# A NOTE ON THE RATE OF CONVERGENCE OF ESTIMATORS OF MIXTURES OF WEIBULLS*

Bo E. Honoré
Department of Economics
Princeton University
Princeton, NJ 08544, U.S.A.

February 1994 (first version)

December 1997 (this version)

## 1 Introduction.

Estimation of duration models has been the subject of some attention in econometrics for almost two decades. This note is concerned with the precision with which it is possible to estimate the parameters of a Weibull regression model with unobserved heterogeneity. This model has a fairly long history in econometrics. Lancaster (1979) estimated a mixture of Weibull regression models in which the mixing distribution was specified as a Gamma–distribution. Elbers and Ridder (1982) proved as a special case that that model is identified even if the mixing distribution is unknown, as long as the mixing distribution has finite mean. Heckman and Singer (1984) proposed a consistent nonparametric maximum likelihood estimator. Honoré (1990) proposed another estimator, based on order statistics, and proved that it is consistent and converges in distribution to a normal distribution at rate $(\log n)^{-1} \cdot n^{-\rho}$ where $\rho$ can be made arbitrarily close to (but cannot exceed)

*The author would like thank James J. Heckman for helpful discussions, and the National Science Foundation and the Alfred P. Sloan Foundation for financial support. The first version of this paper was written while the author visited the University of Chicago in 1993-94. It was revised while the author visited the University of Copenhagen in the Fall of 1997. The hospitality of the Department of Economics at both institutions is gratefully acknowledged.

1/3. This result was derived under the assumption that the mixing distribution has finite second moment.

Much of the statistical and econometric literature investigates estimators that converge to the true parameter at rate $n^{-1/2}$. Hahn (1992) proved that it is not possible to construct such an estimator of the parameters of a mixture of Weibull regression models, and Ishwaran (1996a) has recently shown that the best possible rate of convergence is $n^{-k/(2k+1)}$, where $k+1$ is the number of finite moments assumed on the mixing distribution. Finally, Ishwaran (1996b) proposed an estimator that achieves the optimal rate of convergence.

The contribution of this note is to show that a slightly modified version of the estimator proposed in Honoré (1990) also attains the optimal rate of convergence derived in Ishwaran (1996a), when the number of assumed moments on the mixing distribution is an integer. The result of the paper is derived for the mixtures of Weibulls without covariates. It is well–known that in a mixed Weibull regression model, it is possible to construct $n^{-1/2}$–consistent estimators of the regression coefficients divided by the Weibull parameter (see e.g. Lancaster (1985)). The achievable rate of convergence for the regression coefficients is therefore no less than the achievable rate of convergence of the Weibull parameter.

The idea used here is related to the bias–reduction ideas used in non–parametric kernel estimation of densities and regression functions (see e.g. Prakasa–Rao (1983)). The assumption of higher moments of the mixing distribution ensure that the distribution of the data is "smooth" close to 0. This, in turn, makes it possible to use a Taylor series expansion to characterize the bias of the estimator proposed in Honoré (1990). The largest terms in the bias can then be eliminated using a "jackknife" approach.

Censoring is ignored throughout this paper. If the censoring times are bounded away from zero, then the result can be generalized because only observations close to 0 are used (asymptotically).

## 2 The Result

Assume that $T_1, T_2, \ldots T_n$ are *i.i.d.* observations from the distribution of the mixed Weibull model with parameter $\alpha$ and mixing distribution $G$,

$$P(T_i > t) = \int_0^\infty e^{-\theta t^\alpha}\, dG(\theta).$$

The problem is to estimate $\alpha$ without knowledge of $G$. To simplify the notation, we will initially consider estimation of $\alpha^{-1}$ rather than $\alpha$.

Let $X_i = T_i^{\alpha}$. Then

$$P(X_i > x) = P(T_i^{\alpha} > x) = P(T_i > x^{1/\alpha}) = \int_0^{\infty} e^{-\theta x}\, dG(\theta).$$

In other word, the distribution of $X_i$ is a mixture of exponentials with mixing distribution $G$.

Generalizing the approach in Honoré (1990), we will consider estimation of $\alpha^{-1}$ by a linear combination of order statistics of $T_1, \ldots, T_n$ of the form

$$\widehat{\alpha^{-1}} = \sum_{j=1}^{M} \gamma_j \log T_{jn} \tag{1}$$

where $T_{jn}$ denotes the $n \cdot m_j^{\text{th}}(n)$ order statistic of $T_1, T_1, \ldots, T_n$ with $m_j(n) = a_j n^{-d}$, $a_j > 0$ and $0 < d < 1$. $M$ is some fixed integer.

As

$$\alpha^{-1} \log X_{jn} = \log T_{jn}, \tag{2}$$

the rate of convergence of $\widehat{\alpha^{-1}}$ to $\alpha^{-1}$ equals the rate of convergence of $\sum_{j=1}^{M} \gamma_j \log X_{jn}$ to 1. Note that

$$\log X_i = \log E_i - \log \Theta_i$$

where $E_i$ has a unit exponential distribution and $\Theta_i$ has distribution $G$. Since the estimators considered here all rely on the extreme lower order statistics of $\{\log X_i\}_{i=1}^{n}$, it is not surprising that their asymptotic properties depend on the number of moments of $G$ (i.e., on the left tail of the distribution of $-\log \Theta_i$).

By Lemma 1 in Honoré (1990, page 456),

$$\log X_{jn} = \log \mu_{jn} + O_p\left(n^{-(1-d)/2}\right) \tag{3}$$

where $\mu_{jn}$ is the unique value of $\eta$ that solves $m_j(n) = 1 - \int_0^{\infty} e^{-\theta\eta}\, dG(\theta)$. To gain insight regarding the asymptotic behavior of $\log X_{jn}$, we must therefore examine $\log \mu_{jn}$. Assuming that the mixing distribution, $G$, has $k+1$ finite moments, we can write[1]

$$\log \mu_{jn} = c_0 + \log m_j(n) + \sum_{\ell=1}^{k-1} c_\ell m_j(n)^{\ell} + O\left(m_j(n)^{k}\right) \tag{4}$$

[1] When $k = 1$, this is essentially Lemma 2 of Honoré (1990). The derivation is a straightforward generalization of the proof of that lemma.

where $c_0, c_1, \ldots, c_k$ are functions of the moments of the mixing distribution and, consequently, unknown. Honoré (1990) considered linear combinations of two $X_{jn}$'s based on $m_j(n) = n^{-d_j}$. This eliminated the leading unknown term $c_0$ in (4) and made the second term equal to 1. Here, it will be illustrated that the remaining terms in (4) can be eliminated as well.

Note that with $m_j(n) = a_j n^{-d}$

$$\sum_{j=1}^{M} \gamma_j \log \mu_{jn} = \left(\sum_{j=1}^{M} \gamma_j\right) c_0 + \sum_{j=1}^{M} \gamma_j \log a_j - \left(\sum_{j=1}^{M} \gamma_j\right)(1-d)\log n + \sum_{\ell=1}^{k-1}\left(\sum_{j=1}^{M} \gamma_j a_j^{\ell}\right) c_\ell n^{-\ell d} + O\left(n^{-kd}\right) \tag{5}$$

Hence as long as $(\gamma_1, \ldots, \gamma_M)$ and $(a_1, \ldots, a_M)$ are chosen so that

$$\begin{aligned} \sum_{j=1}^{M} \gamma_j a_j^{\ell} &= 0, \quad \ell = 0, \ldots k-1 \\ \sum_{j=1}^{M} \gamma_j \log a_j &= 1, \end{aligned}$$

(5) reduces to

$$\sum_{j=1}^{M} \gamma_j \log \mu_{jn} = 1 + O\left(n^{-kd}\right)$$

and hence by (3)

$$\sum_{j=1}^{M} \gamma_j \log X_{jn} = 1 + O\left(n^{-kd}\right) + O_p\left(n^{-(1-d)/2}\right)$$

With $d = \frac{1}{2k+1}$, we therefore have

$$\sum_{j=1}^{M} \gamma_j \log X_{jn} = 1 + O_p\left(n^{-k/(2k+1)}\right)$$

and we therefore have the desired result:

$$\widehat{\alpha} = \frac{1}{\sum_{j=1}^{M} \gamma_j \log T_{jn}} = \alpha + O_p\left(n^{-k/(2k+1)}\right)$$

# 3 Discussion.

The estimator proposed in Honoré (1990) assumes two moments of the mixing distribution. The results in Ishwaran (1996a) imply that the optimal rate of convergence in that case is $n^{-1/3}$. The estimator in Honoré (1990) does not achieve this rate, but the rate can be made arbitrarily close to that rate[2]. In contrast, the estimator proposed here does achieve the optimal rate. The reason for the difference between the two estimators is that for the former it was important to have an estimator that was asymptotically normally distributed with a limiting distribution that is centered around the true value. This was achieved by using a jackknife techniques that was based on two estimators that have different rates of convergence. Here, the rate of convergence is the same for all the estimators on the right hand side of (1), and although the jackknifed estimator will be asymptotically normal (by a slight modification of the argument in Honoré (1990)), its limiting distribution will not be centered around the true parameter value. Of course, if the aim of the asymptotic theory is to justify construction of confidence intervals and test–statistics, then it might be preferable to use an estimator which converges at a rate that is slightly slower than the optimal rate, but which has a limiting distribution that is centered around the true value.

Unlike the estimators in Honoré (1990), the estimators considered here are not guaranteed to be positive, even though the true value of $\alpha$ must be. This is because the $\gamma$'s are not, in general, all positive. This, of course, plays no role from the point of view of the rate of convergence of the resulting estimator. However, if it is important to construct a estimator that is always positive, then it might be useful to consider the logarithm of (2),

$$-\log\alpha + \log\left(-\log X_{jn}\right) = \log\left(-\log T_{jn}\right) \tag{6}$$

where the arguments of the logarithms are positive with probability approaching 1 as the sample size increases to infinity. (6) suggests that $-\log\alpha$ can be estimated by linear combinations of the order statistics of[3] $\{\log\left(-\log T_i\right)\}_{i=1}^{n}$.

# 4 References.


Elbers, C. and G. Ridder (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model." *Review of Economic Studies*, 49, pp. 403–409.


[2]In an earlier version of this paper, it was demonstrated that by considering linear combinations of the estimators proposed in Honoré (1990), one can construct estimators that converge at a rate that is arbitrarily close to $n^{-k/(2k+1)}/\log n$ where $k+1$ is the number of finite moments on the mixing distribution.

[3]Again, since we are considering only the extreme low order statistics, the logs will be defined with probability approaching 1.

Hahn, J. (1992): "Efficiency Bounds of the Mixed Proportional Hazard Model," *Review of Economic Studies*, 61, pp. 607 - 629

Heckman, J. J. and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Duration Models," *Econometrica*, 52, pp 271–320.

Honoré, B. E. (1990): "Simple Estimation of a Duration Model with Unobserved Heterogeneity." *Econometrica,* 58, pp. 453–473.

Ishwaran, H. (1996a): "Identifiability and Rates of Convergence of Estimation for Scale Parameters in Location Mixture Models." *Annals of Statistics*, 24, pp. 1560–1571.

Ishwaran, H. (1996b): "Uniform Rates of Estimation in the Semiparametric Weibull Mixture Model." *Annals of Statistics*, 24, pp. 1572–1585.

Lancaster, T. (1979): "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47, pp. 939–956.

Lancaster, T. (1985): "Generalized Residuals and Heterogeneous Duration Models," *Journal of Econometrics*, 28, pp. 155–169.

Prakasa–Rao, B. L. S. (1983): *Nonparametric Functional Estimation.* Orlando: Academic Press.