A note on the consistent estimation of spatial-temporal point process parameters.

UCLA Department of Statistics, 8125 Math-Science Building, Los Angeles, CA 90095-1554.

Supplementary Material

S1 Proof of Lemma 3.1.

When λ is separable in mark m_i, (3) becomes

$$\begin{aligned} L(\theta_i, \theta_{k+1}) &= N(D)\log(\theta_{k+1}) + [\int_D \log\lambda_i(t, m_i; \theta_i) + \log\lambda_{-i}(t, m_{-i}; \theta_{-i})]dN \\ &- \theta_{k+1} \int\limits_{D_0}\int\limits_{D_i} \lambda_i(t, m_i; \theta_i) d\mu_i \int\limits_{D_{-i}} \lambda(t, m_{-i}; \theta_{-i}) d\mu_{-i} dt. \end{aligned}$$

Hence

$$\begin{aligned} 0 &= \frac{\partial L(\theta)}{\partial\theta_i} \\ &= \frac{\partial}{\partial\theta_i}\int_D \log\lambda_i(t, m_i; \theta_i)dN - \theta_{k+1}\frac{\partial}{\partial\theta_i}\int_{D_0}\int_{D_i}\lambda_i(t, m_i; \theta_i)\left[\int_{D_{-i}}\lambda_{-i}(t, m_{-i}; \theta_{-i})d\mu_{-i}\right]d\mu_i dt. \end{aligned}$$

By assumption, $(\tilde\theta_{k+1}, \tilde\theta_i)$ is the unique solution to the pair of equations $\frac{\partial\tilde L_i}{\partial\theta_i} = 0$ and $\frac{\partial\tilde L_i}{\partial\theta_{k+1}} = 0$, and thus satisfies

$$\begin{aligned} 0 &= \frac{\partial\tilde L_i}{\partial\theta_i} \\ &= \frac{\partial}{\partial\theta_i}\int_D \log\lambda_i(t, m_i; \theta_i)dN - \theta_{k+1}\int_{D_0}\int_{D_i}\frac{\partial\lambda_i(t, m_i; \theta_i)}{\partial\theta_i}d\mu_i dt, \end{aligned}$$

which, under condition (5), has the unique solution $\left(\hat\theta_{k+1}\gamma, \hat\theta_i\right)$. If (6) or (7) holds, then neither $\int_{D_0}\int_{D_i}\lambda_i(t, m_i; \theta_i)d\mu_i dt$ nor $\int_{D_0}\int_{D_i}\lambda_i(t, m_i; \theta_i)\left[\int_{D_{-i}}\lambda_{-i}(t, m_{-i}; \theta_{-i})d\mu_{-i}\right]d\mu_i dt$ depends on θ_i, so both $\tilde\theta_i$ and the MLE $\hat\theta_i$ must uniquely satisfy $\frac{\partial}{\partial\theta_i}\int_D \log\lambda_i(t, m_i; \theta_i)dN = 0$. □

S2 Proof of Theorem 3.2.

Reparameterizing the second term in $\tilde{L}_i$ and dropping the subscripts t and m_i for simplicity, one may write

$$\begin{aligned}\theta_{k+1}\int_{D_0}\int_{D_i}\lambda_i(t,m_i;\theta_i)d\mu_i dt &= \theta_{k+1}\int_x\int_y f_1(x;\beta_1)f_2(y;\beta_2)dH(x,y)\\ &= \theta_{k+1}\int_x f_1(x;\beta_1)dH_1(x)\int_y f_2(y;\beta_2)dH_2(y).\end{aligned}$$

Hence $\tilde{\beta}_1$ satisfies

$$\begin{aligned}0 &= \frac{\partial}{\partial\beta_1}L(\tilde{\theta}_i)\\ &= \frac{\partial}{\partial\beta_1}\int_{D_0}\int_{D_i}\log f_1(X(t,m_i);\tilde{\beta}_1)dN(t,m_i) - \theta_{k+1}\int_y f_2(y;\beta_2)dH_2(y)\frac{\partial}{\partial\beta_1}\int_x f_1(x;\beta_1)dH_1(x).\end{aligned}$$

One may similarly reparameterize $\check{L}(\beta_1)$ to obtain

$$\frac{\partial}{\partial\beta_1}\check{L}(\beta_1) = \frac{\partial}{\partial\beta_1}\int_{D_0}\int_{D_i}\log f_1(X(t,m_i);\beta_1)dN(t,m_i) - \theta_{k+1}\frac{\partial}{\partial\beta_1}\int_x f_1(x;\beta_1)dH_1(x).$$

Thus $(\theta_{k+1}\int_y f_2(y;\beta_2)dH_2(y), \tilde{\beta}_1)$ is the unique solution to the equation $\frac{\partial}{\partial\beta_1}\check{L}(\beta) = 0$. Therefore, using Lemma 3.1, $\hat{\beta}_1 = \tilde{\beta}_1 = \check{\beta}_1$. □

S3 Counterexample to Theorem 3.2.

Ogata (1988) fit a purely temporal-magnitude version of ETAS, which was extended to space-time-magnitude in Ogata (1998). Specifically, Ogata (1988) considered models such as

$$\lambda(t,x,y) = \mu + \sum_{i=1}^{n}\exp\{\beta(M_i - M_0)\}K(t - t_i + c)^{-p},$$

and Ogata (1998) considered spatial-temporal extensions including

$$\lambda(t,x,y) = \mu + \sum_{i=1}^{n}\exp\{\beta(M_i - M_0)\}K(t - t_i + c)^{-p}\{(x - x_i)^2 + (y - y_i)^2 + d)^{-q},$$

where $M_i$and t_i are the magnitude and time, respectively, of earthquake i, and M_0 is the lower magnitude cutoff for the catalog. The functional form of the portions of these models governing the temporal clustering and dependence on magnitude are identical, and both were fit by Ogata to a catalog of shallow $M \geq 6.0$ earthquakes off Tohuku, Japan, with only slight differences in the catalog yet substantially differing estimates of these parameters. For example, Ogata

(1988) estimates $\hat{c}$ as 0.01959 days, while Ogata (1998) estimates the same parameter, using the spatial-temporal-magnitude model, as 0.00977 days. Although the only modification in the model is the inclusion of a multiplicative spatial distribution of aftershocks in the triggering function, the spatial term is not separable since the entire function λ is not multiplicative in x and y, and since the spatial information is extremely relevant to the earthquake generation process, it is not surprising that the inclusion of this information results in substantial changes to the parameters governing the temporal and magnitude behavior of the process.

S4 Conditions for Theorem 4.1.

Ogata (1978) gives general conditions under which the MLE of the parameter vector θ governing a stationary point process N is proven to be consistent. The conditions are as follows.

- N is stationary, ergodic, and absolutely continuous with respect to the standard Poisson process on any finite interval.
- The parameter space Θ is a compact metric space and is a subset of $\mathbf{R}^d$.
- $\lambda(0, \omega; \theta_1) = \lambda(0, \omega; \theta_2)$ a.s. if and only if $\theta_1 = \theta_2$.
- $\partial \log \lambda / \partial \theta_i, \partial^2 \log \lambda / \partial \theta_i \partial \theta_j$ and $\partial^3 \log \lambda / \partial \theta_i \partial \theta_j \partial \theta_k$ exist and are continuous in θ for all $i, j, k = 1, 2, ..., d$, for all $t \in \mathbf{R}^+$ almost surely, and $\partial \lambda / \partial \theta_i$ and $\partial^2 \lambda / \partial \theta_i \partial \theta_j$ have finite second moments for any $\theta \in \Theta$.
- For any $\theta \in \Theta$, there is a neighborhood U of θ such that for all $\theta' \in U$, $|\lambda(o, \omega; \theta')| \leq \Lambda_0(\omega)$ and $|\log \lambda(0, \omega; \theta')| \leq \Lambda_1(\omega)$, where Λ_0 and Λ_1 are random variables with finite second moments.
- For any $\theta \in \Theta$, there is a neighborhood U of θ such that $\sup\limits_{\theta' \in U} |\log \lambda_{\theta'}(t, \omega)|$ has finite $(2 + \alpha)$th moment uniformly bounded with respect to t for some $\alpha > 0$.

Note that the last condition is somewhat different in Ogata (1978), which also assumes the conditional intensity, conditioning on the history since time $-\infty$, converges to the conditional intensity since time 0. Here, we assume in Section 2 the point process is only defined since time 0 so this last condition is simplified considerably.

S5 Proof of Theorem 4.1.

To prove Theorem 4.1 formally, first note that by martingale convergence (see e.g. Theorem A3.4.iii of Daley and Vere-Jones 1988 or 3.3 of Lipster and Shiryaev 1977), for any $\theta_i = (\beta_1, \beta_2)$,

$$\frac{1}{T}\left[\int_0^T \int_{D_i} \log \lambda(t, m_i; \theta_i) dN(t, m_i) - \int_0^T \int_{D_i} \log \lambda(t, m_i; \theta_i) \lambda(t, m_i; \theta_i^*) d\mu_i dt\right] \longrightarrow 0 \, a.s.$$

and

$$\frac{1}{T}\left[\int_0^T\int_{D_i}\log\left[f_1(X(t,m_i);\beta_1)\right]dN(t,m_i)-\int_0^T\int_{D_i}\log\left[f_1(X(t,m_i;\beta_1)\right]\lambda(t,m_i;\theta_i^*)d\mu_i dt\right]\longrightarrow 0\,a.s.$$

Thus we can write

$$\begin{aligned}\frac{\tilde{L}_i^{(T)}(\theta_i)}{T} &= \frac{1}{T}\int_0^T\int_{D_i}\log\left[f_1(X(t,m_i);\beta_1)+f_2(Y(t,m_i);\beta_2)\right]dN(m_i,t)\\ &\quad -\frac{1}{T}\int_0^T\int_{D_i}\left[f_1(X(t,m_i);\beta_1)+f_2(Y(t,m_i);\beta_2)\right]d\mu_i dt\\ &\sim \frac{1}{T}\int_0^T\int_{D_i}\log\left[f_1(X(t,m_i);\beta_1)+f_2(Y(t,m_i);\beta_2)\right]\lambda(t,m_i;\theta_i^*)d\mu_i dt\\ &\quad -\frac{1}{T}\int_0^T\int_{D_i}\left[f_1(X(t,m_i);\beta_1)+f_2(Y(t,m_i);\beta_2)\right]d\mu_i dt,\end{aligned}$$

where by $a\sim b$ we mean that $a-b$ converges to zero a.s. as $T\to\infty$.

Similarly,

$$\begin{aligned}\frac{\acute{L}^{(T)}(\beta_1)}{T} &= \frac{1}{T}\int_0^T\int_{D_i}\log[f_1(X(t,m_i);\beta_1)]dN(t,m_i)-\frac{1}{T}\int_0^T\int_{D_i}f_1(X(t,m_i);\beta_1)d\mu_i dt\\ &\sim \frac{1}{T}\int_0^T\int_{D_i}\log[f_1(X(t,m_i);\beta_1)]\lambda(t,m_i;\theta_i^*)d\mu_i dt-\frac{1}{T}\int_0^T\int_{D_i}f_1(X(t,m_i);\beta_1)d\mu_i dt.\end{aligned}$$

Hence, for $\theta\in U$, $\frac{\tilde{L}_i^{(T)}(\theta_i)}{T}-\frac{\acute{L}^{(T)}(\beta_1)}{T}=$

$$\frac{1}{T}\int_0^T\int_{D_i}\lambda(t,m_i;\theta_i^*)\left[\log\left(f_1(X(t,m_i);\beta_1)+f_2(Y(t,m_i);\beta_2)\right)-\log\left(f_1(X(t,m_i);\beta_1)\right)\right]d\mu_i dt$$
$$-\frac{1}{T}\int_0^T\int_{D_i}f_2(Y(t,m_i);\beta_2)d\mu_i dt+o(T).$$

But by assumption, $\frac{1}{T}\int_0^T\int_{D_i}f_2(Y(t,m_i);\beta_2)d\mu_i dt\longrightarrow 0$ in probability. Furthermore, abbreviating $f_1(X(t,m_i);\beta_1)$ and $f_2(Y(t,m_i);\beta_2)$ to f_1 and f_2, respectively, for the moment,

$$\log(f_1+f_2)-\log(f_1)=\log(\frac{f_1+f_2}{f_1})\leq\frac{f_1+f_2}{f_1}-1=\frac{f_2}{f_1},$$

using the well-known relation $\log(x)\leq x-1$, for positive x (see e.g. Abramowitz, 1964). Thus, since by assumption $\frac{1}{T}\int_0^T\int_{D_i}\lambda(t,m_i;\theta_i^*)f_2/f_1 d\mu_i dt$ converges to zero in probability, the same is true of $\tilde{L}_i^{(T)}(\theta_i)/T-\acute{L}^{(T)}(\beta_1)/T$ and this convergence is uniform in θ_i due to the continuity of

f_1 and f_2 and the compactness of Θ_i. Thus for any $\epsilon > 0$, $|\sup_{\theta \in U} \tilde{L}_i^{(T)}(\theta_i)/T - \sup_{\theta \in U} \acute{L}^{(T)}(\beta_1)/T|$ and $|\sup_{\theta \notin U} \tilde{L}_i^{(T)}(\theta_i)/T - \sup_{\theta \notin U} \acute{L}^{(T)}(\beta_1)/T|$ are each less than $\epsilon/2$ with probability going to one, as $T \to \infty$.

By Lemma 3.1, $\tilde{\beta_1} = \hat{\beta_1}$. By relation 3.6 of Ogata (1978), for any $\epsilon > 0$, there exists T_1 such that for $T > T_1$,

$$\sup_{\theta \in U} \tilde{L}_i^{(T)}(\theta_i) \geq \sup_{\theta \notin U} \tilde{L}_i^{T}(\theta_i) + \epsilon T.$$

Let B_1 be any neighborhood of the true parameter β_1^*. We may find $U_1 \subseteq U$, where U_1 is an open subset of U containing θ^* such that B_1 is the restriction of U_1 to the parameter space containing β_1. Then with probability going to one as $T \to \infty$,

$$\sup_{\beta_1 \in B_1} \frac{\acute{L}^{(T)}(\beta_1)}{T} - \sup_{\beta_1 \notin B_1} \frac{\acute{L}^{(T)}(\beta_1)}{T} > \sup_{\theta \in U_1} \frac{\tilde{L}_i^{(T)}(\theta_i)}{T} - \epsilon/2 - \sup_{\theta \notin U_1} \frac{\tilde{L}_i^{(T)}(\theta_i)}{T} - \epsilon/2 \geq 0.$$

Thus, with probability going to 1 as $T \to \infty$, $\acute{\beta} \in B_1$. □

S6 Means and RMSE of PMLEs.

Figure 1 shows the means and RMSE in PMLEs of parameters (μ, α, β) in model (18), using simulations of model (17). Figure 2 shows the means and RMSE in PMLEs of parameters (μ, α, β) in model (16), using simulations of model (19). Table 1 reports the RMSE of parameter estimates for various models simulated in Section 5. Figure 3 shows how ETAS parameters in (24) vary with catalog length, when fit to the data described in Section 6. Figure 4 shows how estimated ETAS parameters in models (24) and (25) vary with catalog length, when fit to the data in Section 6 and when scaled versus their final estimated values using the entire catalog.

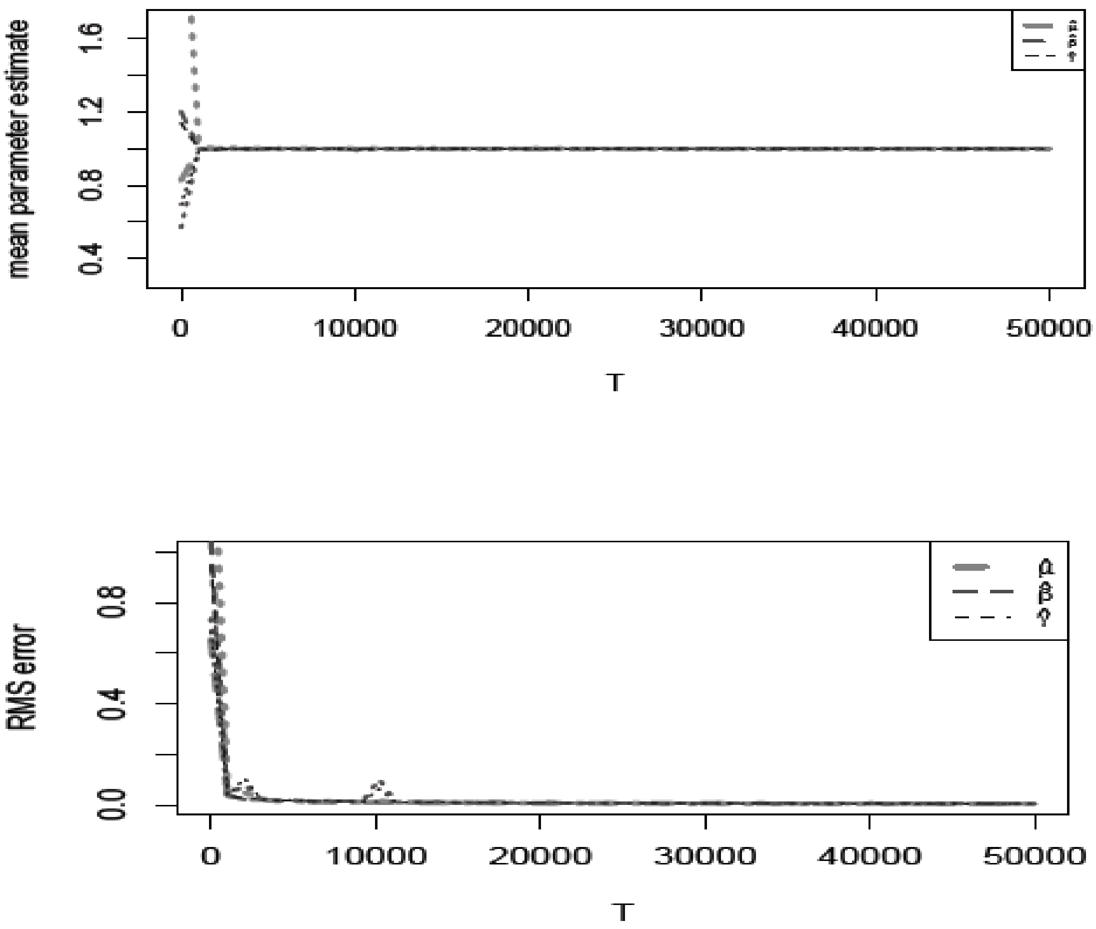

Figure 1: Means and RMSE in PMLEs of parameters (μ, α, β) in model (18), using simulations of model (17) with $(\mu, \alpha, \beta) = (1, 1, 1)$. 100 simulations were performed for each T, for 50 equally spaced values of T between 1 and $50,000$.

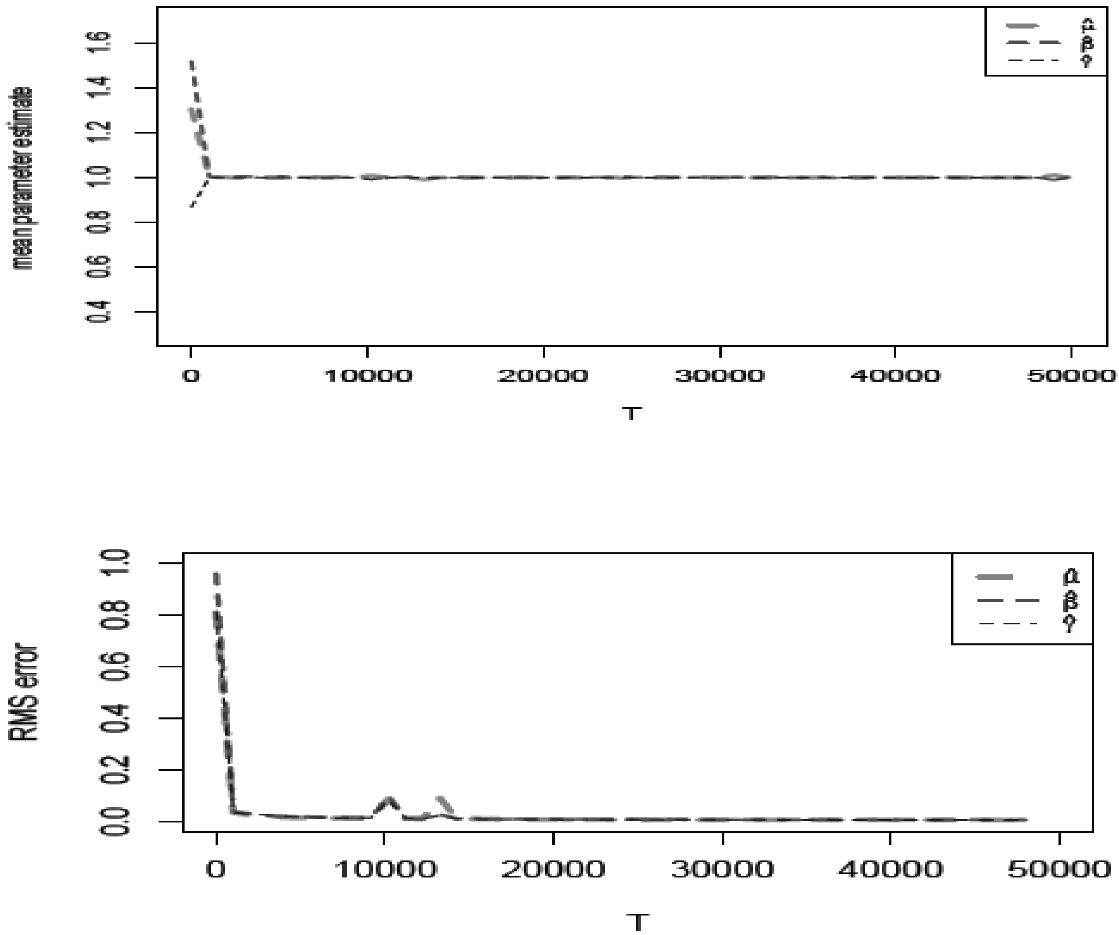

Figure 2: Means and RMSE in partial maximum likelihood estimates of parameters (μ, α, β) in model (16), using simulations of model (19) with $(\mu, \alpha, \beta) = (1, 1, 1)$. 100 simulations were performed for each T, for 50 equally spaced values of T between 1 and $50,000$.

model simulated	model estimated	T	RMSE($\hat{\mu}$)	RMSE($\hat{\alpha}$)	RMSE($\hat{\beta}$)
(23)	(16)	10	0.667	0.997	0.978
		100	0.357	0.432	0.488
		1000	0.102	0.148	0.146
(21)	(21)	10	0.642	1.07	1.00
		100	0.332	0.489	0.535
		1000	0.152	0.169	0.150
(20)	(21)	10	0.764	1.03	1.14
		100	0.361	0.459	0.479
		1000	0.165	0.156	0.152
(22)	(21)	10	0.882	1.04	1.21
		100	0.387	0.471	0.463
		1000	0.173	0.160	0.151
(21b)	(21b)	10	0.650	1.00	1.16
		100	0.330	0.467	0.456
		1000	0.113	0.155	0.161
(20b)	(21b)	10	0.731	1.16	1.01
		100	0.335	0.498	0.460
		1000	0.104	0.179	0.170
(22b)	(21b)	10	0.614	1.11	1.12
		100	0.345	0.461	0.470
		1000	0.0925	0.141	0.149

Table 1: RMSE and mean of parameter estimates for various models. In each case, the model was simulated 100 times, either with or without the covariates Z and W, and for each simulation the parameters for the model without this covariate were estimated by MLE using a Newton-Raphson gradient descent method in R. In each case the true parameters being estimated are $(\mu, \alpha, \beta) = (1, 1, 1)$. For models (20b - 21b), the parameter K was set to 0.01 so the process is clustered, with approximately 1% of the events attributable to this clustering. Model (22b) refer to model (22) with parameter $K = -0.01$ so that the process is inhibitory.

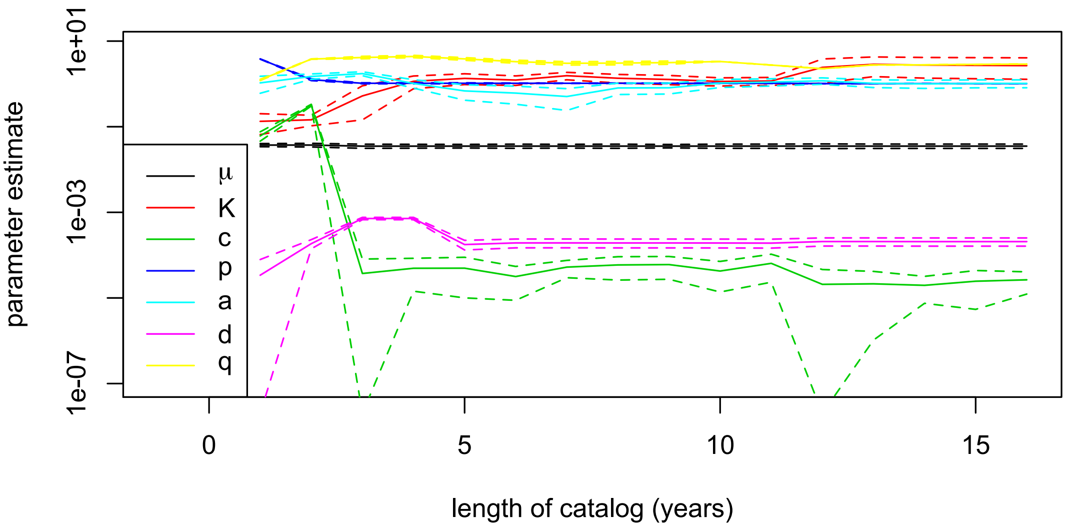

Figure 3: Estimated ETAS parameters, as a function of catalog length, using the progressive approximate MLE technique of Schoenberg (2013). Dashed curves represent estimates ± one standard error (SE), with the SE estimated using the inverse of the Hessian of the loglikelihood.

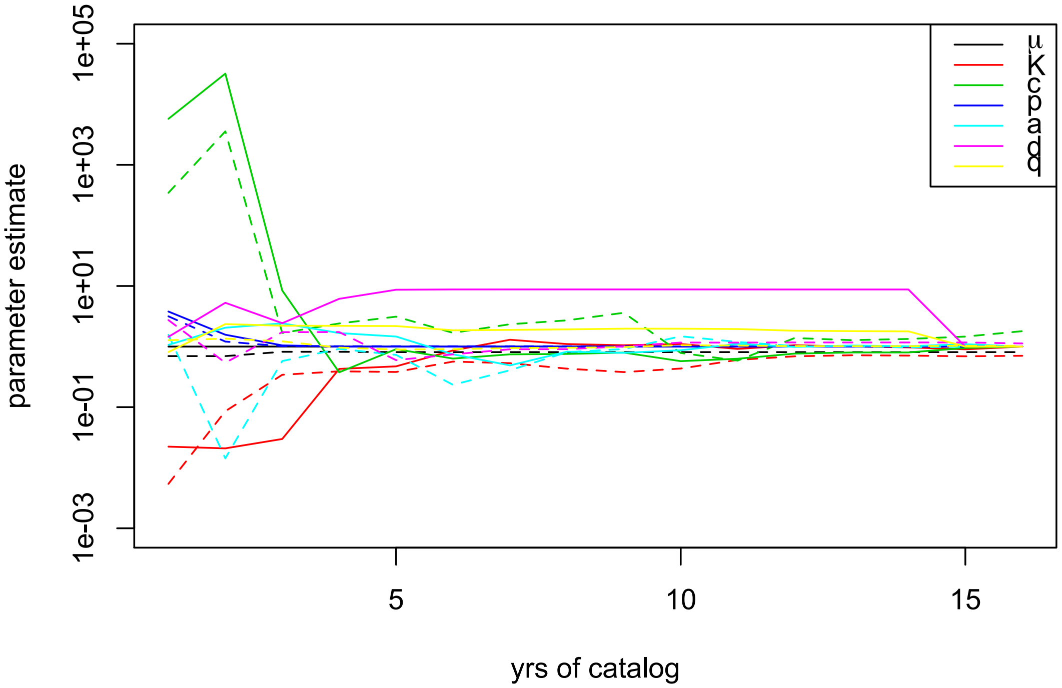

Figure 4: Estimated parameters in models (24) and (25) as functions of catalog length, again using the progressive approximate MLE technique of Schoenberg (2013). Here the parameter estimates are scaled by dividing by the final parameter estimates from model (24). The dashed curves represent estimates of parameters in model (25), and the solid curves correspond to parameter estimates for model (24).