Ann. Inst. Statist. Math.
Vol. 45, No. 4, 721–729 (1993)

# A NOTE ON SMOOTHED ESTIMATING FUNCTIONS

A. THAVANESWARAN[1] AND JAGBIR SINGH[2]

[1] *Department of Statistics, University of Manitoba, Winnipeg, Canada R3T 2N2*
[2] *Department of Statistics, Temple University, Philadelphia, PA 19122, U.S.A.*

(Received December 6, 1991; revised December 21, 1992)

**Abstract.** The kernel estimate of regression function in likelihood based models has been studied in Staniswalis (1989, *J. Amer. Statist. Assoc.*, **84**, 276–283). The notion of optimal estimation for the nonparametric kernel estimation of semimartingale intensity $\alpha(t)$ is proposed. The goal is to arrive at a nonparametric estimate $\hat{\theta}_0$ of $\theta_0 = \alpha(t_0)$ for a fixed point $t_0 \in [0, 1]$. We consider the estimator that is a solution of the smoothed optimal estimating equation $S_{t_0,\theta_0} = \int_0^1 w((t_0-s)/b)dG_s^0 = 0$ where $G_t^0 = \int_0^t a_{s,\theta_0}^0 dM_{s,\theta_0}$ is the optimal estimating function as in Thavaneswaran and Thompson (1986, *J. Appl. Probab.*, **23**, 409–417).

*Key words and phrases*: Censored observations, semimartingales, optimal estimation, smoothing.

## 1. Introduction

In a recent paper, Staniswalis (1989) discussed the problem of obtaining a kernel estimate of a regression function in likelihood based models with independent observations. In this article we are concerned with optimal estimation approach to nonparametric estimation of semimartingale intensity $\alpha(t)$. Our approach is analogous to the nonparametric regression approach that was pioneered by Priestley and Chao (1972), and applied to likelihood based models in Staniswalis (1989).

A semimartingale is a stochastic process which can be represented as the sum of a process of bounded variation and a local martingale. In the case of continuous time, a typical example of semimartingale in such a process $(X(t), t \geq 0)$ with independent increments for which $E|X(t)|$ is finite and a function of locally bounded variation. The class of semimartingales includes point processes, Ito processes, diffusion processes, etc. Consider a continuous time stochastic process $(X(t), t \geq 0)$ defined on $(\Omega, A, P)$ a complete probability space for each $P$ in a family $\{P\}$ of probability measures, and a family $F = [F_t, t \geq 0]$ of $\sigma$ algebras satisfying the standard conditions $F_s \subseteq F_t \subseteq A$ for $s \leq t, F_0$ augmented by sets of measure zero of $A$, and $F_t = F_{t+}$, where $F_{t+} = \bigcap_{s>t} F_s$. We denote by $D$ the space of right-continuous functions $x = (x_t, t \geq 0)$ having limits on the left. We use $X = (X(t), F_t)$ to denote an $F_t$-adapted random processes $(X(t))$ with trajectories in the space $D$. For simplicity we assume that $X(0) = 0$. We shall denote by $M^2_{loc}(F, P)$ a class of locally square integrable martingales $(H(t), F_t)$. Assume that the process $(X(t), F_t)$ is a semimartingale for each $P$, that is for each $P$ it can be represented in the form

$$X(t) = V(t) + H(t) \tag{1.1}$$

where $V(t)$ is a locally bounded variation process and $H(t) \in M^2_{loc}(F, P)$. When we allow $V(t)$ and $H(t)$ to depend on $P \in \{P\}$ only through $\theta$, the model (1.1) can be written as

$$X(t,\theta) = V(t,\theta) + H(t,\theta). \tag{1.2}$$

When $\theta \in R$, optimal as well as recursive estimates have been studied in Thavaneswaran and Thompson (1986). We consider the following semimartingale model of the form

$$dX(t) = \alpha(t)Y(t-)dR(t) + dM(t), \tag{1.3}$$

where $\alpha(t)$ is an unobservable deterministic part of the intensity of the process $X(s)$, $[X(s), Y(s), R(s), q \leq s \leq t]$ are observable processes, $M(t) \in M^2_{loc}(F, P)$ with predictable variance process $\langle M\rangle_t = \int_0^t C(s)dR(s)$ and $C(s)$ is a known function of the observations and $\alpha(s)$. For a similar restriction that the conditional mean and the conditional variance of $X(t)$ are absolutely continuous with respect to $R(t)$ see Hutton and Nelson (1986).

*Example* 1.1. Poisson process: When $X(t)$ is a right continuous process having jumps of size 1 and $\lambda(t) = \int_0^t \alpha(s)Y(s)ds$ with $Y(s) = 1$ is a deterministic function, the semimartingale model (1.3) becomes a nonhomogeneous Poisson process model with cumulative intensity $\lambda(t)$.

*Example* 1.2. Multiplicative intensity model: When $X(t)$ denotes the number of deaths up to time $t$, $\alpha(t)$ is the hazard rate, $Y(t)$ is the number of individuals at risk just before time $t$, the semimartingale model (1.3) takes the form

$$X(t) = \int_0^t \alpha(s)Y(s)ds + M(t) \tag{1.4}$$

where $M(t)$ is a zero mean square integrable martingale with variance process $\langle M\rangle_t = \int_0^t \alpha(s)Y(s)ds$ provided there are no simultaneous deaths. The model (1.4) introduced by Aalen (1978) has been widely applied to such phenomena as the life history data or arrivals at an intensive care unit of a hospital.

## 2. Proposed smoother for $\alpha(t_0)$

For a regression model $y_i = g(x_i) + e_i$ with identically distributed independent errors $e_i$ Staniswalis (1989) proposed an estimator $\hat{\lambda}_0$ of $\lambda_0 = g(x_0)$ as the one which maximizes the weighted likelihood function

$$w(\lambda) = \sum_i^n w\left(\frac{x_0 - x_i}{b}\right) \log f(y_i - \lambda), \tag{2.1}$$

where $f$ is the density of $e_i$, $w$ is a symmetric kernel with compact support and the bandwidth $b$ that controls the degree of smoothing. If the $e_i$'s are normal random variables, then $\hat{\lambda}_0$ becomes the kernel estimator of Priestley and Chao (1972). Here in analogy with (2.1) we propose a smoothed estimator for $\alpha(t_0)$ as a solution of the smoothed optimal estimating equation

$$\int_0^1 w\left(\frac{t_0 - s}{b}\right) dG_s^0 = 0 \tag{2.2}$$

where

(i) $G_t^0 = \int_0^1 a_{s,\theta}^0 dM_{s,\theta}$, as in Thavaneswaran and Thompson (1986) is an optimal estimating function defined through an optimal function $a_{s,\theta}^0 = (Y(s)/C(s)) \cdot J(s)$ where $J(s) = I(Y(s) > 0, C(s) > 0)$.

(ii) $w$ is a non-negative integrable kernel function, and $b$ is a positive bandwidth.

Then the estimate of $\theta_0$ is solution of

$$S_{t_0,\theta} = \int_0^1 w\left(\frac{t_0 - s}{b}\right) J(s)Y(s)C^{-1}(s)(dX(s) - \theta Y(s))dR(s) = 0, \tag{2.3}$$

and is given by

$$\tilde{\theta}_0 = \frac{\int_0^1 w\left(\frac{t_0 - s}{b}\right) C^{-1}(s)J(s)Y(s)dX(s)}{\int_0^1 w\left(\frac{t - s}{b}\right) Y(s)C^{-1}(s)J(s)Y(s)dR(s)}.$$

The proposed smoother $\tilde{\theta}_0$ for a counting process model (1.4) (where $R(t) = t$, $C(t) = \theta Y(t) > 0$ all $t$), becomes

$$\tilde{\theta}_0 = \frac{\int_0^1 w\left(\frac{t_0 - s}{b}\right) dX(s)}{\int_0^1 w\left(\frac{t_0 - s}{b}\right) Y(s)I(Y(s) > 0)ds}.$$

This turns out to be the maximum likelihood estimator for $\alpha$ if $\alpha(t)$ were constant $= \alpha$ and $w(\cdot) \equiv 1$. This cannot be said for the estimator for $\alpha(t_0)$ given in Thavaneswaran (1988) which is

$$\hat{\alpha}(t_0) = \frac{1}{b} \int_0^1 w\left(\frac{t_0 - s}{b}\right) \frac{dX(s)}{Y(s)}.$$

That is, when $\alpha(t)$ is constant, the optimal smoother corresponds to the maximum likelihood estimator in a sense in which the Thavaneswaran (1988) estimator does not.

Moreover, the proposed smoother for a homogeneous Poisson process model indexed by $n$ can be written as

$$\tilde{\theta}_n(t_0) = \frac{\int_0^1 w\left(\dfrac{t_0 - s}{b_n}\right) dN_n(s)}{\int_0^1 w\left(\dfrac{t_0 - s}{b_n}\right) ds} = \theta_0 + \frac{\int_0^1 w\left(\dfrac{t_0 - s}{b_n}\right) dM_n(s)}{\int_0^1 w\left(\dfrac{t_0 - s}{b_n}\right) ds}, \qquad n = 1, 2, \ldots$$

Hence, $E\tilde{\theta}_n(t) = \theta_0$ and the optimal smoother is unbiased for any sample size $n$.

We note that the Ramlau-Hansen (1983) estimate which was introduced for nonparametric purposes, is

$$\hat{\alpha}_n(t_0) = \frac{1}{b_n} \int_0^1 w\left(\frac{t_0 - s}{b_n}\right) dN_n(s),$$

and

$$\begin{aligned} E[\hat{\alpha}_n(t_0)] &= \frac{1}{b_n} \int_0^1 w\left(\frac{t_0 - s}{b_n}\right) \alpha(s) ds \\ &\approx \alpha(t_0) - \alpha^{(1)}(t_0) b_n \int_{-1}^1 u w(u) du + 1/2 \alpha^{(11)}(t) b_n^2 \int_{-1}^1 u^2 w(u) du \end{aligned}$$

i.e. $E[\hat{\alpha}_n(t_0)] \to \alpha(t_0)$ as $n \to \infty$. This implies that in general the kernel estimator $\hat{\alpha}_n(t)$ is not an unbiased estimator of $\alpha(t)$, but the bias tends to zero as $n \to \infty$.

*Note.* Ramlau-Hansen (1983) estimator is not finite-sample unbiased but is optimal for a large class of multiplicative intensity models.

*Example* 2.1. The relationship with maximum likelihood estimator: Nonparametric estimation of the drift function for stochastic differential equations has been studied by Nguyen and Pham (1982) and Leskow and Rozanski (1989). Recently, Leskow and Rozanski (1989) have studied the maximum likelihood estimate of $\alpha(t)$ using the method of sieves. They have considered a sequence of point processes

$$N_k(t) = \int_0^t \alpha(s) Y_k(s) ds + M_k(t) \tag{2.4}$$

and the sieve $S(n)$ defined by discretizing the parameter space

$$S(n) = \left[\alpha \in I \mid \alpha(s) = \sum_{l=1}^{m_n} x_l I_{A_{l,m_n}}(s)\right],$$

where $A_{l,m_n} = [l-1/m_n, l/m_n]$ for fixed $s$, let $l(n,s) = \{1,2,\ldots,m_n\}$ be such that $s \in A_{l(n,s),m_n} = B_n^s$ (say). They showed that

$$\hat{\alpha}_n(s) = \frac{\sum_{k=1}^n N_k(B_n^s)}{\sum_{k=1}^n \int_{B_n^s} Y_k(u)du}. \tag{2.5}$$

It is of interest to note that the smoother obtained by solving (2.3) for a sequence of counting processes in (2.4) reduces to $\hat{\alpha}_n(s)$ as in (2.5) if we set $w((t-s)/b_n) = I_{B_n^s}$ the indicator function of $B_n^s$.

That is, the maximum likelihood estimate of the discretized locally constant (piecewise constant) parameter, obtained using sieves is a special case of the smoother with particular weight.

*Example* 2.2. Diffusion process model: For the diffusion model considered in Thompson and Thavaneswaran (1990)

$$dX(t) = \alpha(t)dt + \sigma(X(t))dW(t), \qquad 0 \le t \le 1$$

where $(X(t), t \ge 0)$ is the one-dimensional observation process, $(W(t), t \ge 0)$ is the standard Wiener process, $\sigma(\cdot)$ is a known function of $X(t)$, and $\alpha(\cdot)$ is an unknown function to be estimated.

The estimate of $\theta_0 = \alpha(t_0)$ is a solution of

$$S_{t,\theta} = \int_0^1 w\left(\frac{t_0 - s}{b}\right) \frac{(dX(s) - \theta ds)}{\sigma^2(X(s))} = 0$$

and is given by

$$\tilde{\theta}_0 = \frac{\int_0^1 w\left(\frac{t_0 - s}{b}\right) dX(s)/\sigma^2(X(s))}{\int_0^1 w\left(\frac{t_0 - s}{b}\right) ds/\sigma^2(X(s))}.$$

*Example* 2.3. Discrete time stochastic processes: Let $Z_1, Z_2, \ldots, Z_t, Z_{t+1}, \ldots, Z_n$ be a series having the conditional moments with respect to $F_{t-1}^z$, the $\sigma$-field generated by $Z_1, \ldots, Z_{t-1}$.

$$E[Z_t \mid F_{t-1}^z] = g(t)h(F_{t-1}^z); \qquad \mathrm{Var}[Z_t \mid F_{t-1}^z] = \sigma^2(F_{t-1}^z).$$

Then, $Z_t$ can be written as

$$Z_t = E[Z_t \mid F_{t-1}^z] + Z_t - E[Z_t \mid F_{t-1}^z] = g(t)h(F_{t-1}^z) + \epsilon_t$$

where $\epsilon_t = Z_t - E[Z_t \mid F_{t-1}^z]$.

*Note.* (i) When $h(\cdot) = 1$, the above model corresponds to a time series model with a time varying parameter $g(t)$.

(ii) Furthermore, if $\epsilon_t$'s are independent $h(\cdot) = 1$ and $\sigma^2(\cdot) =$ constant then the above model corresponds to a regression model considered in Staniswalis (1989).

The smoothed version of least squares estimating function for estimating $\theta_0 = g(t_0)$ can be written as

$$S_n^{LS}(t_0) = \sum_{t=1}^{n} w\left(\frac{t_0 - t}{b}\right) h(F_{t-1}^z)(Z_t - \theta h(F_{t-1}^z)).$$

While the corresponding smoother is the solution of

$$S_n^{opt}(t_0) = \sum_{t=1}^{n} w\left(\frac{t_0 - t}{b}\right) a_{t-1}^0 \epsilon_t = 0$$

where $a_{t-1}^0 = (\partial\epsilon_t/\partial\theta)/\sigma^2(F_{t-1}^z)$ is the optimal value as in Godambe (1985).

*Note.* If $\epsilon_t$'s are independent and having density $f(\cdot)$ then it follows from Godambe (1960) that the optimal estimating function for $\theta_0$ in $y_i = \theta_0 + \epsilon_i$ is the score function

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(y_i - \theta) = 0$$

and the corresponding smoothed optimal estimating function is

$$\sum_{i=1}^{n} w\left(\frac{x_0 - x_i}{b}\right) \frac{\partial}{\partial\theta} \log f(y_i - \theta) = 0$$

which is the same as the one considered in Staniswalis (1989).

## 3. Asymptotics

The optimal smoother for $\alpha(0) = \theta_0$ for fixed $t_0$ from a sequence of semi-martingales indexed by $n$, (i.e.) $dX_n(t) = \alpha(t)Y_n(t)dR_t + dM_n(t)$, $0 \le t \le 1$, can be written as

$$\begin{aligned}
\tilde{\theta}_n(t) &= \frac{\int_0^1 w\left(\frac{t-s}{b_n}\right) Y_n(s) C_n^{-1}(s) J(s) dX_n(s)}{\int_0^1 w\left(\frac{t-s}{b_n}\right) Y_n^2(s) J(s) C_n^{-1}(s) dR(s)} \\
&= \theta_0 + \frac{\int_0^1 w\left(\frac{t-s}{b_n}\right) Y_n(s) J(s) C_n^{-1}(s) dM_n(s)}{\int_0^1 w\left(\frac{t-s}{b_n}\right) Y_n^2(s) J(s) C_n^{-1}(s) dR(s)} \\
&= \theta_0 + \frac{\bar{M}_n}{A_n},
\end{aligned}$$

where

$$\bar{M}_n = \int_0^1 w\left(\frac{t-s}{b_n}\right) Y_n(s)J(s)C_n^{-1}(s)dM_n(s) \qquad \text{and}$$
$$A_n = \int_0^1 w\left(\frac{t-s}{b_n}\right) Y_n^2(s)J(s)C_n^{-1}(s)dR(s).$$

*Strong Consistency.*

THEOREM 3.1. *Let $m_n = \sum_{i=1}^n(\Delta\bar{M}_i/A_i)$ and under the assumption that* (i) *the predictable variance process of $m_n$, $\langle m\rangle_\infty < \infty$ a.s. and* (ii) *the corresponding predictable variance process of $M_n$, $A_\infty = \infty$ a.s. The optimal smoother $\tilde{\theta}_n(t_0) \to \theta_0$ a.s. i.e. $\tilde{\theta}_n(t) \to \alpha(t)$ a.s. for all fixed $t$ as $n \to \infty$.*

*Note.* The assumptions (i) and (ii) are somewhat restrictive for a general semimartingale model. However these can be easily verified for an autoregressive model of order one as in Shiryayev ((1984), p. 489).

PROOF. Let $\bar{M}_n = \int_0^1 H_n(s)dM_n(s)$ where $H_n(s) = w((t-s)/b_n)Y_n(s)J(s)\cdot C_n^{-1}(s)$ is a predictable process and $M_n(s)$ is a zero mean square integrable martingale and hence, the stochastic integral $\int_0^1 H_n(s)dM_n(s)$ is a zero mean square integrable martingale. Furthermore, $A_n(t) = \int_0^1 w((t-s)/b_n)Y_n^2(s)C_n^{-1}(s)dR(s)$ is a Lebseque-Stieltjes integral of a predictable process with respect to a bounded variation function $R(s)$ and is predictable. Therefore, $\bar{M}_n/A_n$ is a martingale sequence. The proof now follows by applying the strong law of large numbers for the martingale sequence $\bar{M}_n/A_n$, as in Shiryayev ((1984), p. 487).

*Asymptotic Normality.*
We have shown that $\tilde{\theta}_{n,0}(t) - \theta = \bar{M}_n/A_n$.

THEOREM 3.2. *Assume that*
(i) *$J(s)Y(s)/C_n(s) \to 1/\sigma(s)$, as $n \to \infty$, in probability,*
(ii) *the functions $\alpha$ and $\sigma$ are continuous at the point $t$,*
(iii) *$R(t) = t$, $\gamma_s = \lim_n J(s)Y_n^2(s)/nC_n(s)$, in probability uniformly in a neighbourhood of $t$.*
*Then, $\sqrt{n}[\tilde{\theta}_n(t) - \theta]$ converges in distribution to a normal distribution with mean 0 and variance*

$$\Sigma_t = \left(\int_{-1}^{+1} w^2(u)du\right) \Big/ \left(\int_{-1}^{+1} w(u)du\right)^2 \frac{1}{\gamma_t}.$$

PROOF. Recall that

$$\tilde{\theta}_n(t_0) - \theta = \frac{\bar{M}_n}{A_n} = \frac{\int_0^1 H_n(s)dM_n(s)/\sqrt{nb_n}}{A_n \mid nb_n},$$

where $H_n(s)$ and $A_n$ are as defined earlier, $\{\bar{M}_n\}$ is a sequence of martingales and

(a)
$$\begin{aligned}\langle \bar{M}_n \rangle &= \frac{1}{nb_n}\int_0^1 H_n^2(s)d\langle M\rangle_n(s) = \frac{1}{nb_n}\int_0^1 H_n^2(s)C_n(s)ds \\ &= \int w^2(u)\frac{Y_n^2(t-b_nu)J(t-b_nu)}{C_n^2(t-b_nu)}\frac{C_n(t-b_nu)}{n}du \\ &\xrightarrow{P} \gamma_t \int_{-1}^{+1} w^2(u)du = \Sigma_{1t},\end{aligned}$$

(b)
$$[|H_n(s)] > \epsilon] = \left[\left| w\left(\frac{t-s}{b_n}\right)\frac{J(s)Y_n(s)}{C_n(s)}\right| > \epsilon\sqrt{(nb_n)}\right]$$

as $n \to \infty$, $b_n \to 0$, $Y_n(s)/C_n(s) \xrightarrow{P} 1/\sigma(s)$ uniformly in a neighbourhood of $t$ and $1/\sigma(t)$ is bounded in this neighbourhood, thus

$$\boldsymbol{I}[|H_n(t)| > \epsilon] \to 0 \quad \text{in probability.}$$

Then, applying the martingale central limit theorem (Shiryayev (1984), Theorem 4, p. 511), as $n \to \infty$, $\bar{M}_n(t) \to N(0, \Sigma_{1t})$ in distribution. Moreover, $A_n/nb_n \to \gamma_t \int_{-1}^{+1} w(u)du = \Sigma_{2t}$ in probability. Hence, $\sqrt{nb_n}(\theta_n^0 - \theta) \to N(0, \Sigma_t)$ in distribution where $\Sigma_t = (\int_{-1}^{+1} w^2(u)du)/(\int_{-1}^{+1} w(u)du)^2(1/\gamma_t)$. □

## Acknowledgements

Our sincere thanks are extended to the referee and to Professor M. E. Thompson for valuable comments leading to a substantial improvement in the presentation of the paper.

## References

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes, *Ann. Statist.*, **6**, 701–726.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood equations, *Ann. Math. Statist.*, **31**, 1208–1211.

Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes, *Biometrika*, **72**, 419–428.

Hutton, J. E. and Nelson, P. I. (1986). Quasi-likelihood estimation for semimartingales, *Stochastic Process. Appl.*, **22**, 245–257.

Leskow, J. and Rozanski, R. (1989). Histogram maximum likelihood estimate in the multiplicative intensity model, *Stochastic Process. Appl.*, **31**, 151–159.

Nguyen, H. T. and Pham, T. D. (1982). Identification of nonstationary diffusion model by the method of sieves, *SIAM J. Control Optim.*, **20**, 603–611.

Priestley, M. B. and Chao, M. T. (1972). Non-parametric function fitting, *J. Roy. Statist. Soc. Ser. B*, **34**, 385–392.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions, *Ann. Statist.*, **11**, 453–466.

Shiryayev, A. N. (1984). *Probability*, Springer, New York.

Staniswalis, J. G. (1989). The kernel estimate of a regression function in likelihood-based models, *J. Amer. Statist. Assoc.*, **84**, 276–283.

Thavaneswaran, A. (1988). Smoothing signals for semimartingales, *Stochastic Process. Appl.*, **28**, 81–89.

Thavaneswaran, A. and Thompson, M. E. (1986). Optimal estimation for semimartingales, *J. Appl. Probab.*, **23**, 409–417.

Thompson, M. E. and Thavaneswaran, A. (1990). Optimal nonparametric estimates for stochastic differential equations, *Journal of Applied Mathematico and Computations*, **37**, 169–183.