Ann. Inst. Statist. Math.
34 (1982), Part B, 373-387

THE APPLICATION OF LINEAR INTENSITY MODELS TO THE INVESTIGATION OF CAUSAL RELATIONS BETWEEN A POINT PROCESS AND ANOTHER STOCHASTIC PROCESS

Y. OGATA, H. AKAIKE AND K. KATSURA

(Received Mar. 26, 1981; revised Oct. 6, 1981)

Summary

The computational aspect of the fitting of a parametric model for the analysis of the influence of an input to a point process output is discussed. The feasibility of the procedure is demonstrated by an artificial example. Its practical utility is illustrated by applying it to the analysis of the causal relation between two earthquake series data from certain seismic regions of Japan.

1. Introduction

Consider a point process defined by the intensity process

$$\lambda(t)=\mu+\int_0^t g(t-s)dN_s+\int_0^t h(t-s)dX_s\ , \tag{1.1}$$

where $\{N_t\}$ denotes the point process and $\{X_t\}$ the input process which may be either a point process or a cumulative process

$$X_t=\int_0^t x(s)ds$$

of a stochastic process $x(t)$. Given bivariate data $\{N_t, X_t;\ 0\leqq t\leqq T\}$, we are interested in estimating the response functions $g(t)$ and $h(t)$ in (1.1). When $h(t)\equiv 0$, this means that there is no causal relation between the input $\{X_t\}$ and the output $\{N_t\}$, while $g(t)\equiv 0$ means that the output process is a doubly stochastic Poisson process whose intensity is modulated only by $\{X_t\}$.

In [7] we proposed a parametrization by the Laguerre type polynomials

$$g(t)=\sum_{k=1}^{K} a_k t^{k-1}e^{-ct} \qquad \text{and} \qquad h(t)=\sum_{k=1}^{L} b_k t^{k-1}e^{-ct}\ . \tag{1.2}$$

The log partial likelihood in the sense of Cox [3] is then defined by

$$\log L_T(\theta)=\int_0^T \log \lambda_\theta(t)dN_t-\int_0^T \lambda_\theta(t)dt \,, \tag{1.3}$$

where θ stands for $(\mu, c, a_1, \cdots, a_K, b_1, \cdots, b_L)$. The estimation of the parameters can be realized by applying the method of maximum likelihood to the partial likelihood and the selection of the orders K and L is realized by the minimum AIC procedure [1] which minimizes

$$\text{AIC}=(-2)\max(\text{log partial likelihood}) + 2(\text{number of parameters}) \,. \tag{1.4}$$

In this paper we will discuss the numerical aspect of the process of parameter estimation and order selection and demonstrate the performance of the procedure by applying it to both artificial and real data.

2. The likelihood computation and the minimum AIC procedure

Given a pair of records of occurrence times of two events $\{t_i;\ i=1,\cdots,I\}$ and $\{\tau_m;\ m=1,\cdots,M\}$ over the time interval $[0,T]$ the partial log likelihood (1.3), with $\{t_i\}$ as the output and $\{\tau_m\}$ as the input, is given by

$$\log L_T(\theta)=\sum_{i=1}^{I}\log\left\{\mu+\sum_{k=1}^{K}a_kP_k(i)+\sum_{k=1}^{L}b_kQ_k(i)\right\} -\left\{\mu T+\sum_{k=1}^{K}a_k\sum_{i=1}^{I}R_k(T-t_i)+\sum_{k=1}^{L}b_k\sum_{m=1}^{M}R_k(T-\tau_m)\right\} \tag{2.1}$$

where

$$P_k(i)=\sum_{t_j<t_i}(t_i-t_j)^{k-1}e^{-c(t_i-t_j)} \,, \tag{2.2}$$

$$Q_k(i)=\sum_{\tau_m<t_i}(t_i-\tau_m)^{k-1}e^{-c(t_i-\tau_m)} \tag{2.3}$$

and

$$R_k(t)=\int_0^t t^{k-1}e^{-ct}dt \,. \tag{2.4}$$

When a continuous record $\{x(t)\}$ over the time interval $[0,T]$ is given as the record of the input, the intensity process (1.1) is approximated by

$$\lambda(t)=\mu+\sum_{t_i<t}g(t-t_i)+\sum_{\sigma_m<t}h(t-\sigma_m)x(\sigma_m)\cdot(M/T) \,, \tag{2.5}$$

where M is a properly chosen large integer and $\sigma_m=(m/M)T-1/(2MT)$,

$m=1,\cdots,M$. With the parametrization (1.4), the approximate partial log likelihood is given by

$$\text{(2.6)} \quad \log L_T^*(\theta)=\sum_{i=1}^{I}\log\left\{\mu+\sum_{k=1}^{K}a_kP_k(i)+\sum_{k=1}^{L}b_kU_k(i)\right\} -\left\{\mu T+\sum_{k=1}^{K}a_k\sum_{i=1}^{I}R_k(T-t_i)+\sum_{k=1}^{L}b_k\sum_{m=1}^{M}x(\sigma_m)R_k(T-\sigma_m)\right\},$$

where $P_k(i)$ and $R_k(t)$ are given by (2.2) and (2.3), and

$$\text{(2.7)} \quad U_k(i)=\sum_{\sigma_m<t_i}x(\sigma_m)(t_i-\sigma_m)^{k-1}e^{-c(t_i-\sigma_m)}.$$

The derivation of these partial log likelihoods is discussed in [7].

Given a pair of orders (K, L), the maximum likelihood estimates of the parameters can be obtained by using a non-linear optimization technique developed by Fletcher and Powell [4] which requires only the gradient for the optimization. The gradients of the partial log likelihood functions can easily be obtained by differentiating the above likelihood functions with respect to the parameters. They are given as follows:

$$\text{(2.8)} \quad \frac{\partial \log L_T}{\partial\mu}=\sum_{i=1}^{I}1/\Lambda(i)-T,$$

$$\text{(2.9)} \quad \frac{\partial \log L_T}{\partial a_k}=\sum_{i=1}^{I}P_k(i)/\Lambda(i)-\sum_{i=1}^{I}R_k(T-t_i), \qquad k=1,\cdots,K,$$

$$\text{(2.10)} \quad \frac{\partial \log L_T}{\partial b_k}=\sum_{i=1}^{I}Q_k(i)/\Lambda(i)-\sum_{m=1}^{M}R_k(T-\tau_m), \qquad k=1,\cdots,L,$$

and

$$\text{(2.11)} \quad \frac{\partial \log L_T}{\partial c}=-\sum_{i=1}^{I}\left\{\sum_{k=1}^{K}a_kP_{k+1}(i)+\sum_{k=1}^{L}b_kQ_{k+1}(i)\right\}\Big/\Lambda(i) +\sum_{k=1}^{K}a_k\sum_{i=1}^{I}R_{k+1}(T-t_i)+\sum_{k=1}^{L}b_k\sum_{m=1}^{M}R_{k+1}(T-\tau_m),$$

where

$$\text{(2.12)} \quad \Lambda(i)=\mu+\sum_{k=1}^{K}a_kP_k(i)+\sum_{k=1}^{L}b_kQ_k(i).$$

If we set

$$\text{(2.13)} \quad \Lambda(i)=\mu+\sum_{k=1}^{K}a_kP_k(i)+\sum_{k=1}^{L}b_kU_k(i),$$

the gradients of $\log L_T^*$ for μ and a_k's are obtained by the same formula as (2.8) and (2.9), respectively, and

(2.14) $$\frac{\partial \log L_T^*}{\partial b_k}=\sum_{i=1}^{I} U_k(i)/\Lambda(i)-\sum_{m=1}^{M} x(\sigma_m)R_k(T-\sigma_m)\ , \qquad k=1,\cdots,L\ ,$$

and

(2.15) $$\frac{\partial \log L_T^*}{\partial c}=-\sum_{i=1}^{I}\left\{\sum_{k=1}^{K} a_k P_{k+1}(i)+\sum_{k=1}^{L} b_k Q_{k+1}(i)\right\}\Big/\Lambda(i) + \sum_{k=1}^{K} a_k \sum_{i=1}^{I} R_{k+1}(T-t_i)+\sum_{k=1}^{L} b_k \sum_{m=1}^{M} x(\sigma_m)R_{k+1}(T-\sigma_m)\ .$$

For the statistics P, Q, U and R the following recursive relations, which are useful for the efficient calculation of the likelihood, are available:

(2.16) $$P_k(i+1)=(t_{i+1}-t_i)^{k-1}e^{-c(t_{i+1}-t_i)} + \sum_{j=1}^{k}\binom{k-1}{j-1}P_j(i)(t_{i+1}-t_i)^{k-j}e^{-c(t_{i+1}-t_i)}\ ,$$

(2.17) $$Q_k(i+1)=D_k(t_i,\ t_{i+1})+\sum_{j=1}^{k}\binom{k-1}{j-1}Q_j(i)(t_{i+1}-t_i)^{k-j}e^{-c(t_{i+1}-t_i)}\ ,$$

where

$$D_k(t_i,\ t_{i+1})=\sum_{t_i\leqq\tau_m<t_{i+1}}(t_{i+1}-\tau_m)^{k-1}e^{-c(t_{i+1}-\tau_m)}\ .$$

Also

(2.18) $$U_k(i+1)=F_k(t_i,\ t_{i+1})+\sum_{j=1}^{k}\binom{k-1}{j-1}U_j(i)(t_{i+1}-t_i)^{k-j}e^{-c(t_{i+1}-t_i)}\ ,$$

where

$$F_k(t_i,\ t_{i+1})=\sum_{t_i\leqq\sigma_m<t_{i+1}}x(\sigma_m)(t_{i+1}-\sigma_m)^{k-1}e^{-c(t_{i+1}-\sigma_m)}\ .$$

Finally,

(2.19) $$R_k(t)=\{(k-1)R_{k-1}(t)-t^{k-1}e^{-ct}\}/c\ .$$

The AIC defined by (1.4) for the order determination is given by

(2.20) $$\mathrm{AIC}\,(K,\ L)=(-2)\max_{\theta}\log L_T(\theta)+2(K+L+2)\ ,$$

where θ stands for $(\mu,\ c,\ a_1,\cdots,\ a_K,\ b_1,\cdots,\ b_L)$. We choose $(K,\ L)$ that minimizes AIC. A very practical computationally efficient procedure is realized by restricting the exponential coefficient c to some finite number of candidate values c_j and minimizing

(2.21) $$\mathrm{AIC}\,(c_j;\ K,\ L)=(-2)\max_{\zeta}\log L_T(c_j;\ \zeta)+2(K+L+1)\ ,$$

where ζ stands for $(\mu,\ a_1,\cdots,\ a_K,\ b_1,\cdots,\ b_L)$. We choose the triplet $(c_j,$

K, L) that minimizes AIC $(c_j; K, L)$.

3. Relation to the second order analysis

The problems treated in this paper are related to the spectral analysis of input-output systems of point processes studied by Brillinger [2]. In the discussion of this paper, Cox suggested that an alternative approach would be a likelihood analysis based on models where the intensity function is modulated by the input process. Our present paper is exactly in this line of approach. Given the response functions $g(t)$ and $h(t)$, we can obtain the auto- and cross-covariance functions, $\mu_{11}(t)$ and $\mu_{12}(t)$ for $t \geqq 0$, numerically by using the recursive relation of Hawkes [5]

$$\begin{aligned}\mu_{11}(t) &= \lambda_1 g(t) + \int_0^t g(t-v)\mu_{11}(v)dv + \int_0^\infty g(t+v)\mu_{11}(v)dv \\ &\quad + \int_0^\infty h(t+v)\mu_{12}(v)dv \\ \mu_{12}(t) &= \lambda_2 h(t) + \int_0^t g(t-v)\mu_{12}(v)dv + \int_0^t h(t-v)\mu_{22}(v)dv \\ &\quad + \int_0^\infty h(t+v)\mu_{22}(v)dv\,,\end{aligned} \tag{3.1}$$

where $\mu_{22}(v)$ is the auto-covariance of the input process. In the next section we will use this relation to get initial guesses of the parameters. We will also use the relation for the comparison of the parametric and non-parametric estimates of the covariance functions.

4. Illustration by artificial examples

We wish to illustrate the procedure described in the previous section by some artificial examples. In the first example the input process $\{x(\sigma_m)\}$ was generated by the relation $x(\sigma_m) = \exp\{0.3Y_m\}$, where $\sigma_m = m$ and $Y_m = 1.394Y_{m-1} - 0.752Y_{m-2} + \varepsilon_m$ and ε_m is a standard normal white noise. The record of $\{x(\sigma_m)\}$ is shown in Fig. 1. The true response functions are given by (1.2) defined with the parameters given in Table 1. We obtained a series of events with the total number of occurrences $N_T = 1514$ in the time interval $[0, T] = [0, 1000.0]$ (see [6] for the method of generation). Models with $L = K$ and up to 6th order were fitted. The maximum of the log likelihood for each order was obtained as follows: As the initial guess of the parameters for the case with $L = K = 1$ we used $(\mu, c, a_0, b_0) = (\hat{\mu}, \hat{c}, 0, 0)$, where $\hat{\mu} = \hat{c} = N_T/T = 1.514$. The case with $L = K = 2$ was started with the initial guess

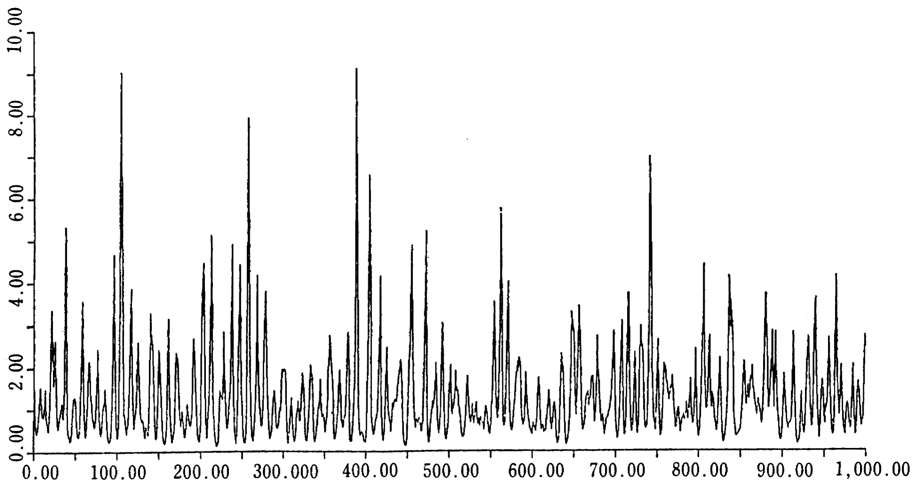

Fig. 1. The record of input process

Table 1

	μ	c	a_1	a_2	a_3	a_4	a_5	b_1	b_2	b_3	b_4	b_5
true	0.10	1.00	0.23	−1.00	1.74	−0.83	0.12	0.00	0.49	−0.36	0.08	0.02
m.l.e.	0.04	0.99	0.20	−0.84	1.29	−0.55	0.07	−0.03	−0.09	0.86	−0.55	0.11

$(\mu, c, a_0, b_0, a_1, b_1)=(\hat{\mu}, \hat{c}, \hat{a}_0, \hat{b}_0, 0, 0)$, where $\hat{\mu}$, $\hat{c}$, $\hat{a}_0$ and $\hat{b}_0$ were the maximum likelihood estimates of the case with $L=K=1$. A similar procedure was repeated up to the 6th order model ($K=L=7$). The values of AIC in the sense of (2.20) are listed in Table 3 which shows that the model with $L=K=5$ is the best. The maximum likelihood estimates of the corresponding coefficients are given in Table 1. The true

Table 2

	μ_1	c	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
true	0.08	5.00	0.69	−12.23	97.13	−289.65	338.69	−65.73	−111.68	50.79
m.l.e.	0.08	4.00	0.61	−10.21	76.13	−215.89	272.90	−153.23	32.53	—
			b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8
true			0.89	−13.28	75.43	−170.56	90.29	220.64	−258.23	74.26
m.l.e.			0.93	−12.85	72.53	−202.57	294.36	−186.53	41.41	—

Table 3

k	1	2	3	4	5	6	7
AIC (k, k)	1635.18	1588.26	1588.49	1587.76	1585.29*	1589.13	1592.81
estimated c	0.15	0.31	0.49	0.76	0.99	1.00	1.06

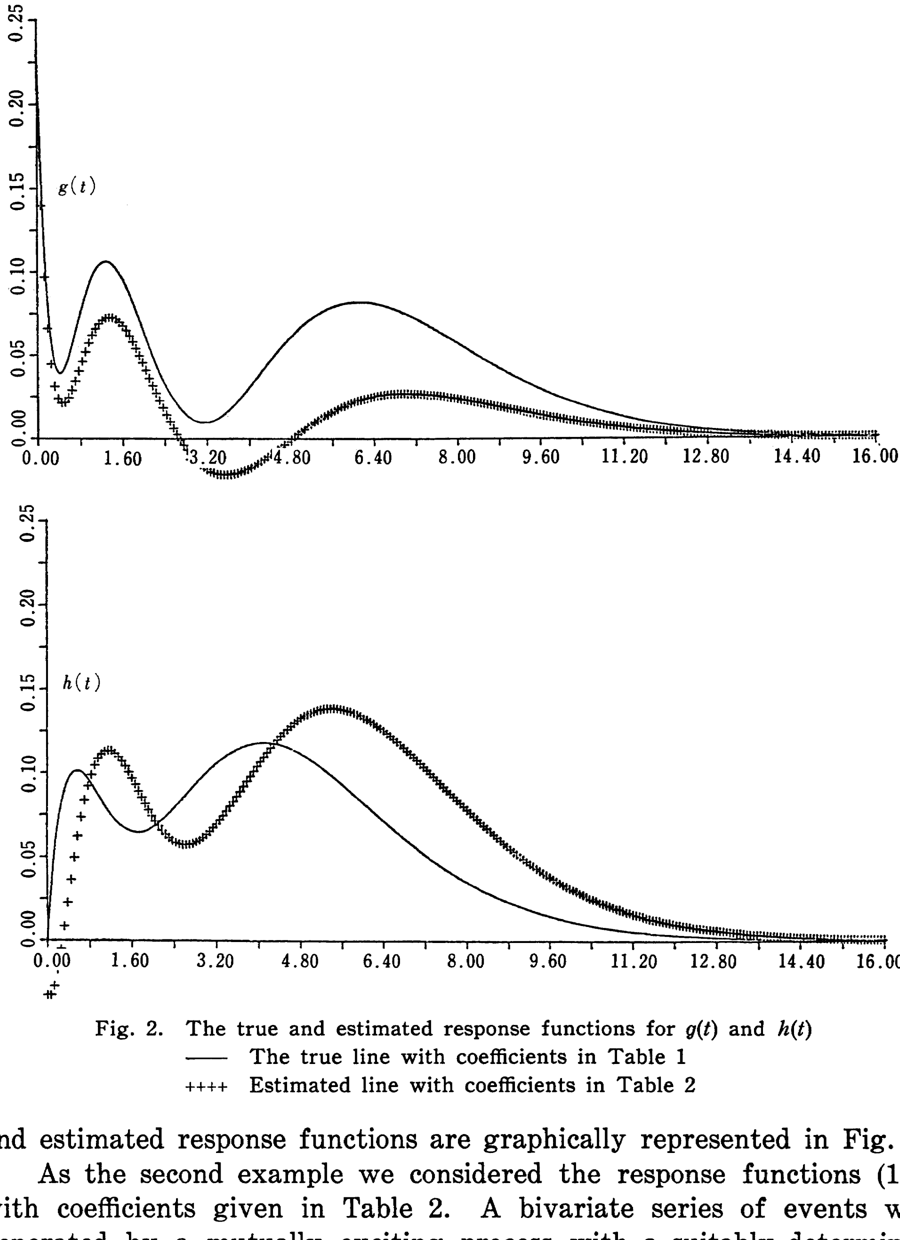

Fig. 2. The true and estimated response functions for $g(t)$ and $h(t)$
—— The true line with coefficients in Table 1
++++ Estimated line with coefficients in Table 2

and estimated response functions are graphically represented in Fig. 2.

As the second example we considered the response functions (1.2) with coefficients given in Table 2. A bivariate series of events was generated by a mutually exciting process with a suitably determined intensity for the second component (see [6] for the algorithm of the generation). The input and output series were with the total numbers of occurrences $I=1165$ and $M=1010$, respectively. They were obtained on the time interval [0, 10000.0]. For the initial guess of the exponential coefficient c the auto- and cross-correlograms were obtained (see Fig. 4). Significant values of the correlations are observed up to the time lag 4.0. The relation (3.1) suggests that the shapes of the re-

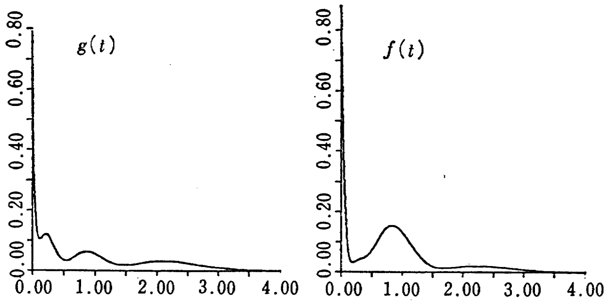

Fig. 3. The true response functions with coefficients in Table 2

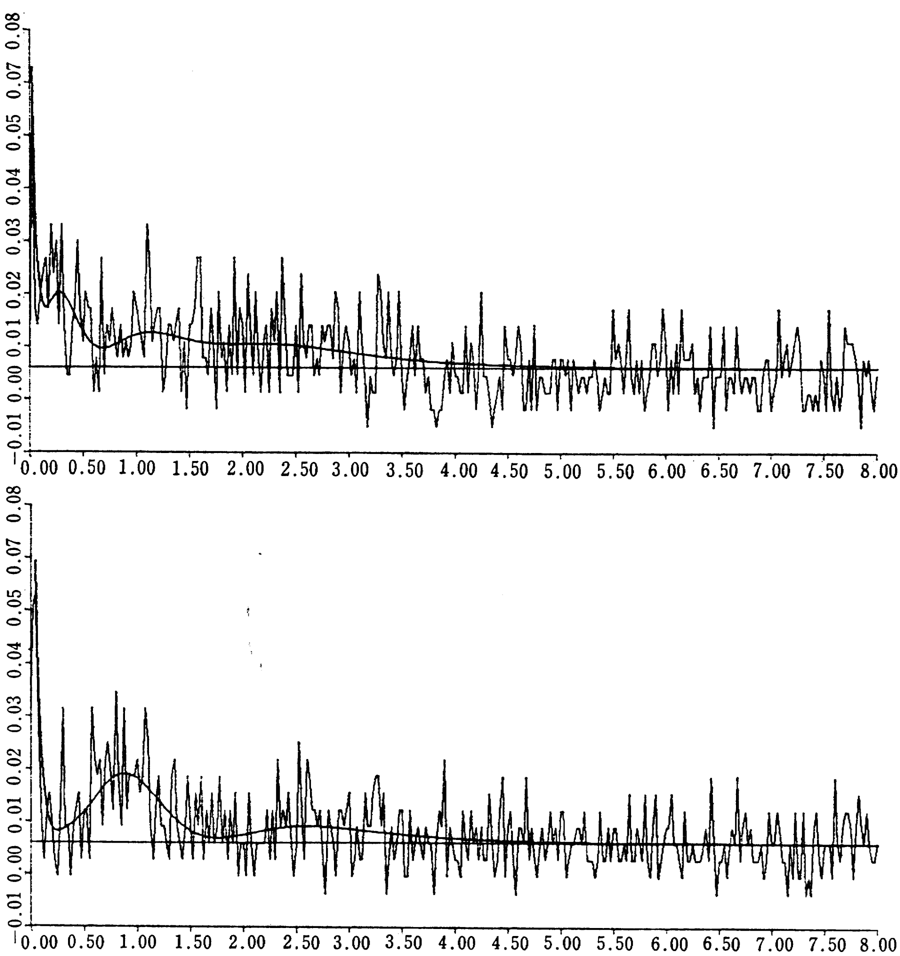

Fig. 4. Estimations of auto- and cross-covariance functions

The smooth lines are obtained numerically by solving the equation (3.1) with estimated parameters in the Table 2.

sponse and covariance functions are not very much different. Thus we guessed that the time range R of the significant response would be around half of this time lag 4.0, i.e., $R=2.0$. By assuming the model with $K=3$ we obtained a rough initial guess $c_0=4/R=2.0$. Alternatively, $c_0=0.98$ was obtained by the direct maximum likelihood estimation of the 0th order model, starting with the initial guess $(\mu, c, a_0, b_0)=(0.1, 0.1, 0.0, 0.0)$. Using $c_0=2.0$ as our initial guess of the exponential coefficient we successively tried $c=0.5$, 1.0, 4.0, 8.0 and 6.0. The AIC values defined by (2.21) were obtained for the models up to the 14th order for each c. They are listed in Table 4. The minimum AIC value is attained at $c=4.0$ and $L=K=7$. The estimated coefficients for this case are given in Table 2. Fig. 5 shows the graphs of the estimated response functions with the minimum AIC for each choice of c. It can be seen that the graphs for $c=4.0$ which gives the overall minimum of the AIC's are very close to those of the true response functions (Fig. 3). The covariance functions calculated by the relation (3.1) show good fit to the corresponding non-parametrically estimated covariance functions (Fig. 4).

Table 4 List of AIC $(c_j; K, K)$

K \ c_j	0.5	1.0	2.0	4.0	6.0	8.0
0 (Poisson)	7465.16	7465.16	7465.16	7465.16	7465.16	7465.16
1	7199.38	7173.59	7198.15	7237.29	7252.05	7259.12
2	7177.50*	7177.51	7188.97	7231.10	7250.01	7254.38
3	7181.10	7180.47	7156.68	7137.24	7168.77	7202.06
4	7184.08	7169.53	7149.49	7133.53	7154.44	7181.38
5	7184.65	7168.95	7128.70	7122.85	7126.22	7140.83
6	7187.10	7168.22	7118.62*	7126.15	7127.21	7135.07
7	7180.06	7166.09	7122.35	7112.87**	7122.41	7127.84
8	7178.65	7144.72	7126.20	7112.92	7119.08	7126.51
9	7182.21	7137.27	7130.14	7116.31	7116.79	7125.00
10	7185.23	7139.54	7129.98	7117.02	7112.99*	7118.11*
11	7186.23	7132.61*	7126.95	7120.75	7116.65	7122.11
12	7189.77	7133.12	7127.16	7122.77	7119.70	7123.87
13	7193.54	7136.67	7128.54	7121.53	7115.22	7122.20
14	7191.76	7140.56	7124.15	7123.47	7119.22	7126.20
15	7195.76	7144.32	7127.90	7127.30	7122.92	7128.64

* shows the minimum AIC for each c_j and ** shows the overall minimum AIC.

5. Analysis of causality of earthquake data

Utsu [9] discussed the correlation between the intermediate earthquakes in Hida (the region around Takayama city) and the shallower

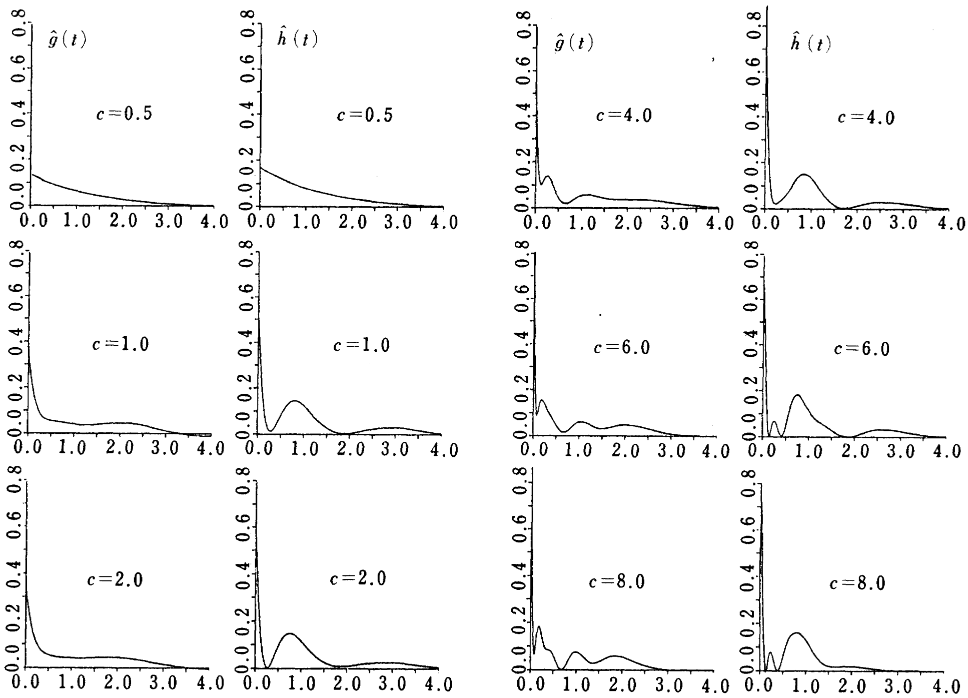

Fig. 5. Estimated response functions with the minimum AIC's

earthquakes in the central Kwanto (the region around Tokyo). He tested the independence between these earthquakes, and concluded that there was a significant dependence which could be attributed to some mechanical connection between the two seismic regions.

Utsu's data is composed of 61 earthquakes with Richter magnitude $M \geqq 5.5$ in central Kwanto and 16 earthquakes of $M \geqq 5.0$ in Hida during the 51 years from 1924 through 1974. Here the data is reproduced after the transformation into i-day, i.e., the time scale unit is one-day with the origin 0 of the time being equated to the 1st of January, 1924 (see Table 6). The model (1.1) was fitted to both sets of earthquake data. The values of AIC of the models given by (2.20) with the earthquake series in the central Kwanto area as the output and the series in the Hida area as the input are listed in Table 7. The minimum AIC is attained at $L=1$ and $M=1$. The corresponding estimates of the parameters are listed in Table 5. The graphs of the estimated response functions are given in Fig. 6. The result clearly

Table 5

	μ	c	a_1	b_1
m.l.e.	1.42×10^{-3} (shocks/day)	6.33×10^{-3} (1/day)	1.01×10^{-3} (shocks/day)	8.66×10^{-3} (shocks/day)

Table 6 List of intermediate earthquakes ($M \geqq 5.0$) in the Hida region and shallow earthquakes ($M \geqq 5.5$) in the central Kwanto region

			Central Kwanto			
1109	1272	1313	1356	1458	1469	1484
2172	2556	2598	2697	2834	3129	3813
3819	3842	3910	3915	3922	3927	3967
5163	5385	5968	6246	6365	6938	7135
7419	8054	8054	8216	8326	8567	8763
8770	9062	9160	10965	11263	11444	11450
12069	12108	12208	12434	12622	12827	12899
13056	13091	14257	15011	16097	16166	16221
16878	17348	19265	19266	19573		
			Hida			
1443	2505	3804	5217	6675	8218	11297
11746	11866	12187	12661	12753	14521	15100
16150	19219					

Time scale unit is one-day.

shows that earthquakes in Hida area do stimulate the occurrence of earthquakes in Kwanto area.

The graph of the estimated intensity process of the Kwanto earthquakes is given in Fig. 7 where the symbols K and H indicate the occurrence times of the earthquakes in Kwanto and Hida area, respectively. Thus it seems that Hida earthquakes will play a significant role in earthquake risk prediction (see Vere-Jones [10]) in the Kwanto area.

To see if a similar effect exists in the opposite direction, the values of AIC of the models with the series in the Hida area as the output and that of Kwanto area as the input were calculated. However, for some choices of the orders (K, L) the maximum log likelihood diverged and the estimated intensity took significantly negative values. In particular either $g(0)$ or $h(0)$ was tending to minus infinity. To avoid this difficulty the parameters a_1 and b_1 were restricted to non-negative values. This also kept the estimated intensity process positive. This problem is discussed in [8]. The AIC values with * in Table 8 were obtained under these restrictions. The overall minimum of AIC was attained by the Poisson model with $g(t) \equiv h(t) \equiv 0$. This suggests that the earthquakes in the Kwanto area do not stimulate the occurrence of those in the Hida area. To check the validity of the obtained model, we performed simulations of earthquakes of the Kwanto area by using the estimated intensity function with the occurrence times for the Hida area given by the Utsu data of Table 6. One numerical result is given in Fig. 8; see [7] for the method of simulation. Fig. 8 exhibits a similarity with the result given in Fig. 7. We obtained the

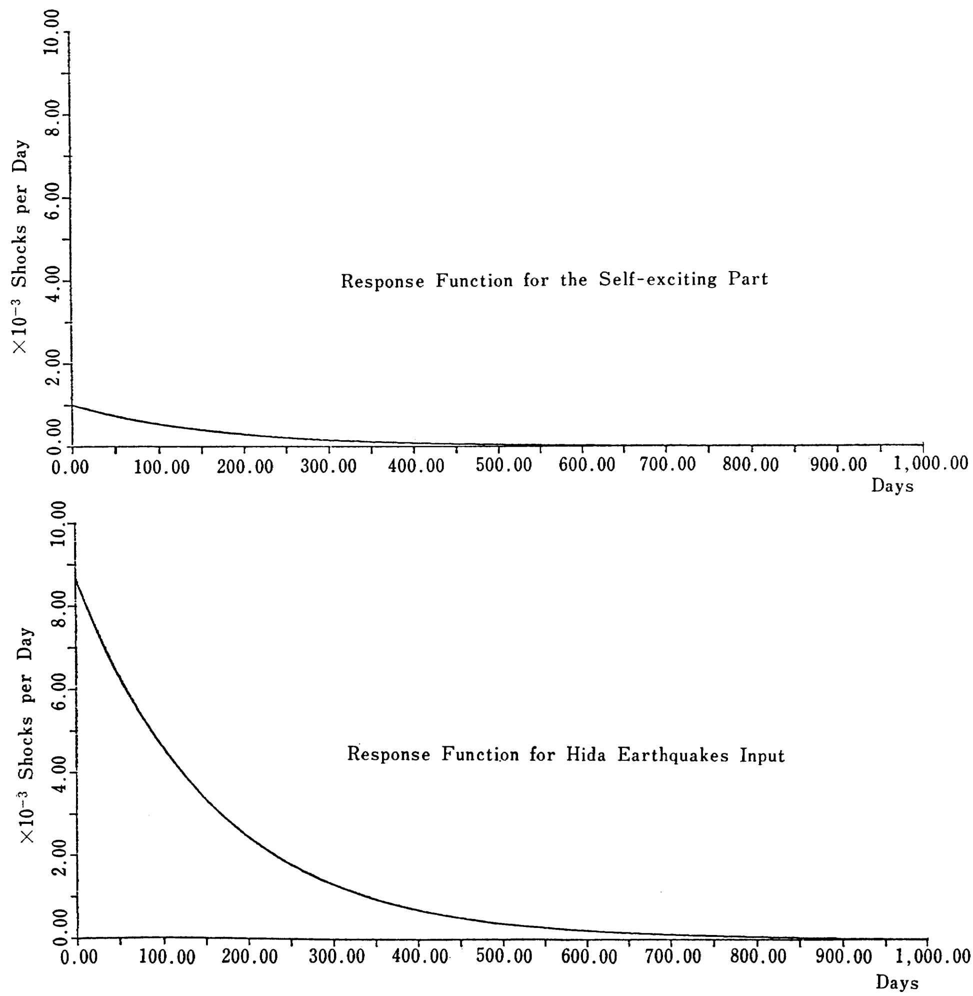

Fig. 6. Estimated response functions of Kwanto earthquakes with Hida earthquakes input

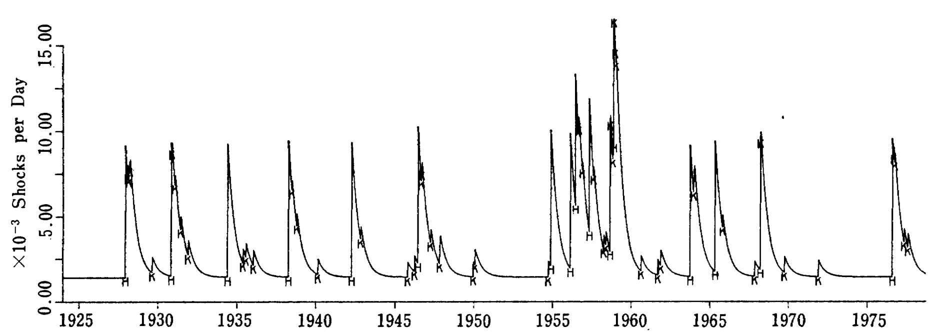

Fig. 7. Estimated intensity process of Kwanto earthquakes
The symbols K and H indicate the occurrence times of earthquakes in Kwanto and Hida area, respectively.

Table 7

Input=Hida data

	AIC (L, K)	Self-exciting	$L=1$	$L=2$	$L=3$	$L=4$
Output =Kwanto	Poisson	−12.0	−33.0	−31.6	−29.6	−27.7
	$K=1$	−20.8	−33.6	−31.9	−30.1	−28.0
	$K=2$	−18.8	−32.7	−30.7	−28.8	−26.8
	$K=3$	−20.1	−30.7	−28.7	−27.3	−25.7
	$K=4$	−18.2	−28.7	−26.7	−25.1	−23.4

The underlines show the minimum AIC.

Table 8

Input=Kwanto data

	AIC (L, K)	Self-exciting	$L=1$	$L=2$	$L=3$	$L=4$
Output=Hida	Poisson	41.1	43.6	45.4	47.4	49.1
	$K=1$	44.6	45.6*	47.4	49.2	51.2
	$K=2$	46.3	47.6*	48.5*	50.1*	52.1*
	$K=3$	47.9*	49.6*	51.9*	50.6*	52.1*
	$K=4$	44.2	51.6*	53.9	51.9*	49.2*

The underlines show the minimum AIC, and
* shows that AIC was obtained under the restriction $g(0)\geqq 0$ or $h(0)\geqq 0$.

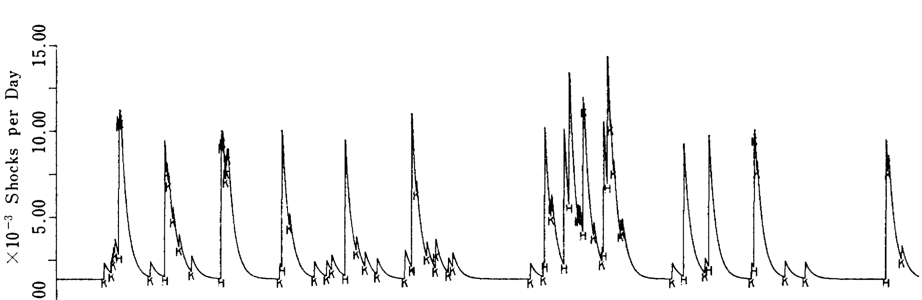

Fig. 8. A simulated example of earthquakes with the parameters in Table 5

cumulative distributions of the interval lengths of the series of Figs. 7 and 8 and checked the observation of Utsu [9] that the cumulative distribution displays a broken line type behavior. The behavior can be observed in both Figs. 9 and 10. This suggests that our estimated model is reproducing a basic characteristic of the Utsu data. The broken line phenomenon might be caused by the significant change of intensity after a Hida earthquake.

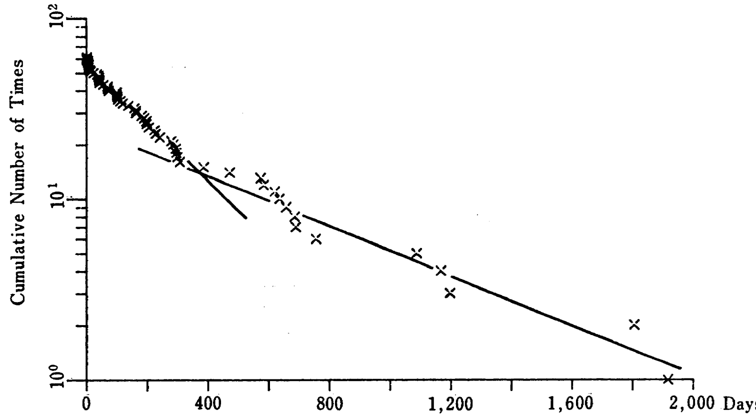

Fig. 9.

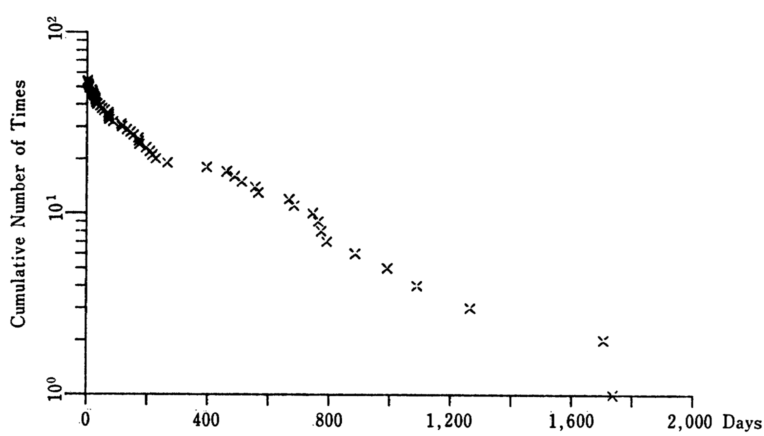

Fig. 10.

6. Concluding remarks

Conventionally the model selection is realized by successively applying the likelihood ratio test. The relationship between the AIC and the likelihood ratio statistic is given by

$$(6.1) \qquad (-2)\log \lambda(H_0; H_1) = \mathrm{AIC}\,(H_0) - \mathrm{AIC}\,(H_1) - 2k$$

where the model H_1 contains the model H_0 as a restricted family of distributions of H_1 and k denotes the degrees of freedom of the chi-square distribution of the likelihood ratio test statistic. Owing to the difficulty in selecting appropriate significance levels, we did not follow this conventional approach. However, if the reader is interested in the testing procedure the AIC's can be translated into the log likelihood ratios by (6.1). For example, to test the causal relation in the earthquake data, the AIC's given in Table 7 and the relation (6.1) provide necessary information.

For maximum likelihood computation, the selection of the exponential coefficient of the response function is important. By our experience of simulation study, too large or too small values of the exponential coefficient c cause the increase of the order of the model with minimum AIC. However, it may happen that the optimum c is not unique or that the likelihood increases indefinitely as $c \to 0$. In these cases we may use prior information to select an appropriate finite interval for c, since c is, roughly speaking, a scale parameter of the influential range of the response function. For example, if we consider an earthquake series, useful prior information for the interval could be obtained from some seismological finding such as Omori's law [11].

Acknowledgement

The authors are grateful to the referees for the helpful comments on the organization of the present paper.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Contr.*, AC-**19**, 716-723.

[2] Brillinger, D. (1975). The identification of point process systems, *Ann. Prob.*, **3**, 909-929.

[3] Cox, D. R. (1975). Partial likelihood, *Biometrika*, **62**, 269-276.

[4] Fletcher, R. and Powell, M. J. D. (1963). A rapidly convergent descent method for minimization, *Computer J.*, **6**, 163-168.

[5] Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes, *Biometrika*, **58**, 83-90.

[6] Ogata, Y. (1981). On Lewis' simulation method for point processes, *IEEE Trans. Inform. Theory*, IT-**27**, 23-31.

[7] Ogata, Y. and Akaike, H. (1982). On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes, *J. R. Statist. Soc.*, B, **44**, Part 1, to appear.

[8] Ogata, Y. and Katsura, K. (1981). Statistical identification for series of events which possesses the self-inhibitory property, *Research Memorandum*, 202, Institute of Statistical Mathematics, Tokyo.

[9] Utsu, T. (1975). Correlation between shallow earthquakes in Kwanto region and intermediate earthquakes in Hida region, central Japan, *Zisin* (*J. Seism. Soc. Japan*), 2nd Ser., **28**, 303-311 (in Japanese).

[10] Vere-Jones, D. (1978). Earthquake prediction—A statistician's view, *J. Phys. Earth*, **26**, 129-146.

[11] Omori, F. (1894). On the aftershocks of earthquake, *J. Coll. Sci. Imp. Univ. Tokyo*, **7**, 111-200.