Endogenous Sampling in Duration Models

Takeshi Amemiya

This paper considers the problem of endogenous sampling in the duration model. This is an important problem in the duration analysis of bank failures and loan defaults because it is common for the researchers in these areas to use only the default sample or non-default sample or both at a certain ratio, rather than using a random sample. The properties of endogenous sampling have been considered in various models, notably in qualitative response models, but not in duration models as far as I am aware. In this paper, I obtain the asymptotic distribution of the endogenous sampling maximum likelihood estimator and compare it with that of the random sampling maximum likelihood estimator and indicate when efficiency gain may result. I also show that the random sampling maximum likelihood estimator is inconsistent if the data are collected by endogenous sampling.

Key words: Duration models; Endogenous sampling; Bank failure; Loan default; Insolvency; Maximum likelihood estimator; Asymptotic distribution

Edward Ames Edmonds Professor of Economics, Department of Economics, Stanford University (E-mail: amemiya@leland.stanford.edu)

The first draft of the paper was written while I was visiting the Institute for Monetary and Economic Studies, Bank of Japan, and presented at the institute's seminar on July 13, 2000. I am indebted to Professor S. Moridaira, Professor Keunkwan Ryu, an anonymous referee, and the participants of the seminar for helpful comments.

I. Introduction

Endogenous sampling in the duration model occurs when the statistician uses only the default (non-right-censored) sample or only the non-default (right-censored) sample or both at a certain predetermined ratio. This is an important problem in the duration analysis of banks and loans because it is quite common for the researchers in these areas to use only the default sample or non-default sample or both at a certain ratio. For example, Lee and Urrutia (1996) use both kinds of data in an equal proportion in the analysis of insurer insolvencies. See other references cited there.

The properties of endogenous sampling have been considered in various models, most notably in qualitative response models (see Amemiya [1985]), but not in duration models as far as I am aware. Kim *et al.* (1995), in their study of insurer insolvencies, recognize the problem and cite Manski and Lerman (1977), who addressed the problem of endogenous sampling in the qualitative response model, but do not correctly deal with it. In fact, endogenous sampling in the duration model is so basically different from that in the qualitative response model that the results in one model cannot be readily applied to the other.

The order of the presentation is as follows: In Section II, I consider the asymptotic properties of the endogenous sampling maximum likelihood estimator (ESMLE) in the model where defaults and non-defaults are sampled in a certain proportion. I show that the random sample maximum likelihood estimator (RSMLE) is inconsistent under this scheme. Next, I compare ESMLE and RSMLE under their respective favorable conditions. A problem with ESMLE is its necessity to estimate a starting time distribution. In Section III, I propose a conditional ESMLE that alleviates this problem. In Section IV, I consider estimating the starting time distribution from a separate sample. In Sections V and VI, I consider ESMLE and Conditional ESMLE in the models with left censoring. Generalizations to the case of heterogeneous samples are given in Section VII.

II. Sampling Defaults and Non-Defaults in a Certain Proportion

A. Asymptotic Properties of ESMLE

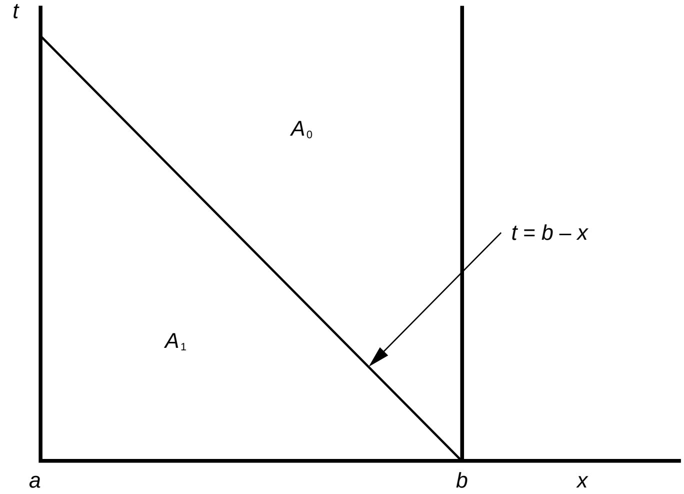
The duration data are generated according to the following scheme: a spell starts in an interval (a, b), and the starting time X is distributed according to density $h(x)$ and distribution function $H(x)$. The duration T of the spell is distributed according to density $f(t)$ and distribution function $F(t)$. We assume that X and T are independent. A spell is a default if it ends before b $(D = 1)$ and a non-default if it continues to b $(D = 0)$. Thus,

$$D = 1 \Leftrightarrow t < b - x \equiv A_1$$

$$D = 0 \Leftrightarrow t \geq b - x \equiv A_0.$$

The above is diagrammatically represented in Figure 1.

Figure 1 Partition of the *x*-*t* Plane



We assume that the statistician samples defaults with probability λ_1 and non-defaults with probability λ_0 (= 1 – λ_1). In order to write the likelihood function of the model, we first note

$$f(x, t \mid D = 1) = h(x)f(t)/P_1,$$

where $P_1 = P(T < b - X) = \int_a^b F(b-x)h(x)dx$ and

$$f(x \mid D = 0) = h(x)[1 - F(b-x)]/P_0,$$

where $P_0 = \int_a^b h(x)[1 - F(b-x)]dx.$

Therefore, the likelihood function is

$$L = P_1^{-N_1}\prod_1 h(x_i)f(t_i)\cdot\lambda_1^{N_1}\cdot P_0^{-N_0}\prod_0 h(x_i)[1 - F(b-x_i)]\cdot\lambda_0^{N_0}, \qquad (1)$$

where $\prod_1$ and $\prod_0$ mean taking the product over the default and non-default samples, respectively, and N_1 and N_0 are the numbers of the default and non-default samples. Note that N_1 is a random variable distributed as binomial (N, λ_1). Assume that the parameter β characterizes f but not h. (For simplicity of the notation, I will assume that β is a scalar, but all the subsequent formulas can be easily generalized to the vector case.) Ignoring the terms that do not depend on β, we have

$$\log L = N_1 \log P_1 + \sum_1 \log f(t_i) - N_0 \log P_0 + \sum_0 \log[1 - F(b - x_i)]. \tag{2}$$

To show the consistency of the ESMLE of β, consider

$$\frac{1}{N}\frac{\partial \log L}{\partial \beta} = -\frac{N_1}{N}\frac{1}{P_1}\frac{\partial P_1}{\partial \beta} + \frac{1}{N}\sum D \frac{1}{f}\frac{\partial f}{\partial \beta}$$

$$-\frac{N_0}{N}\frac{1}{P_0}\frac{\partial P_0}{\partial \beta} + \frac{1}{N}\sum (1-D)\frac{1}{1-F}\frac{\partial (1-F)}{\partial \beta}, \tag{3}$$

where $\sum$ means the summation over the whole sample and D, f, and F depend implicitly on i. The consistency of ESMLE is essentially equivalent to the condition that the expression in equation (3) converges to zero in probability. In order to verify this condition, note

$$\text{plim}\frac{1}{N}\sum D \frac{1}{f}\frac{\partial f}{\partial \beta} = ED\frac{1}{f}\frac{\partial f}{\partial \beta} = \lambda_1 E\left(\frac{1}{f}\frac{\partial f}{\partial \beta}\,\Big|\, D = 1\right)$$

$$= \frac{\lambda_1}{P_1}\int_{A_1}\frac{1}{f}\frac{\partial f}{\partial \beta} hfdtdx = \frac{\lambda_1}{P_1}\frac{\partial P_1}{\partial \beta}, \tag{4}$$

$$E(1-D)\frac{1}{1-F}\frac{\partial (1-F)}{\partial \beta} = \lambda_0 E\left(\frac{1}{1-F}\frac{\partial (1-F)}{\partial \beta}\,\Big|\, D = 0\right)$$

$$= \frac{\lambda_0}{P_0}\int_a^b \frac{1}{1-F}\frac{\partial (1-F)}{\partial \beta} h(1-F)dx$$

$$= \frac{\lambda_0}{P_0}\frac{\partial P_0}{\partial \beta}. \tag{5}$$

Thus, the consistency follows from equations (3), (4), and (5). From the above results, we see that both ESMLE using only the default sample (λ_1 = 1) and the ESMLE using only the non-default sample (λ_0 = 1) are consistent.

Next, we will derive the asymptotic variance using the well-known formula (see Amemiya [1985, p. 121]):

$$AV[\sqrt{N}(\hat{\beta} - \beta)] = \left[E\frac{1}{N}\left(\frac{\partial \log L}{\partial \beta}\right)^2\right]^{-1} = \left[-E\frac{1}{N}\frac{\partial^2 \log L}{\partial \beta^2}\right]^{-1}. \tag{6}$$

Rearranging the terms of the right-hand side (RHS) of equation (3) and multiplying them by $\sqrt{N}$,

$$\frac{1}{\sqrt{N}}\frac{\partial \log L}{\partial \beta} = \frac{1}{\sqrt{N}}\sum D\left(\frac{1}{f}\frac{\partial f}{\partial \beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial \beta}\right)$$

$$+ \frac{1}{\sqrt{N}}\sum (1-D)\left(\frac{1}{1-F}\frac{\partial (1-F)}{\partial \beta} - \frac{1}{P_0}\frac{\partial P_0}{\partial \beta}\right). \qquad (7)$$

Since

$$E\left[\left(\frac{1}{f}\frac{\partial f}{\partial \beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial \beta}\right)^2 \middle| D = 1\right] = \frac{1}{P_1}\int_{A_1}\left(\frac{1}{f}\frac{\partial f}{\partial \beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial \beta}\right)^2 hfdtdx$$

$$= \frac{1}{P_1}\int_{A_1}\left[\frac{1}{f^2}\left(\frac{\partial f}{\partial \beta}\right)^2 + \frac{1}{P_1^2}\left(\frac{\partial P_1}{\partial \beta}\right)^2 - \frac{2}{f}\frac{\partial f}{\partial \beta}\frac{1}{P_1}\frac{\partial P_1}{\partial \beta}\right]hfdtdx$$

$$= \frac{1}{P_1}\left[\int_{A_1}\frac{h}{f}\left(\frac{\partial f}{\partial \beta}\right)^2 dtdx - \frac{1}{P_1}\left(\frac{\partial P_1}{\partial \beta}\right)^2\right], \qquad (8)$$

we have

$$ED\left(\frac{1}{f}\frac{\partial f}{\partial \beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial \beta}\right)^2 = \frac{\lambda_1}{P_1}\left[\int_{A_1}\frac{h}{f}\left(\frac{\partial f}{\partial \beta}\right)^2 dtdx - \frac{1}{P_1}\left(\frac{\partial P_1}{\partial \beta}\right)^2\right]. \qquad (9)$$

Since

$$E\left[\left(\frac{1}{1-F}\frac{\partial (1-F)}{\partial \beta} - \frac{1}{P_0}\frac{\partial P_0}{\partial \beta}\right)^2 \middle| D = 0\right]$$

$$= \frac{1}{P_0}\int_a^b\left(\frac{1}{1-F}\frac{\partial (1-F)}{\partial \beta} - \frac{1}{P_0}\frac{\partial P_0}{\partial \beta}\right)^2 h(1-F)dx$$

$$= \frac{1}{P_0}\left[\int_a^b\frac{h}{1-F}\left(\frac{\partial (1-F)}{\partial \beta}\right)^2 dx - \frac{1}{P_0}\left(\frac{\partial P_0}{\partial \beta}\right)^2\right], \qquad (10)$$

we have

$$E(1-D)\left(\frac{1}{1-F}\frac{\partial (1-F)}{\partial \beta} - \frac{1}{P_0}\frac{\partial P_0}{\partial \beta}\right)^2$$

$$= \frac{\lambda_0}{P_0}\left[\int_a^b\frac{h}{1-F}\left(\frac{\partial (1-F)}{\partial \beta}\right)^2 dx - \frac{1}{P_0}\left(\frac{\partial P_0}{\partial \beta}\right)^2\right]. \qquad (11)$$

Therefore, from equations (6), (7), (9), and (11),

$$AV(ESMLE)^{-1} = E\frac{1}{N}\left(\frac{\partial \log L}{\partial \beta}\right)^2$$

$$= \frac{\lambda_1}{P_1}\left[\int_{A_1}\frac{h}{f}\left(\frac{\partial f}{\partial \beta}\right)^2 dtdx - \frac{1}{P_1}\left(\frac{\partial P_1}{\partial \beta}\right)^2\right]$$

$$+ \frac{\lambda_0}{P_0}\left[\int_a^b \frac{h}{1-F}\left(\frac{\partial(1-F)}{\partial \beta}\right)^2 dx - \frac{1}{P_0}\left(\frac{\partial P_0}{\partial \beta}\right)^2\right]. \tag{12}$$

We can also verify the second equality of equation (6).

B. Inconsistency of RSMLE

We will show that RSMLE is inconsistent under the endogenous sampling scheme described in the beginning of Section II.A. The likelihood function to be maximized to obtain RSMLE is

$$L_R = \prod_1 h(x_i)f(t_i)\cdot\prod_0 h(x_i)[1-F(b-x_i)]. \tag{13}$$

Therefore, we have

$$\log L_R = \sum D_i \log f(t_i) + \sum(1-D_i)\log[1-F(b-x_i)], \tag{14}$$

$$\frac{1}{N}\frac{\partial \log L_R}{\partial \beta} = \frac{1}{N}\sum D\frac{1}{f}\frac{\partial f}{\partial \beta} + \frac{1}{N}\sum(1-D)\frac{1}{1-F}\frac{\partial(1-F)}{\partial \beta}, \tag{15}$$

$$E\frac{1}{N}\frac{\partial \log L_R}{\partial \beta} = \frac{\lambda_1}{P_1}\frac{\partial P_1}{\partial \beta} + \frac{\lambda_0}{P_0}\frac{\partial P_0}{\partial \beta}. \tag{16}$$

The inconsistency follows from the fact that the RHS of equation (16) is not zero unless $\lambda_1 = P_1$ (hence $\lambda_0 = P_0$).

We will evaluate the degree of the inconsistency of RSMLE in a simple example. For this purpose, we must treat the β that appears on the RHS of equation (15) as the domain of the function and take the expectation using the true value β^*. Note that in equation (16) I was implicitly evaluating the function at the true value without defining a new symbol. Then, we have, instead of equation (16),

$$\frac{\lambda_1}{P_1^*}\int_{A_1}\frac{1}{f}\frac{\partial f}{\partial \beta}hf^*dtdx + \frac{\lambda_0}{P_0^*}\int_a^b\frac{1}{1-F}\frac{\partial(1-F)}{\partial \beta}h(1-F^*)dx, \tag{16*}$$

where the functions with * are evaluated at β^*. Note that equation (16*) is reduced to equation (16) when we remove * from the RHS. Given the true value β^*, the probability limit of RSMLE is given by solving for β the equation obtained by equating equation (16*) to zero. The simple example we will consider is defined by $a = 0$, $b = 1$, $h(x) = U(0, 1)$, $f(t) = \beta\exp(-\beta t)$, and $\lambda_0 = \lambda_1 = 0.5$. We will assume $\beta^* = 1$. Then we can calculate the probability limit of RSMLE to be 0.468.

C. Comparison of ESMLE and RSMLE

We will now compare the asymptotic variances of RSMLE and RSMLE derived under their respective correct models. We have already done so for ESMLE in Section II.A, so we now do the same for RSMLE. Note that equations (17), (18), and (19) below are analogous to equations (7), (9), and (11). From equation (15), we obtain

$$\frac{1}{\sqrt{N}}\frac{\partial \log L_R}{\partial\beta} = \frac{1}{\sqrt{N}}\sum\left(D\frac{1}{f}\frac{\partial f}{\partial\beta} - \frac{\partial P_1}{\partial\beta}\right) + \frac{1}{\sqrt{N}}\sum\left((1-D)\frac{1}{1-F}\frac{\partial(1-F)}{\partial\beta} - \frac{\partial P_0}{\partial\beta}\right). \tag{17}$$

Note a slight difference between equations (17) and (7). Analogous to equation (9), we have

$$E\left(D\frac{1}{f}\frac{\partial f}{\partial\beta} - \frac{\partial P_1}{\partial\beta}\right)^2 = \int_{A_1}\frac{h}{f}\left(\frac{\partial f}{\partial\beta}\right)^2 dtdx - \left(\frac{\partial P_1}{\partial\beta}\right)^2, \tag{18}$$

and analogous to equation (11) we have

$$E\left[(1-D)\frac{1}{1-F}\frac{\partial(1-F)}{\partial\beta} - \frac{1}{P_0}\frac{\partial P_0}{\partial\beta}\right]^2 = \int_a^b\frac{h}{1-F}\left(\frac{\partial(1-F)}{\partial\beta}\right)^2 dx - \left(\frac{\partial P_0}{\partial\beta}\right)^2. \tag{19}$$

Unlike the derivation in the case of ESMLE, however, here we need to calculate the expectation of the cross product:

$$E\left(D\frac{1}{f}\frac{\partial f}{\partial\beta} - \frac{\partial P_1}{\partial\beta}\right)\left((1-D)\frac{1}{1-F}\frac{\partial(1-F)}{\partial\beta} - \frac{1}{P_0}\frac{\partial P_0}{\partial\beta}\right) = \left(\frac{\partial P_1}{\partial\beta}\right)^2. \tag{20}$$

Therefore, from equations (17) through (20) we obtain

$$AV(RSMLE)^{-1} = E\frac{1}{N}\left(\frac{\partial \log L_R}{\partial\beta}\right)^2$$

$$= \int_{A_1}\frac{h}{f}\left(\frac{\partial f}{\partial\beta}\right)^2 dtdx + \int_a^b \frac{h}{1-F}\left(\frac{\partial(1-F)}{\partial\beta}\right)^2 dx. \qquad (21)$$

It is interesting to note that if we put $\lambda_1 = P_1$ and $\lambda_0 = P_0$ in equation (12), we do not get equation (21). In fact, what we obtain by putting $\lambda_1 = P_1$ and $\lambda_0 = P_0$ in equation (12) is smaller than equation (21). For some values of λ_1 and λ_0, however, equation (12) may be larger than equation (21), allowing for the possibility that ESMLE may be more efficient than RSMLE. From equation (12), it is clear that the RHS of equation (12) is maximized either at $\lambda_1 = 1$ or $\lambda_0 = 1$ depending on which of the coefficients on p_1 and p_0 is greater. Thus, contrary to intuition, the optimum does not occur in between.

To get a concrete idea about the difference in asymptotic efficiency between ESMLE and RSMLE, we will evaluate their asymptotic variances in the same simple example we considered at the end of the preceding section: namely, $a = 0$, $b = 1$, $h(x) = U(0, 1)$, and $f(t) = \beta\exp(-\beta t)$. Define ESMLE1 to be the estimator using only the default sample and ESMLE0 using only the non-default sample. Their asymptotic variances are given by equations (8) and (10). Then, inserting the values specified by the simple example into equations (8) and (10), we obtain

$$AV(ESMLE1)^{-1} = \frac{\beta - 3 + (3 + 2\beta + \beta^2)e^{-\beta}}{\beta^3 - \beta^2 + \beta^2 e^{-\beta}} - \frac{(1 - e^{-\beta} - \beta e^{-\beta})^2}{(\beta^2 - \beta + \beta e^{-\beta})^2}, \qquad (22)$$

$$AV(ESMLE0)^{-1} = \frac{2 - (\beta^2 + 2\beta + 2)e^{-\beta}}{\beta^2 - \beta^2 e^{-\beta}} - \frac{(1 - e^{-\beta} - \beta e^{-\beta})^2}{(\beta - \beta e^{-\beta})^2}, \qquad (23)$$

$$AV(RSMLE)^{-1} = \frac{1}{\beta^2} - \frac{1}{\beta^3} + \frac{e^{-\beta}}{\beta^3}. \qquad (24)$$

In Table 1, we have evaluated these three inverses of the asymptotic variances for some values of β.

Table 1 Asymptotic Variances of Three Estimators

β	0.5	1	5
$AV(RSMLE)^{-1}$	0.852	0.368	0.032
$AV(ESMLE1)^{-1}$	0.052	0.048	0.020
$AV(ESMLE0)^{-1}$	0.082	0.079	0.033

III. Conditional ESMLE Using Defaults

ESMLE using only the default sample maximizes

$$L_1 = P_1^{-N_1}\prod_1 h(x_i)f(t_i), \tag{25}$$

and its asymptotic variance is given by equation (8). A problem with this estimator is the fact that P_1 depends on $h(x)$ and hence $h(x)$ cannot be ignored even if one wanted to estimate only the parameter β that characterizes $f(t)$. Conditional ESMLE (CESMLE) alleviates this difficulty. This estimator is analogous to the conditional maximum likelihood estimator used in the duration model with left censoring (see Amemiya [1999]).

The conditional density of t given x in A_1 is given by

$$f(t|x) = \frac{h(x)f(t)}{\int_0^{b-x} h(x)f(t)dt} = \frac{f(t)}{F(b-x)}. \tag{26}$$

Therefore, CESMLE maximizes

$$L_C = \Pi \frac{f(t_i)}{F(b-x_i)}, \tag{27}$$

or, equivalently,

$$\log L_C = \sum \log f - \sum \log F. \tag{28}$$

Consider

$$\frac{1}{N}\frac{\partial \log L_C}{\partial \beta} = \frac{1}{N}\sum \frac{1}{f}\frac{\partial f}{\partial \beta} - \frac{1}{N}\sum \frac{1}{F}\frac{\partial F}{\partial \beta}. \tag{29}$$

Taking the probability limit,

$$\begin{aligned} \text{plim}\frac{1}{N}\frac{\partial \log L_C}{\partial \beta} &= E\left(\frac{1}{f}\frac{\partial f}{\partial \beta}\Big| D=1\right) - E\left(\frac{1}{F}\frac{\partial F}{\partial \beta}\Big| D=1\right) \\ &= \frac{1}{P_1}\frac{\partial P_1}{\partial \beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial \beta} = 0. \end{aligned} \tag{30}$$

Therefore, CESMLE is consistent.

To evaluate the asymptotic variance, consider

$$\frac{1}{\sqrt{N}}\frac{\partial \log L_C}{\partial\beta} = \frac{1}{\sqrt{N}}\sum\left(\frac{1}{f}\frac{\partial f}{\partial\beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial\beta}\right)$$

$$- \frac{1}{\sqrt{N}}\sum\left(\frac{1}{F}\frac{\partial F}{\partial\beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial\beta}\right). \qquad (31)$$

We need to evaluate the mean of the square of each term and the cross product. The mean of the square of the first term has been derived in equation (8). We have

$$E\left[\left(\frac{1}{f}\frac{\partial f}{\partial\beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial\beta}\right)^2 \middle| D = 1\right] = \frac{1}{P_1}\left[\int_a^b \frac{h}{F}\left(\frac{\partial F}{\partial\beta}\right)^2 dx - \frac{1}{P_1}\left(\frac{\partial P_1}{\partial\beta}\right)^2\right] \qquad (32)$$

and

$$E\left[\left(\frac{1}{f}\frac{\partial f}{\partial\beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial\beta}\right)\left(\frac{1}{F}\frac{\partial F}{\partial\beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial\beta}\right) \middle| D = 1\right]$$

$$= \frac{1}{P_1}\left[\int_a^b \frac{h}{F}\left(\frac{\partial F}{\partial\beta}\right)^2 dx - \frac{1}{P_1}\left(\frac{\partial P_1}{\partial\beta}\right)^2\right]. \qquad (33)$$

Therefore,

$$AV(CESMLE)^{-1} = E\left(\frac{1}{\sqrt{N}}\frac{\partial \log L_C}{\partial\beta}\right)^2$$

$$= \frac{1}{P_1}\left[\int_{A_1} \frac{h}{f}\left(\frac{\partial f}{\partial\beta}\right)^2 dtdx - \int_a^b \frac{h}{F}\left(\frac{\partial F}{\partial\beta}\right)^2 dx\right]. \qquad (34)$$

The above can be shown to be equal to

$$-E\frac{1}{N}\frac{\partial^2 \log L_C}{\partial\beta^2}.$$

We will show that equation (34) is smaller than the inverse of the asymptotic variance of ESMLE using only the default sample, namely, what we obtain by putting $p_1 = 1$ in equation (12). For this, we need to verify

$$\int_{A_1}\frac{h}{f}\left(\frac{\partial f}{\partial\beta}\right)^2 dtdx-\frac{1}{P_1}\left(\frac{\partial P_1}{\partial\beta}\right)^2\geq\int_{A_1}\frac{h}{f}\left(\frac{\partial f}{\partial\beta}\right)^2 dtdx-\int_a^b\frac{h}{F}\left(\frac{\partial F}{\partial\beta}\right)^2 dx, \qquad (35)$$

or equivalently

$$P_1\int_a^b\frac{h}{F}\left(\frac{\partial F}{\partial\beta}\right)^2 dx\geq\left(\frac{\partial P_1}{\partial\beta}\right)^2, \qquad (36)$$

or equivalently

$$\int_a^b Fhdx\cdot\int_a^b\frac{1}{F}\left(\frac{\partial F}{\partial\beta}\right)^2 hdx\geq\left(\int_a^b\frac{\partial F}{\partial\beta}hdx\right)^2, \qquad (37)$$

which follows from the Cauchy-Schwartz inequality

$$EV^2\cdot EU^2\geq(EUV)^2, \qquad (38)$$

if we put

$$U=\frac{1}{\sqrt{F}}\frac{\partial F}{\partial\beta},\ V=\sqrt{F},$$

and the expectation is taken with respect to x.

ESMLE using only the non-default sample maximizes

$$L_0=P_0^{-N_0}\prod_0 h(x_i)[1-F(b-x_i)]. \qquad (39)$$

Thus, each density has the form

$$\frac{h(x)[1-F(b-x)]}{\int_a^b h(x)[1-F(b-x)]dx}. \qquad (40)$$

Note that equation (40) depends only on x. Therefore, there is no CESMLE using only the non-default sample.

IV. Separate Estimation of $h(x)$

We now consider the case where we can estimate the density $h(x)$ or the distribution function $H(x)$ using an augmented sample independent of that used to estimate β.

We maximize equation (25) after estimating h from a separate independent sample. That is, maximize

$$W = \Pi f(t_i)\hat{P}_1^{-1}, \tag{41}$$

where

$$\hat{P}_1 = \int_a^b F(b-x)d\hat{H}(x) \tag{42}$$

and $\hat{H}$ is the empirical distribution function based on K separate observations. Thus,

$$\hat{P}_1 = K^{-1}\sum_{k=1}^{K} F(b-x_k). \tag{43}$$

If we denote this estimator by $\tilde{\beta}_1$, its asymptotic distribution can be obtained from

$$\sqrt{N}(\tilde{\beta}_1 - \beta) = -\left(\frac{1}{\sqrt{N}}\frac{\partial \log W}{\partial \beta}\right)\left(\frac{1}{N}\frac{\partial^2 \log W}{\partial \beta^2}\right)^{-1}. \tag{44}$$

The second-derivative term above divided by N will converge to the same limit as if H were not estimated. So, here, we will consider only the first derivative part. We have, ignoring the terms that do not depend on β,

$$\log W = \sum_{i=1}^{N} \log f(t_i) - N\log\sum_{k=1}^{K} F(b-x_k). \tag{45}$$

Since

$$\frac{\partial \log W}{\partial \beta} = \sum_{i=1}^{N}\frac{1}{f}\frac{\partial f}{\partial \beta} - \frac{N}{\sum_{k=1}^{K} F}\sum_{k=1}^{K}\frac{\partial F}{\partial \beta}, \tag{46}$$

$$\frac{1}{\sqrt{N}}\frac{\partial \log W}{\partial \beta} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{f}\frac{\partial f}{\partial \beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial \beta}\right)$$

$$-\sqrt{N}\left[\frac{\frac{1}{K}\sum_{k=1}^{K}\frac{\partial F}{\partial \beta} - \frac{\partial P_1}{\partial \beta}}{\frac{1}{K}\sum_{k=1}^{K}F} - \frac{\frac{\partial P_1}{\partial \beta}\left(\frac{1}{K}\sum_{k=1}^{K}F - P_1\right)}{P_1\frac{1}{K}\sum_{k=1}^{K}F}\right]$$

$$\underset{=}{LD}\,\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{f}\frac{\partial f}{\partial \beta} - \frac{1}{P_1}\frac{\partial P_1}{\partial \beta}\right)$$

$$-\sqrt{N}\left[\frac{\frac{1}{K}\sum_{k=1}^{K}\frac{\partial F}{\partial \beta} - \frac{\partial P_1}{\partial \beta}}{P_1} - \frac{\frac{\partial P_1}{\partial \beta}\left(\frac{1}{K}\sum_{k=1}^{K}F - P_1\right)}{P_1^2}\right]. \qquad (47)$$

Note that the first term after $\underset{=}{LD}$ above is

$$\frac{1}{\sqrt{N}}\frac{\partial \log L_1}{\partial \beta}$$

in the case of using only the default sample, as can be seen from the first term on the RHS of equation (7), and the second term arises from estimating H. Therefore, if we define

$$B = E\left[\frac{1}{\sqrt{N}}\frac{\partial \log L_1}{\partial \beta}\right]^2 = \frac{1}{P_1}\left[\int_{A_1}\frac{h}{f}\left(\frac{\partial f}{\partial \beta}\right)^2 dtdx - \frac{1}{P_1}\left(\frac{\partial P_1}{\partial \beta}\right)^2\right], \qquad (48)$$

the asymptotic variance of $\sqrt{N}(\tilde{\beta}_1 - \beta)$ is given by

$$AV\left[\sqrt{N}(\tilde{\beta}_1 - \beta)\right] =$$

$$B^{-1}\left[B + \frac{N}{K}\frac{1}{P_1^2}V\frac{\partial F}{\partial \beta} + \frac{N}{K}\frac{1}{P_1^4}\left(\frac{\partial P_1}{\partial \beta}\right)^2 VF - 2\frac{N}{K}\frac{1}{P_1^3}\frac{\partial P_1}{\partial \beta}\text{cov}\left(F, \frac{\partial F}{\partial \beta}\right)\right]B^{-1}. \quad (49)$$

Thus, if $N/K \to 0$, the estimator is as efficient as if H were known. Otherwise, K must go to infinity at least as fast as N in order for the above to remain finite.

If we estimate the density h by a kernel estimator of the form

$$\hat{h}(x) = \sum_{i=1}^{K} g\left(\frac{x_i - x}{d}\right)\frac{1}{Kd}, \qquad (50)$$

we can get the same asymptotic result as above provided that the kernel function *g* and the rate of convergence of *d* to zero satisfy certain conditions. But the proof is more involved in this case. See, for example, Ait-Sahalia (1994).

We will now obtain an analogous result for the case of ESMLE using only the non-default sample. Here we maximize

$$W = \Pi[1 - F(b - x_i)]\hat{P}_0^{-1}, \tag{51}$$

where

$$\hat{P} = \frac{1}{K}\sum_{k=1}^{K}[1 - F(b - x_k)]. \tag{52}$$

We have, ignoring the terms that do not depend on β,

$$\log W = \sum_{i=1}^{N}\log[1 - F(b - x_i)] - N\log\sum_{k=1}^{K}[1 - F(b - x_k)]. \tag{53}$$

Since

$$\frac{\partial \log W}{\partial \beta} = \sum_{i=1}^{N}\frac{1}{1-F}\frac{\partial(1-F)}{\partial\beta} - \frac{N}{\sum_{k=1}^{K}(1-F)}\sum_{k=1}^{K}\frac{\partial(1-F)}{\partial\beta}, \tag{54}$$

$$\frac{1}{\sqrt{N}}\frac{\partial \log W}{\partial \beta} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{1-F}\frac{\partial(1-F)}{\partial\beta} - \frac{1}{P_0}\frac{\partial P_0}{\partial\beta}\right)$$

$$-\sqrt{N}\left[\frac{\frac{1}{K}\sum_{k=1}^{K}\frac{\partial(1-F)}{\partial\beta} - \frac{\partial P_0}{\partial\beta}}{\frac{1}{K}\sum_{k=1}^{K}(1-F)} - \frac{\frac{\partial P_0}{\partial\beta}\left(\frac{1}{K}\sum_{k=1}^{K}(1-F) - P_0\right)}{P_0\frac{1}{K}\sum_{k=1}^{K}(1-F)}\right]$$

$$\underset{=}{LD} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{(1-F)}\frac{\partial(1-F)}{\partial\beta} - \frac{1}{P_0}\frac{\partial P_0}{\partial\beta}\right)$$

$$-\sqrt{N}\left[\frac{\frac{1}{K}\sum_{k=1}^{K}\frac{\partial(1-F)}{\partial\beta} - \frac{\partial P_0}{\partial\beta}}{P_0} - \frac{\frac{\partial P_0}{\partial\beta}\left(\frac{1}{K}\sum_{k=1}^{K}(1-F) - P_0\right)}{P_0^2}\right]. \tag{55}$$

Note that the first term after $\underset{=}{LD}$ above is

$$\frac{1}{\sqrt{N}}\frac{\partial \log L_0}{\partial \beta}$$

in the case of using only the default sample, as can be seen from the second term on the RHS of equation (7), and the second term arises from estimating H. Therefore, if we define

$$C = E\left[\frac{1}{\sqrt{N}}\frac{\partial \log L_0}{\partial \beta}\right]^2 = \frac{1}{P_0}\left[\int_a^b \frac{h}{1-F}\left(\frac{\partial(1-F)}{\partial \beta}\right)^2 dx - \frac{1}{P_0}\left(\frac{\partial P_0}{\partial \beta}\right)^2\right], \qquad (56)$$

the asymptotic variance of $\sqrt{N}(\tilde{\beta}_0 - \beta)$ is given by

$$AV[\sqrt{N}(\tilde{\beta}_0 - \beta)] =$$

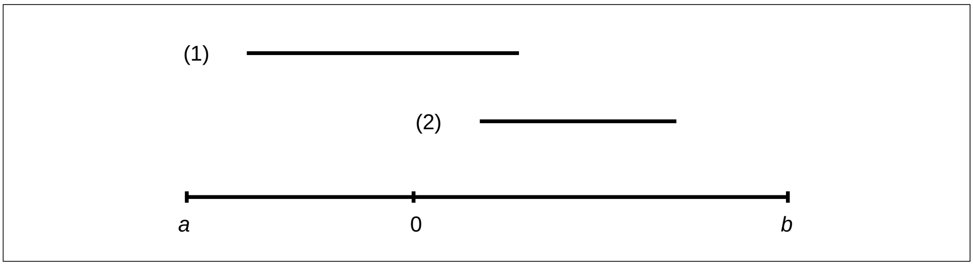
$$C^{-1}\left[C + \frac{N}{K}\frac{1}{P_0^2}V\frac{\partial F}{\partial \beta} + \frac{N}{K}\frac{1}{P_0^4}\left(\frac{\partial P_0}{\partial \beta}\right)^2 VF - 2\frac{N}{K}\frac{1}{P_0^3}\frac{\partial P_0}{\partial \beta}\operatorname{cov}\left(F, \frac{\partial F}{\partial \beta}\right)\right]C^{-1}. \qquad (57)$$

V. ESMLE with Left Censoring

At the beginning of Section II.A, we defined the range of the starting time x of a spell as (a, b), where a is a certain time in the past and b is the present time. Now we consider a time within this interval and denote it as zero. The assumption of the present section is that we sample only those spells which either are continuing at time zero or start after zero. (This is the problem of left censoring studied by Amemiya [1999].) Moreover, we sample only defaults, that is, only those spells which end before b. (This is the problem of endogenous sampling.) In this section, we consider a simultaneous occurrence of left censoring and endogenous sampling.

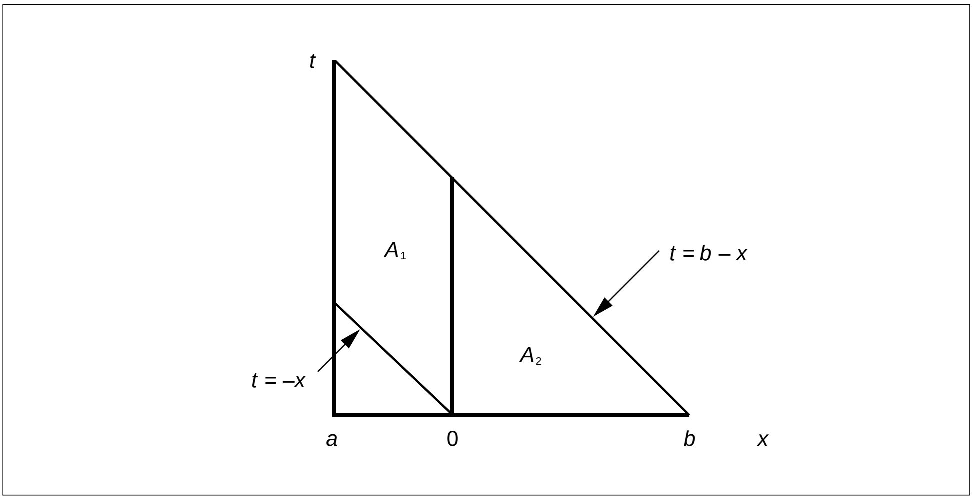
Consider the following two types of spells: (1) those that start in $(a, 0)$ and continue to zero but end before b and (2) those that start in $(0, b)$ and end before b. The two types of spells are described in Figure 2.

Figure 2 Two Types of Spells



In Figure 3, the domains of the two types of spells are described on the x-t plane.

Figure 3 Domains of the Two Types of Spells



We have

$$P_1 \equiv P(A_1) = \int_a^0 \int_{-x}^{b-x} fhdtdx = \int_a^0 [F(b-x) - F(-x)]hdx, \quad (58)$$

$$P_2 \equiv P(A_2) = \int_0^b \int_0^{b-x} fhdtdx = \int_0^b F(b-x)hdx. \quad (59)$$

The question that we now wish to address is: should we divide Type (1) sample by P_1 and Type (2) sample by P_2, or all the samples by $P \equiv P_1 + P_2$?

If we divide all the samples by P, we maximize

$$L = P^{-N} \prod_1 h(x_i) f(t_i) \prod_2 h(x_i) f(t_i). \quad (60)$$

Ignoring the terms that do not depend on β, we have

$$\log L = -N \log P + \sum \log f, \quad (61)$$

$$\frac{1}{\sqrt{N}} \frac{\partial \log L}{\partial \beta} = \frac{1}{\sqrt{N}} \sum \left(\frac{1}{f} \frac{\partial f}{\partial \beta} - \frac{1}{P} \frac{\partial P}{\partial \beta} \right). \quad (62)$$

Thus, in analogy to equation (12),

$$E\frac{1}{N}\left(\frac{\partial \log L}{\partial \beta}\right)^2 = \frac{1}{P}\left[\int_A \frac{h}{f}\left(\frac{\partial f}{\partial \beta}\right)^2 dtdx - \frac{1}{P}\left(\frac{\partial P}{\partial \beta}\right)^2\right]. \tag{63}$$

where $A = A_1 \cup A_2$.

If we divide Type (1) sample by P_1 and Type (2) sample by P_2, we maximize

$$L^* = P_1^{-N_1}\prod_1 h(x_i)f(t_i)P_2^{-N_2}\prod_2 h(x_i)f(t_i). \tag{64}$$

Ignoring the terms that do not depend on β, we have

$$\log L = -N_1 \log P_1 - N_2 \log P_2 + \sum \log f, \tag{65}$$

$$\frac{1}{\sqrt{N}}\frac{\partial \log L^*}{\partial \beta} = \frac{N_1}{\sqrt{N}}\frac{1}{P_1}\frac{\partial P_1}{\partial \beta} - \frac{N_2}{\sqrt{N}}\frac{1}{P_2}\frac{\partial P_2}{\partial \beta} + \frac{1}{\sqrt{N}}\sum\frac{1}{f}\frac{\partial f}{\partial \beta}$$

$$= -\frac{1}{P_1}\frac{\partial P_1}{\partial \beta}\frac{1}{\sqrt{N}}\sum\left(D_1 - \frac{P_1}{P}\right) - \frac{1}{P_2}\frac{\partial P_2}{\partial \beta}\frac{1}{\sqrt{N}}\sum\left(D_2 - \frac{P_2}{P}\right)$$

$$+ \frac{1}{\sqrt{N}}\sum\left(\frac{1}{f}\frac{\partial f}{\partial \beta} - \frac{1}{P}\frac{\partial P}{\partial \beta}\right), \tag{66}$$

where $D_1 = 1$ if the spell is of Type (1) and $D_2 = 1$ if it is of Type (2). Since the last term above is equal to the RHS of equation (62),

$$E\frac{1}{N}\left(\frac{\partial \log L^*}{\partial \beta}\right)^2 = \frac{1}{P_1^2}\left(\frac{\partial P_1}{\partial \beta}\right)^2\frac{P_1}{P}\left(1 - \frac{P_1}{P}\right) + \frac{1}{P_2^2}\left(\frac{\partial P_2}{\partial \beta}\right)^2\frac{P_2}{P}\left(1 - \frac{P_2}{P}\right)$$

$$-\frac{2}{P^2}\frac{\partial P_1}{\partial \beta}\frac{\partial P_2}{\partial \beta} - \frac{2}{P_1}\frac{\partial P_1}{\partial \beta}\left(\frac{1}{P}\frac{\partial P_1}{\partial \beta} - \frac{P_1}{P^2}\frac{\partial P}{\partial \beta}\right)$$

$$-\frac{2}{P_2}\frac{\partial P_2}{\partial \beta}\left(\frac{1}{P}\frac{\partial P_2}{\partial \beta} - \frac{P_2}{P^2}\frac{\partial P}{\partial \beta}\right) + E\frac{1}{N}\left(\frac{\partial \log L}{\partial \beta}\right)^2$$

$$= \frac{1}{P^2}\left(\frac{\partial P}{\partial \beta}\right)^2 - \frac{1}{PP_1}\left(\frac{\partial P_1}{\partial \beta}\right)^2 - \frac{1}{PP_2}\left(\frac{\partial P_2}{\partial \beta}\right)^2 + E\frac{1}{N}\left(\frac{\partial \log L}{\partial \beta}\right)^2. \tag{67}$$

The above can be shown to be equal to

$$-E\frac{1}{N}\frac{\partial^2 \log L^*}{\partial \beta^2}.$$

To see that L is the better likelihood function than L^*, verify

$$E\frac{1}{N}\left(\frac{\partial \log L}{\partial \beta}\right)^2 - E\frac{1}{N}\left(\frac{\partial \log L^*}{\partial \beta}\right)^2$$

$$= \frac{1}{PP_1}\left(\frac{\partial P_1}{\partial \beta}\right)^2 + \frac{1}{PP_2}\left(\frac{\partial P_2}{\partial \beta}\right)^2 - \frac{1}{P^2}\left(\frac{\partial P}{\partial \beta}\right)^2 \frac{1}{P^2}\left[\sqrt{\frac{P_2}{P_1}}\left(\frac{\partial P_1}{\partial \beta}\right) - \sqrt{\frac{P_1}{P_2}}\left(\frac{\partial P_2}{\partial \beta}\right)\right]^2 \geq 0, \tag{68}$$

VI. Conditional ESMLE with Left Censoring

Conditional density of t given x in A ($\equiv A_1 \cup A_2$) is given by

$$f(t \mid x) = \frac{h(x)f(t)}{\int_A h(x)f(t)dt} = \frac{f(t)}{G(x)}, \tag{69}$$

where

$$G(x) = F(b-x) - X_{(a,0)}(x)F(-x). \tag{70}$$

Note that $X_{(a,0)}(x) = 1$ if $x \in (a, 0)$ and $= 0$ otherwise. All the results of Section III will go through by replacing F by G. Thus,

$$AV(CESMLE)^{-1} = \frac{1}{P}\left[\int_A \frac{h}{f}\left(\frac{\partial f}{\partial \beta}\right)^2 dtdx - \int_a^b \frac{h}{G}\left(\frac{\partial G}{\partial \beta}\right)^2 dx\right]. \tag{71}$$

The above can be shown to be less than equation (63) by replacing F by G on the Cauchy-Schwartz inequality equation (36).

VII. Generalizations to the Case of Heterogenous Samples

So far, we have assumed that we have i.i.d. observations on the random variables X and T. In actual applications, however, their densities, h and f, are likely to depend on vectors of exogenous variables s_i and z_i so that we can write $h(x_i - s_i'\theta)$ and $f(t_i - z_i'\beta)$. We will indicate how the foregoing results should be modified to take into account these specifications. Below, we will indicate necessary modifications to some of the preceding equations.

(1) Replace h and f with $h(x_i - s_i'\theta)$ and $f(t_i - z_i'\beta)$.

(12) Add $\lim \frac{1}{N}\sum_{i=1}^{N}$ to every term after the above replacement.

(21) Same as above.

(34) Same as above.

(41) $W = \prod_{i=1} f(t_i - z_i'\beta)\hat{P}_{1i}^{-1}$.

(42) $\hat{P}_{1i} = \int_a^b F(b - x - z_i'\beta)d\hat{H}(x)$,

where $\hat{H}$ is a step function with a jump of size $1/K$ at $x_k - s_k'\hat{\theta}$, $\hat{\theta}$ being the least squares estimator of the regression of x_k on s_k. Therefore,

(43) $\hat{P}_{1i} = \frac{1}{K}\sum_{k=1}^{K} F(b - x_k + s_k'\theta - z_i'\beta)$,

A further error is introduced by the estimation of θ, but the rate of convergence is the same as in the case of the homogenous sample.

References

Ait-Sahalia, Y., "The Delta Method for Nonparametric Kernel Functionals," mimeo (revised), University of Chicago Graduate School of Business, August, 1994.

Amemiya, T., *Advanced Econometrics*, Cambridge, Massachusetts: Harvard University Press, 1985.

———, "A Note on Left Censoring" in C. Hsiao, K. Lahiri, L.-F. Lee, and M. Pesaran, eds. *Analysis of Panels, and Limited Dependent Variable Models*, Cambridge University Press, 1999, pp. 7–22.

Kim, Y.-D., D. R. Anderson, T. L. Amburgey, and J. C. Hickman, "The Use of Event History Analysis to Examine Insurer Insolvencies," *Journal of Risk and Insurance*, 62 (1), 1995, pp. 94–110.

Lee, S. H. and J. L. Urrutia, "Analysis and Prediction of Insolvency in the Property-Liability Insurance Industry: A Comparison of Logit and Hazard Models," *Journal of Risk and Insurance*, 63 (1), 1996, pp. 121–130.

Manski, C. F., and S. R. Lerman, "Estimation of Choice Probabilities from Choice-Based Samples," *Econometrica*, 45, 1977, pp. 1977–1989.