJ}_A$, $\hat\beta^{BJ}_B$, $\hat\beta^{BJ}_C$, $\hat\beta^{K}$ and $\hat\beta^{SD}$) required the use of a smoothing value h; in all cases three different h were used: $h_1 = 3n^{-1/4}$, $h_2= 8n^{-1/4}$ and $h_3 = 1000$. The latter amounts to the same thing as if we considered $G_1 = G$ and estimated this function with a KM estimate rather than with a kernel-conditional KM estimate. The values h_1 and

h_2 have been selected inspecting the performance of the nonparametric estimates in a small number of simulated data; we use $n^{-1/4}$ because this rate of convergence satisfies all assumptions in Dabrowska (1989) (note that the usual rate $n^{-1/5}$ does not satisfy the assumption required in Corollary 2.2.iii of this paper). One of the estimators ($\hat{\beta}^{GC}$) requires the use of two smoothing values $(h^A,h^B) \equiv h^*$, where h^A is used to compute $\hat{G}_{KM}(.|x)$ and h^B is used to compute $\hat{m}(X)$ as defined in (2.36); as before, we used three different h^*: h_1^A, h_2^A, h_3^A are the same as h_1, h_2, h_3 above, and $h_1^B \equiv 2n^{-1/4}$, $h_2^B \equiv 3n^{-1/4}$, $h_3^B = 4n^{-1/4}$. Finally, $\hat{\beta}^K$ requires the use of a trimming value; as Koul et al. (1981) suggest we used $M \equiv 3(log(n))^{2/5}$.

In Tables 1-5 we report the bias and variance (VAR) of the estimates of the intercept (β_1) and slope (β_2). The columns labelled "MSE" contain $10000 \times \{MSE(\hat{\beta}_1) + MSE(\hat{\beta}_2)\}$ (MSE denotes Mean Squared Error). In Table 6 we report corresponding results for Model 6; note that in this Model the slope is (β_2,β_3); in this table the column labelled "MSE" contains $10000 \times \{MSE(\hat{\beta}_1) + MSE(\hat{\beta}_2) + MSE(\hat{\beta}_3)\}$, and the columns labelled "Slope-Bias" and "Slope-VAR" contain $\{|Bias(\hat{\beta}_2)| + |Bias(\hat{\beta}_3)|\}/2$ and $\{VAR(\hat{\beta}_2) + VAR(\hat{\beta}_3)\}/2$, respectively.

The main conclusion from the results in Tables 1-6 is that $\hat{\beta}_A^{BJ}$ and $\hat{\beta}^{SD}$ are the preferred estimators. In fact, comparison of all estimators from the point of view of mean squared error provides the following ranking,

Group 1: $\hat{\beta}_A^{BJ}$, $\hat{\beta}^{SD}$;
Group 2: $\hat{\beta}_B^{BJ}$, $\hat{\beta}^{MI}$;
Group 3: $\hat{\beta}_C^{BJ}$; $\hat{\beta}^{GC}$;
Group 4: $\hat{\beta}^K$,

where the estimators in Group i are preferable to those in Group j whenever $i<j$, regardless of the characteristics of the simulated model. Observe that $\hat{\beta}^{LAM}$ is not included in this ranking -we discuss below the behaviour of this estimator.

In most models, $\hat{\beta}^{SD}$ performs slightly better than $\hat{\beta}_A^{BJ}$, but the behaviour of the former depends heavily on the smoothing value, whereas the choice of this value is not so crucial when using $\hat{\beta}_A^{BJ}$. If we look further into Tables 1-6, when comparing $\hat{\beta}^{SD}$ and $\hat{\beta}_A^{BJ}$ we observe that the former is generally more biased than the latter when estimating the slope coefficient, but $\hat{\beta}^{SD}$ is less biased than $\hat{\beta}_A^{BJ}$ when estimating the intercept coefficient. However, this does not seem to imply that $\hat{\beta}_A^{BJ}$ behaves better than $\hat{\beta}^{SD}$ in multiple regression (see Table 6), because the higher bias of $\hat{\beta}^{SD}$ for the slope coefficients is compensated by a lower variance of this estimate for the intercept coefficient.

The two other estimators which have been introduced following Buckley and James's procedure ($\hat{\beta}_B^{BJ}$, $\hat{\beta}_C^{BJ}$) perform worse than $\hat{\beta}_A^{BJ}$ in all models, and the difference between the latter and the two others is remarkable under a high degree of censorship is high (Models 1 and 6). In fact, this is not a surprising feature of the estimators, because $\hat{\beta}_A^{BJ}$ uses more information from data (it requires the use of an initial value), and in the models we simulated the information provided by this previous estimation was accurate.

Miller estimator performs almost as well as those in Group 1 only when the degree of censorship is low (Models 2 and 3). The main advantage of this estimate is that it requires no selection of smoothing value. Moreover, the results in our simulation are very similar irrespective of the initial value which was used. Another advantage of Miller estimator is that it is computationally much cheaper than the others.

The estimator introduced by Gonzalez-Manteiga and Cadarso-Suarez ($\hat{\beta}^{GC}$) and the estimator based on Koul, Susarla and Van Ryzin's procedure ($\hat{\beta}^{K}$) perform very poorly when the sample size is $n=50$; when the sample size is $n=400$ their behaviour is closer to that of the other estimators. Their main drawback is that they both are based on the equation $E[\delta Z(1-G(Z|X))^{-1}|X=x] = x'\beta$, and, hence, their properties depend crucially on the estimation of $(1-G(Z|X))^{-1}$; in practice, this value is very poorly estimated and, as a result, $\hat{\beta}^{GC}$ and $\hat{\beta}^{K}$ do not perform well. The behaviour of the former is usually better than the behaviour of the latter because in the former there is a second process of smoothing which partly reduces the negative effects which the poor estimate of $(1-G(Z|X))^{-1}$ causes. In fact, in some cases (Models 2, 4 and 6) $\hat{\beta}^{GC}$ performs almost as well as Miller estimator. As $\hat{\beta}^{GC}$ depends on two smoothing values, it is likely that its performance improves with data-driven bandwidths and, then, be comparable to those who appear as better in our previous ranking.

The estimator based on the linear attribute method ($\hat{\beta}^{LAM}$) seems to perform well when the degree of censorship is low (especially in Model 2) but very badly otherwise. It is easy to explain the results which we have obtained for this estimator. We must just take into account that $\hat{\beta}^{LAM}$ slightly modifies the OLS estimator. In Models 1-4 and 6, the OLS procedure yields a consistent estimate for the slope, but an inconsistent estimate for the intercept. And the bigger the degree of censorship, the greater the difference between β_1 and the limit in probability of its OLS estimate. The estimator $\hat{\beta}^{LAM}$ simply reduces the negative effects of censorship, but does not remove them, that is to say, the corresponding estimator of the intercept is less biased than the OLS estimator but not consistent. In Table 2 (very low degree of censorship) we observe that $\hat{\beta}^{LAM}$ is actually the best estimate when $n=40$, but the bias of the estimate of the intercept does not converge to *0*. In Tables 1, 3, 4 and 6 we observe again that this is an unbiased estimate of the slope but a biased estimated of the intercept (exactly the same as the OLS estimator), and the bigger the degree of censorship the worse $\hat{\beta}^{LAM}$ performs (again, the same as the OLS estimator). Finally, in Model 5 both the slope and the intercept estimates are biased and their bias does not converge to *0* because neither the OLS estimate for the intercept nor the OLS estimate for the slope are consistent. To sum up, the linear attribute method can be advisable for a practitioner as an easy-to-compute procedure which produces estimates which are close to the true parameters when the degree of censorship is low (what is not a surprise because the higher the degree of censorship, the more serious is the problem of inconsistency); but it cannot be seriously taken into account because it does not produce consistent estimates.

The results of our Monte Carlo experiment also provide the following further information:

(a) All methods are more effective under a stronger regression model, but this feature does not affect the relative behaviour of the estimators.

(b) As expected, all methods are more effective when the degree of censorship is small. And, except for $\hat{\beta}^{LAM}$, this feature of the underlying model does not affect either the relative behaviour of the estimates.

(c) When censoring is equal for all observations (Model 5), the four procedures designed for unequal censoring (all except the linear attribute method and Miller's estimator) continue to perform adequately. However, if censoring is unequal for all observations (Models 1-4 and 6) and this fact is ignored, then the performance of the estimates may worsen dramatically (see, for instance, the bad performance of $\hat{\beta}^{SD}$ when the smoothing value is h_3, that is, when it was not taken into account that $G(.|x)$ varies with x). Therefore, it seems highly advisable to use a method which permits for unequal censoring when the researcher is not sure of the equal censoring assumption.

(d) In multiple regression models all estimators worsen their behaviour, but, again, their relative performance remains unchanged.

4. CONCLUDING REMARKS

The objective of this chapter is to propose estimation procedures in the semiparametric linear regression model with censored data and stochastic regressors. Specifically, two important assumptions made in the literature on the topic have been relaxed: we do not consider that regressors are fixed non-random variables, and we do not consider, either, that the distribution function of the censoring variable is the same for all observations ("equal censoring"). The former assumption does not affect the estimation procedures, but the latter does influence the behaviour of the estimators. We have shown that combining kernel-conditional KM estimators with previous techniques, the resulting estimators perform much better than the original ones. And this fact is meaningful because the assumption of equal censoring is not very realistic in practice. In most situations where these models apply (for example, the Stanford heart transplant program data contained in Miller and Halpern 1982, or the survival times of two groups of cancer patients contained in Leurgans 1987) it is unreasonable to assume that censoring is not affected by the characteristics of individuals. As our simulations show, inferences based on this false assumption may be entirely wrong[6], producing invalid results (see,

---

[6] Leurgans (1987), aware that the unequal censoring assumption is unreasonable in her examples, proposed to group data and then apply the synthetic data procedure. In fact, the procedure we propose in section

for example, the estimate obtained with Koul et al. procedure in Leurgans 1987, Table 1). The alternatives proposed in this chapter seem to provide good results.

Our study confirms some previous results. It was already known that in the model with equal censoring for all observations the estimator based on Koul, Susarla and Van Ryzin's procedure performs badly (see, for instance, the first example provided by Leurgans 1987), whereas the estimator based on Buckley and James's procedure seems to be the preferred one (see Heller and Simonoff 1990). But our conclusions are beyond this. Our results show that it is necessary take into account the possibility of unequal censoring among observations and we describe how estimation can be performed in this situation. Under equal censoring, the new estimators permitting for unequal censoring produce results which are comparable to those obtained with estimators designed for equal censoring (see Table 5), but the converse is not true.

However, it must not be forgotten that we have made no attempt to analyse the influence of the smoothing value (bandwidth). Further research is required on the problem of optimal selection of this value in this model. Our results might be affected by this selection because most of the procedures we have studied require a smoothing value. The performance of estimates may improve with data-driven bandwidths.

APPENDIX.- Proofs

Proposition 1.- If $E|T|<\infty$, (1.2), (2.7) hold and $x\in\mathbb{R}^p$ is such that $P(C_x\le T_x)>0$, then (2.20) holds.

PROOF: Given $s\in\mathbb{R}$, $P(T_x\le s|C_x\le T_x) = \psi_x(s)/P(C_x\le T_x)$, where

$$\psi_x(s) \equiv \int I(u\le s)\{\int I(v\le u)dG(v|x)\}dF^T(u|x) = \int I(u\le s)G(u|x)dF^T(u|x).$$

2.3 generalises this procedure in a less naive way.

On the other hand,

$$P(C_x \le T_x) = \int\{\int I(v \le u) dG(v|x)\} dF^T(u|x) = \int G(u|x) dF^T(u|x).$$

Thus, with the notation introduced in Section 2.2,

$$J(x) = E[T_x | C_x \le T_x] = \frac{\int u d\psi_x(u)}{P(C_x \le T_x)} = \frac{\int u G(u|x) dF^T(u|x)}{\int G(u|x) dF^T(u|x)}. \quad \blacksquare$$

Proposition 2.- With the same conditions as in previous proposition, if also (1.1) holds, then (2.15) holds.

PROOF: Given $(a,b) \in \mathbb{R}\times\mathbb{R}$, denote $f(a,b) \equiv a \times I(a<b)$. Then $\delta Z = f(T,C)$. As $E|T|<\infty$ and $\delta Z \le T$ *a.s.*, we can obtain

$$E[\delta Z | X=x] = E[f(T,C)|X=x] = \int\{\int f(u,v) dG(v|x)\} dF^T(u|x) =$$

$$\int u\{\int I(u<v) dG(v|x)\} dF^T(u|x) = \int u(1-G(u|x)) dF^T(u|x). \tag{A.1}$$

On the other hand, $E[(1-\delta)J(X)|X=x] = J(x) \times P(C_x \le T_x)$, and both factors were obtained in proposition 1. Hence,

$$E[(1-\delta)J(X)|X=x] = \int u G(u|x) dF^T(u|x). \tag{A.2}$$

Now, (2.15) follows from (A.1), (A.2) and (1.1). ■

Proposition 3.- With the same conditions as in previous proposition, if also (2.16) holds, then (2.17) holds.

PROOF: From (1.1) and (2.16) we deduce that $F^T(u|x) = F^\varepsilon(u-x'\beta)$. Thus, this proposition follows from proposition 1 if we simply make the change of variable $s=u-x'\beta$. ■

Proposition 4.- If $E|T|<\infty$, (1.2) holds and $(t,x)\in\mathbb{R}\times\mathbb{R}^p$ is such that $P(t \le T_x)>0$, then (2.23) holds.

PROOF: Given $s\in\mathbb{R}$, $P(T_x \le s | t \le T_x) = \varphi_{tx}(s)/P(t \le T_x)$, where $\varphi_{tx}(s) \equiv \int I(t<u\le s) dF^T(u|x)$. Thus, with the notation introduced in Section 2.2,

$$L(t,x) = E[T_x | t \le T_x] = \frac{\int u d\varphi_{tx}(u)}{P(t \le T_x)} = \frac{\int u I(t \le u) dF^T(u|x)}{\int I(t \le u) dF^T(u|x)}. \qquad ■$$

Proposition 5.- With the same conditions as in previous proposition, if also (1.1) and (2.7) hold and x is such that $E[(1-\delta)|L(C,X)| \,|X=x]<\infty$, then (2.22) holds.

PROOF: Given $(a,b,c) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}^p$, denote $g(a,b,c) \equiv L(b,c) \times I(b \le a)$. Then $(1-\delta)L(Z,X) = g(T,C,X)$. As $E[|g(T,C,X)| \,|X=x]<\infty$, we can obtain

$$E[(1-\delta)L(Z,X)|X=x] = E[g(T,C,X)|X=x] =$$

$$= \int\left(\int g(u,v,x)dF^T(u|x)\right)dG(v|x) = \int L(v,x)\left(\int I(v \le u)dF^T(u|x)\right)dG(v|x).$$

By (2.23), $L(v,x)\left(\int I(v \le u)dF^T(u|x)\right) = \int uI(v \le u)dF^T(u|x)$. Hence,

$$E[(1-\delta)L(Z,X)|X=x] = \int\left(\int uI(v \le u)dF^T(u|x)\right)dG(v|x) =$$

$$\int u\left(\int I(v \le u)dG(v|x)\right)dF^T(u|x) = \int uG(u|x)dF^T(u|x) \qquad \text{(A.3)}$$

Now, (2.22) follows from (A.1), (A.3) and (1.1). ■

Proposition 6.- If (1.1), (1.2), (2.7), (2.8) hold and $E[|T|(1-G(T|X))^{-1}|X=x]<\infty$, then (2.25) holds.

PROOF: For $(a,b,c) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}^p$, denote $h(a,b,c) \equiv a(1-G(a|c))^{-1} \times I(a<b)$. Then $\delta Z(1-G(Z|X)^{-1} = h(T,C,X)$. As $E[|T|(1-G(T|X))^{-1}|X=x]<\infty$, then $E[|h(T,C,X)| \,|X=x]<\infty$; hence, we can obtain

$$E[\delta Z(1-G(Z|X))^{-1}|X=x] = E[h(T,C,X)|X=x] =$$

$$= \int\left(\int h(u,v,x)dG(v|x)\right)dF^T(u|x) = \int u(1-G(u|x))^{-1}\left(\int I(u<v)dG(v|x)\right)dF^T(u|x).$$

Now, $\int I(u<v)dG(v|x) = 1-G(u|x)$. Hence, this expression implies that

$$E[\delta Z(1-G(Z|X))^{-1}|X=x] = \int udF^T(v|x) = x'\beta,$$

where the last equality holds by (1.1). ■

Proposition 7.- With the notation introduced in Section 2.4, if there are no ties among the observations of Z then (2.31) holds.

PROOF: Let us consider the ordered sample $\{(Z_{(i)},\delta_{(i)},X_{(i)})\}$, where the new order in the sample is such that $Z_{(1)}<...<Z_{(n)}$. If now $W_{(j)}$ denotes the size of the jump down of $\hat{H}_{n(j)}$ in $Z_{(j)}$ and, for $i<j$, $w_{(j)}(i)$ denotes the size of the jump up of $\hat{H}_{n(j)}$ in $Z_{(i)}$ ($\hat{H}_{n(j)}$ is defined as $\hat{H}_{nj}$ replacing (Z_j,δ_j,X_j) by $(Z_{(j)},\delta_{(j)},X_{(j)})$), then the summation in (2.30) may be rewritten as

$$\sum_j\{W_{(j)}(Z_{(j)}-b'X_{(j)})^2-\sum_{i=1}^{j-1}w_{(j)}(i)(Z_{(i)}-b'X_{(j)})^2\} =$$

$$= \sum_j\{W_{(j)}Z^2_{(j)}+W_{(j)}(b'X_{(j)})^2-2W_{(j)}Z_{(j)}b'X_{(j)}-\sum_{i=1}^{j-1}w_{(j)}(i)Z^2_{(i)}-$$

$$-\sum_{i=1}^{j-1}w_{(j)}(i)b'X_{(j)}+2\sum_{i=1}^{j-1}w_{(j)}(i)Z_{(i)}b'X_{(j)}\} \tag{A.4}$$

As $\hat{H}_{n(j)}(.)$ is a step-function whose value is 1 until its first discontinuity point and 0 in $[Z_{(j)},\infty)$ and their discontinuity points are contained in the set $\{Z_{(1)}, ..., Z_{(j)}\}$, then

$$W_{(j)} - \sum_{i=1}^{j-1}w_{(j)}(i) = 1. \tag{A.5}$$

Let us define $\tilde{W}_j(1) \equiv 1$ and, for $1<i\leq j$, $\tilde{W}_j(i) \equiv 1 + \sum_{k=1}^{i-1}w_{(j)}(k)$. (We will prove below that this expression coincides with (2.33)). Thus, $w_{(j)}(i) = \tilde{W}_j(i+1)-\tilde{W}_j(i)$, and, by (A.5), $\tilde{W}_j(j) = W_{(j)}$. Taking this into account, (A.4) may be transformed as follows:

$$\text{(A.4)} = \sum_j\{W_{(j)}Z^2_{(j)} + (b'X_{(j)})^2 - 2W_{(j)}Z_{(j)}b'X_{(j)} -$$

$$-\sum_{i=1}^{j-1}(\tilde{W}_j(i+1)-\tilde{W}_j(i))Z^2_{(i)}+2b'X_{(j)}\sum_{i=1}^{j-1}(\tilde{W}_j(i+1)-\tilde{W}_j(i))Z_{(i)}\} =$$

$$= \sum_j\{\sum_{i=1}^{j}\tilde{W}_j(i)Z^2_{(i)}-\sum_{i=1}^{j-1}\tilde{W}_j(i+1)Z^2_{(i)}+(b'X_{(j)})^2-$$

$$-2b'X_{(j)}[\sum_{i=1}^{j}\tilde{W}_j(i)Z_{(i)}-\sum_{i=1}^{j-1}\tilde{W}_j(i+1)Z_{(i)}]\}. \tag{A.6}$$

Now, if we denote $Z_{(0)} \equiv 0$ and rearrange terms, (A.6) becomes

$$\text{(A.6)} = \sum_j\{\sum_{i=1}^{j}\tilde{W}_j(i)(Z^2_{(i)}-Z^2_{(i-1)}) + (b'X_{(j)})^2-$$

$$-2b'X_{(j)}[\sum_{i=1}^{j}\tilde{W}_j(i)(Z_{(i)}-Z_{(i-1)})]\} =$$

$$= \sum_j\{\sum_{i=1}^{j}\tilde{W}_j(i)(Z^2_{(i)}-Z^2_{(i-1)})+(b'X_{(j)})^2-2b'X_{(j)}\hat{Z}^{SD}_{(j)}\} =$$

$$= \hat{P}_n + \hat{Q}_n(b),$$

where $\hat{Q}_n(b) \equiv \sum_j(\hat{Z}^{SD}_{(j)} - b'X_{(j)})^2$ and $\hat{P}_n \equiv \sum_j\{\sum_{i=1}^{j}\tilde{W}_j(i)(Z^2_{(i)} - Z^2_{(i-1)}) - (\hat{Z}^{SD}_{(j)})^2\}$. Obviously, as $\hat{P}_n$ does not depend on b,

$$\hat{\beta}^{SD} \equiv \underset{b}{argmin}\ (\hat{P}_n + \hat{Q}_n(b)) = \underset{b}{argmin}\ \sum_j(\hat{Z}^{SD}_{(j)} - b'X_{(j)})^2 \tag{A.7}$$

Now, (2.31) follows straightforwardly from (A.7). It only remains to prove that the expression for the weights $\tilde{W}_j(i)$ defined above coincides with the right-hand expression in (2.33). As $\tilde{W}_j(i) = 1 + \sum_{k=1}^{i-1} w_{(j)}(k)$, from the definition of $w_{(j)}(k)$ we deduce that

$$\tilde{W}_j(i) = \hat{H}_{n(j)}(Z_{(i-1)}) = (1-\hat{G}_{KC}(Z_{(i-1)}|X_{(j)}))^{-1}. \tag{A.8}$$

Now, (2.33) follows from (A.8) and (2.10). ■

REFERENCES

Aalen, O.O. (1978), "Nonparametric inference for a family of counting processes", *Annals of Statistics*, 6, 701-726.

Beran, R. (1981), "Nonparametric regression with randomly censored data", Technical Report, University of California, Berkeley.

Buckley, J. and James, I.R. (1979), "Linear regression with censored data", *Biometrika*, 66, 429-436.

Chatterjee, S. and McLeish, D.L. (1986), "Fitting linear regression models to censored data by least squares and maximum likelihood methods", *Communications in Statistics, A*, 15, 3227-3243.

Cox, D.R. (1972), "Regression models and life tables", *Journal of the Royal Statistical Society, B*, 33, 187-202.

Dabrowska, D. (1987), "Nonparametric regression with randomly censored survival data", *Scandinavian Journal of Statistics*, 17, 181-197.

Dabrowska, D. (1989), "Uniform consistency of the kernel conditional Kaplan-Meier estimate", *Annals of Statistics*, 17, 1157-1167.

Gonzalez-Manteiga, W. and Cadarso-Suarez, C. (1991), "Linear regression with randomly right-censored data using prior nonparametric estimation", in *Nonparametric Functional Estimation and Related Topics*, 315-328, G. Roussas (ed), Kluwer Academic Publishers.

Gonzalez-Manteiga, W. and Cadarso-Suarez, C. (1994), "Asymptotic properties of a generalized Kaplan-Meier estimator with some applications", *Nonparametric Statistics*, forthcoming.

Heckman, J.J. and Singer, B. (1984), "Econometric Duration Analysis", *Journal of Econometrics*, 24,63-132.

Heckman, J.J. and Walker, J. (1990), "The relationship between wages and income and the timing and spacing of births: evidence from Swedish longitudinal data", *Econometrica*, 58, 1411-1441.

Heller, G. and Simonoff, J.S. (1990), "A comparison of estimators for regression with a censored response variable", *Biometrika*, 77, 515-520.

James, I.R., and Smith, P.J. (1984), "Consistency results for linear regression with censored data", *Annals of Statistics*, 81, 1276-1288.

Kalbfleisch, J.D. and Prentice, R.L. (1980), *The Statistical Analysis of Failure Time Data*, Wiley, New York.

Kaplan, E.L. and Meier, P. (1958), "Nonparametric estimation from incomplete observations", *Journal of the American Statistical Association*, 53, 457-481.

Koul, H., Susarla, V. and Van Ryzin, J. (1981), "Regression analysis with randomly right-censored data", *Annals of Statistics*, 9, 1276-1288.

Lai, T.L. and Ying, Z. (1991), "Large sample theory of a modified Buckley-James estimator in regression analysis with censored data", *Annals of Statistics*, 19, 1370-1402.

Leurgans, S. (1987), "Linear models, random censoring and synthetic data", *Biometrika*, 74, 301-309.

Miller, R.G. (1976), "Least squares regression which censored data", *Biometrika*, 63, 449-464.

Miller, R.G. and Halpern, J. (1982), "Regression with censored data", *Biometrika*, 69, 521-531.

Nelson, W. (1972), "Theory and applications of hazard plotting for censored failure data", *Technometrics*, 14, 945-966.

Ritov, Y. (1990), "Estimation in a linear regression model with censored data", *Annals of Statistics*, 18, 303-328.

Susarla, V. and Van Ryzin, J. (1980), "Large sample theory for an estimator of the mean survival time from censored samples", *Annals of Statistics*, 8, 1002-1008.

TABLE 1

Monte Carlo Results for Model 1

		n=50					*n=400*				
		Intercept		Slope			Intercept		Slope		
Estim.		Bias	VAR	Bias	VAR	MSE[1]	Bias	VAR	Bias	VAR	MSE[1]
$\hat{\beta}^{BJ}_{A}$	h_1	-.151	.035	-.017	.019	776	-.076	.005	-.008	.003	144
	h_2	-.172	.036	-.050	.020	881	-.078	.005	-.021	.003	151
	h_3	-.180	.037	-.060	.020	925	-.113	.005	-.024	.003	213
$\hat{\beta}^{BJ}_{B}$	h_1	-.285	.027	-.144	.030	1597	-.142	.005	-.091	.004	374
	h_2	-.182	.035	-.363	.022	2219	-.123	.005	-.112	.004	367
	h_3	-.141	.046	-.501	.024	3414	-.156	.009	-.407	.021	2200
$\hat{\beta}^{BJ}_{C}$	h_1	-.187	.050	-.227	.062	1982	-.097	.007	-.122	.011	423
	h_2	.027	.076	-.356	.059	2622	.024	.007	-.157	.013	452
	h_3	.087	.096	-.424	.051	3342	-.053	.008	-.225	.012	734
$\hat{\beta}^{K}$	h_1	-.537	.077	-.123	.108	4884	-.322	.015	-.060	.016	1384
	h_2	-.320	.121	.414	.341	7352	-.263	.016	.344	.032	2354
	h_3	-.123	.212	.813	.768	--[2]	.154	.045	1.355	.261	--[2]
$\hat{\beta}^{SD}$	h_1	-.097	.026	.121	.028	785	-.049	.004	.068	.004	143
	h_2	.088	.041	.526	.071	3996	.010	.005	.314	.006	1127
	h_3	.264	.074	.857	.171	--[2]	.329	.012	.965	.034	--[2]
$\hat{\beta}^{MI}$	A	-.203	.046	-.002	.048	1344	-.073	.007	.000	.013	253
	B	-.238	.045	.010	.045	1470	-.085	.008	.025	.010	261
$\hat{\beta}^{LAM}$		-.408	.015	.000	.016	1981	-.398	.002	.005	.002	1625
$\hat{\beta}^{GC}$	h^*_1	-.471	.076	-.215	.091	4344	-.126	.011	-.057	.014	452
	h^*_2	-.324	.100	-.138	.201	4245	-.087	.013	.189	.018	754
	h^*_3	-.216	.130	.259	.329	5732	-.032	.019	.856	.204	9584

[1] $MSE \equiv 10000 \times (MSE(\hat{\beta}_1)+MSE(\hat{\beta}_2))$

[2] The corresponding value is greater than 9999

TABLE 2

Monte Carlo Results for Model 2

		n=50					*n=400*				
		Intercept		Slope			Intercept		Slope		
Estim.		Bias	VAR	Bias	VAR	MSE[1]	Bias	VAR	Bias	VAR	MSE[1]
$\hat{\beta}_A^{BJ}$	h_1	-.020	.020	-.006	.022	431	-.006	.002	.000	.002	46
	h_2	-.026	.021	-.012	.022	438	-.008	.002	-.003	.002	49
	h_3	-.029	.020	-.015	.023	441	-.009	.002	-.007	.003	58
$\hat{\beta}_B^{BJ}$	h_1	-.028	.021	-.033	.024	474	-.012	.002	-.021	.003	63
	h_2	-.003	.022	-.054	.023	475	.002	.003	-.035	.003	70
	h_3	.006	.022	-.076	.023	537	.016	.002	-.078	.002	111
$\hat{\beta}_C^{BJ}$	h_1	-.035	.024	-.044	.026	531	-.017	.003	-.031	.003	73
	h_2	.000	.025	-.055	.027	550	.001	.003	-.042	.003	78
	h_3	.009	.026	-.064	.028	584	.007	.003	-.051	.002	86
$\hat{\beta}^{K}$	h_1	-.137	.025	-.071	.039	883	-.055	.003	-.010	.003	93
	h_2	-.097	.026	.036	.053	902	-.050	.003	.076	.004	147
	h_3	-.066	.030	.099	.068	1118	.011	.003	.232	.008	654
$\hat{\beta}^{SD}$	h_1	-.010	.020	.009	.021	420	-.007	.002	.006	.002	47
	h_2	.011	.022	.052	.023	477	.000	.002	.030	.002	57
	h_3	.030	.024	.086	.026	579	.030	.003	.089	.003	144
$\hat{\beta}^{MI}$	A	-.027	.022	.015	.022	451	-.008	.003	.000	.003	52
	B	-.032	.021	.010	.023	451	-.008	.003	.001	.003	53
$\hat{\beta}^{LAM}$		-.046	.018	-.002	.020	404	-.045	.002	.000	.002	63
$\hat{\beta}^{GC}$	h_1^*	-.080	.025	-.092	.023	631	-.039	.003	-.035	.003	81
	h_2^*	-.047	.024	-.083	.027	598	-.002	.003	-.038	.003	74
	h_3^*	-.028	.026	-.159	.034	854	-.014	.003	-.142	.005	283

[1] $MSE \equiv 10000 \times (MSE(\hat{\beta}_1)+MSE(\hat{\beta}_2))$

TABLE 3

Monte Carlo Results for Model 3

		n=50					*n=400*				
		Intercept		Slope			Intercept		Slope		
Estim.		Bias	VAR	Bias	VAR	MSE[1]	Bias	VAR	Bias	VAR	MSE[1]
$\hat{\beta}^{BJ}_{A}$	h_1	-.036	.005	-.009	.005	120	-.018	.001	-.003	.001	18
	h_2	-.054	.005	-.017	.005	137	-.025	.001	-.008	.001	26
	h_3	-.058	.005	-.020	.005	141	-.024	.001	-.011	.001	27
$\hat{\beta}^{BJ}_{B}$	h_1	-.021	.006	-.067	.012	233	-.017	.001	-.025	.001	29
	h_2	.045	.008	-.165	.011	486	.028	.001	-.091	.001	110
	h_3	.072	.011	-.233	.017	874	.061	.002	-.204	.002	494
$\hat{\beta}^{BJ}_{C}$	h_1	-.008	.010	-.095	.022	418	-.007	.002	-.086	.002	114
	h_2	.086	.015	-.140	.027	685	.024	.002	-.088	.002	123
	h_3	.120	.019	-.174	.029	924	.073	.002	-.141	.002	292
$\hat{\beta}^{K}$	h_1	-.212	.020	.012	.040	1046	-.147	.002	.006	.003	266
	h_2	-.050	.026	.360	.099	2568	-.092	.002	.252	.005	784
	h_3	-.057	.042	.565	.179	5435	.146	.005	.334	.027	1583
$\hat{\beta}^{SD}$	h_1	-.004	.005	.041	.006	131	-.009	.001	.020	.001	18
	h_2	.092	.009	.188	.013	654	.035	.001	.101	.001	133
	h_3	.164	.016	.308	.027	1647	.170	.002	.319	.003	1364
$\hat{\beta}^{MI}$	A	-.034	.006	.005	.007	141	-.011	.001	.000	.001	19
	B	-.044	.006	-.007	.007	150	-.013	.001	-.002	.001	19
$\hat{\beta}^{LAM}$		-.076	.004	-.002	.005	142	-.076	.001	.000	.001	67
$\hat{\beta}^{GC}$	h^*_1	-.150	.015	-.090	.024	686	-.087	.002	-.041	.002	132
	h^*_2	-.057	.014	.087	.042	666	-.041	.002	-.038	.003	81
	h^*_3	-.031	.020	.081	.086	1116	-.017	.004	-.121	.009	279

[1] $MSE \equiv 10000 \times (MSE(\hat{\beta}_1) + MSE(\hat{\beta}_2))$

TABLE 4

Monte Carlo Results for Model 4

		n=50					n=400				
		Intercept		Slope			Intercept		Slope		
Estim.		Bias	VAR	Bias	VAR	MSE[1]	Bias	VAR	Bias	VAR	MSE[1]
$\hat{\beta}_A^{BJ}$	h_1	-.095	.093	-.014	.082	1836	-.043	.011	-.023	.010	230
	h_2	-.106	.093	-.043	.083	1886	-.052	.011	-.027	.010	244
	h_3	-.112	.093	-.055	.084	1921	-.056	.011	-.039	.010	256
$\hat{\beta}_B^{BJ}$	h_1	-.198	.085	-.071	.080	2089	-.097	.009	-.032	.010	294
	h_2	-.157	.084	-.173	.070	2087	-.091	.009	-.064	.010	313
	h_3	-.143	.088	-.241	.072	2382	-.090	.011	-.227	.011	817
$\hat{\beta}_C^{BJ}$	h_1	-.151	.108	-.105	.099	2407	-.084	.017	-.057	.016	433
	h_2	-.065	.112	-.155	.090	2305	-.044	.017	-.063	.016	389
	h_3	-.046	.116	-.180	.089	2394	-.041	.017	-.112	.015	462
$\hat{\beta}^{K}$	h_1	-.567	.091	-.306	.107	6140	-.291	.013	-.155	.014	1360
	h_2	-.516	.099	-.154	.151	5387	-.278	.013	-.015	.018	1085
	h_3	-.478	.109	-.058	.192	5336	-.210	.015	.265	.032	1606
$\hat{\beta}^{SD}$	h_1	-.059	.086	.050	.085	1774	-.029	.011	.029	.012	244
	h_2	.000	.095	.209	.104	2423	-.012	.011	.133	.012	415
	h_3	.051	.104	.319	.128	3361	.066	.013	.349	.017	1561
$\hat{\beta}^{MI}$	A	-.135	.097	-.002	.117	2324	-.053	.014	.013	.018	343
	B	-.161	.107	.059	.117	2539	-.060	.013	.060	.020	404
$\hat{\beta}^{LAM}$		-.310	.071	-.001	.069	2355	-.309	.008	.007	.009	1129
$\hat{\beta}^{GC}$	h_1^*	-.279	.117	-.135	.100	3121	-.121	.011	.041	.011	383
	h_2^*	-.190	.130	-.020	.132	2987	-.089	.012	.017	.010	302
	h_3^*	-.146	.138	-.025	.161	3209	-.081	.014	.122	.026	614

[1] $MSE \equiv 10000 \times (MSE(\hat{\beta}_1)+MSE(\hat{\beta}_2))$

TABLE 5

Monte Carlo Results for Model 5

		n=50					n=400				
		Intercept		Slope			Intercept		Slope		
Estim.		Bias	VAR	Bias	VAR	MSE[1]	Bias	VAR	Bias	VAR	MSE[1]
$\hat{\beta}^{BJ}_{A}$	h_1	-.081	.028	-.127	.026	769	-.035	.003	-.060	.003	112
	h_2	-.079	.028	-.129	.026	767	-.033	.003	-.060	.003	107
	h_3	-.079	.028	-.129	.026	765	-.033	.003	-.061	.003	106
$\hat{\beta}^{BJ}_{B}$	h_1	-.148	.030	-.193	.037	1260	-.068	.003	-.084	.004	187
	h_2	-.156	.029	-.269	.024	1501	-.071	.003	-.093	.003	194
	h_3	-.166	.030	-.318	.025	1839	-.072	.003	-.097	.003	203
$\hat{\beta}^{BJ}_{C}$	h_1	-.108	.040	-.196	.050	1402	-.041	.003	-.082	.006	159
	h_2	-.080	.039	-.225	.036	1327	-.038	.004	-.095	.006	158
	h_3	-.078	.041	-.254	.036	1479	-.034	.004	-.096	.005	193
$\hat{\beta}^{K}$	h_1	-.242	.056	-.293	.098	2982	-.099	.006	-.150	.010	482
	h_2	-.176	.058	-.233	.151	2945	-.066	.006	-.112	.016	378
	h_3	-.157	.062	-.216	.189	3216	-.051	.006	-.090	.026	429
$\hat{\beta}^{SD}$	h_1	-.041	.030	-.068	.035	710	-.019	.003	-.031	.004	84
	h_2	-.017	.031	-.034	.054	870	-.010	.004	-.015	.006	93
	h_3	-.011	.032	-.024	.067	998	-.006	.004	-.007	.009	124
$\hat{\beta}^{MI}$	A	-.062	.030	-.086	.041	827	-.026	.003	-.096	.004	172
	B	-.041	.032	-.088	.033	747	-.022	.003	-.091	.004	163
$\hat{\beta}^{LAM}$		-.284	.023	-.189	.026	1652	-.287	.003	-.179	.003	1199
$\hat{\beta}^{GC}$	h^*_1	-.192	.054	-.319	.071	2634	-.033	.005	-.151	.006	349
	h^*_2	-.130	.053	-.346	.091	2810	-.021	.005	-.140	.008	330
	h^*_3	-.108	.055	-.425	.090	3374	-.008	.005	-.162	.009	403

[1] $MSE \equiv 10000 \times (MSE(\hat{\beta}_1)+MSE(\hat{\beta}_2))$

TABLE 6

Monte Carlo Results for Model 6

		n=50					*n*=400				
		Intercept		Slope			Intercept		Slope		
Estim.		Bias	VAR	Bias	VAR	MSE[1]	Bias	VAR	Bias	VAR	MSE[1]
$\hat{\beta}^{BJ}_{A}$	h_1	-.166	.044	.035	.022	1188	-.136	.005	.021	.004	279
	h_2	-.193	.044	.025	.020	1239	-.154	.005	.019	.004	330
	h_3	-.206	.044	.024	.021	1289	-.158	.005	.019	.004	343
$\hat{\beta}^{BJ}_{B}$	h_1	-.309	.041	.200	.047	3106	-.151	.005	.111	.009	491
	h_2	-.094	.055	.370	.031	4034	-.078	.005	.142	.007	382
	h_3	-.030	.080	.511	.036	6756	-.041	.017	.410	.021	2078
$\hat{\beta}^{BJ}_{C}$	h_1	-.263	.073	.286	.073	4525	-.129	.005	.134	.013	526
	h_2	.172	.121	.356	.072	5492	.080	.006	.151	.013	482
	h_3	.272	.165	.420	.063	7244	.211	.009	.280	.011	1429
$\hat{\beta}^{K}$	h_1	-.441	.046	.235	.060	4703	-.270	.005	.114	.007	1174
	h_2	-.261	.063	.028	.110	3527	-.161	.005	.074	.010	622
	h_3	-.158	.085	.291	.165	6093	.079	.011	.416	.030	4234
$\hat{\beta}^{SD}$	h_1	-.160	.025	.082	.024	1114	-.113	.003	.044	.003	254
	h_2	.197	.056	.490	.061	7056	.043	.004	.274	.005	1656
	h_3	.505	.124	.866	.195	--[2]	.566	.016	.942	.032	--[2]
$\hat{\beta}^{MI}$	A	-.217	.051	.000	.047	1916	-.078	.008	.005	.011	363
	B	-.270	.046	.029	.052	2242	-.103	.008	.030	.012	442
$\hat{\beta}^{LAM}$		-.407	.017	.002	.016	2152	-.401	.002	.001	.002	1663
$\hat{\beta}^{GC}$	h^*_1	-.354	.077	.209	.080	4497	-.162	.004	.092	.006	591
	h^*_2	-.290	.081	.058	.129	4298	-.151	.004	.029	.007	425
	h^*_3	-.082	.215	.224	.530	9928	.040	.009	.406	.039	4183

[1] MSE $\equiv 10000 \times (MSE(\hat{\beta}_1)+MSE(\hat{\beta}_2)+MSE(\hat{\beta}_3))$

[2] The corresponding value is greater than 9999