A GENERALIZED THRESHOLD REGRESSION MODEL FOR ANALYZING NON-NORMAL NONLINEAR TIME SERIES: PLAGUE IN KAZAKHSTAN AS AN ILLUSTRATION

By NOELLE I. SAMIA

Department of Statistics & Actuarial Science,

University of Iowa, Iowa City, IA 52242, U.S.A.

nsamia@stat.uiowa.edu

KUNG-SIK CHAN

Department of Statistics & Actuarial Science,

University of Iowa, Iowa City, IA 52242, U.S.A.

kung-sik-chan@uiowa.edu

AND NILS CHR. STENSETH

Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology,

University of Oslo, P.O. Box 1050 Blindern, N-0316 Oslo, Norway

n.c.stenseth@bio.uio.no

SUMMARY

We introduce the (mixed-effect) Generalized Threshold Regression (GTR) model for piecewise linear stochastic regression with (possibly) non-normal time-series data. Specifically, it is assumed that the conditional probability distribution of the response variable belongs to the exponential family, and the conditional mean response is linked to some piecewise linear stochastic regression function. We study the specific case where the response variable equals zero in the lower regime. Some large-sample properties of a likelihood-based estimation scheme are derived. Our approach is motivated by the need for modeling nonlinearity in serially correlated epizootic events. Data coming from monitoring conducted in a natural plague focus in Kazakhstan are used to illustrate this model by obtaining biologically meaningful conclusions regarding the threshold relationship between prevalence of plague and some covariates including past abundance of great gerbils and other climatic variables.

Some key words: Binomial distribution; Delay; Epizootic events (plague outbreaks); Exponential family; Stochastic regression.

1. Introduction and the Model Formulation

We introduce a new statistical method motivated by the need for studying the biotic and abiotic factors affecting the prevalence of plague among the great gerbils in Kazakhstan. The great gerbil (*Rhombomys opimus*) populations constitute several natural foci to plague (caused by the bacteria *Yersinia pestis*) in Kazakhstan where the disease may be transmitted to humans by vectors, mainly, fleas (mainly of the genus *Xenopsylla*). A long-term monitoring study of this natural plague system was undertaken from 1949–1995, for tracking the prevalence of plague in the great gerbil populations. Monitoring efforts consisted primarily of trapping the great gerbils, together with their fleas, and testing both rodents and fleas for plague using a bacteriological test and a serological test; see Davis *et al.* (2004) and Park *et al.* (2005). Bacteriological tests may detect the presence of plague bacteria and hence plague disease in great gerbils at the time of sampling. On the other hand, serological tests may detect the presence of antibodies to plague bacteria and hence are indicative of past infections; see Park *et al.* (2005). In this paper, the development of new statistical methods for analyzing the bacteriological test data will be our main focus. Plague is still prevalent in several Asian, African, and American countries including the USA, and is today one of the re-emerging diseases; see Gage and Kosoy (2005).

The major difficulty and/or novelty of the problem is that epizootics occur only if the expected number of secondary infections arising from a primary infection (known as the basic reproductive ratio R_0) is greater than 1; see Dickmann and Heesterbeek (2000, chapter 1) and Keeling and Gilligan (2000). The parameter R_0 generally depends on the size and social structure (contact rates) of the study population; see chapter 5 of Dickmann and Heesterbeek (2000). Assuming that R_0 increases monotonically with population size (which holds for several popular epidemiological models including the SIR model, see p. 16 of Dickmann and Heesterbeek, 2000), the preceding condition of epizootics translates to the requirement that in host-pathogen dynamics with a fixed social structure, there is an unknown threshold population abundance below which an infectious disease is unlikely to invade a fully susceptible host population and persist. Specifically, above the threshold, R_0 is greater than 1 and the disease spreads and persists while below the threshold, R_0 is less than 1 and the disease dies out. Note that, in practice, the threshold effect may be delayed. For general surveys on the modeling of infectious diseases, see Anderson and May (1991), Grenfell and Dobson (1995), and Hudson *et al.* (2002).

We introduce the Generalized Threshold Regression (GTR) model that incorporates the preceding threshold condition for analyzing epidemiological time series that conditionally belongs to the exponential family. The general model is best illustrated by considering a specific GTR model for analyzing the Kazakhstan monitoring data. Let N_t be the number of great gerbils examined at time $t, t = 1, 2, \cdots, T$, and B_t be the (sample) proportion of great gerbils testing positive for plague under bacteriological tests at time t. Let $P_t = E(B_t)$ be the prevalence of plague at time t. The R_0 dichotomy discussed above implies that the true prevalence $P_t = 0$ when the (possibly delayed) abundance is below a threshold, and, otherwise, the prevalence is assumed to be dependent on some vector covariate process $Y = \{Y_t, t = 1, \cdots, T\}$, where Y_t consists of the abundance X_t of the great gerbils and its lags, as well as some other covariates and their lags. Furthermore, let $\epsilon = \{\epsilon_t, t \in Z\}$ be a latent process that may be used to account for possible overdispersion and some missing covariates such as the virulence of bacteria (infectivity variable). Conditional on N, Y, and ϵ, we may model $N_t B_t$ as independent binomial random variables with parameters (N_t, P_t), where

$$P_t = \begin{cases} 0, & \texttt{if } X_{t-d} < r \\ \texttt{logit}^{-1}\left(\beta' Y_t + \epsilon_t\right), & \texttt{if } X_{t-d} \geq r; \end{cases} \tag{1}$$

$t = 1, \cdots, T$. Besides the logit link function, other non-constant smooth link function may be employed.

The error terms ϵ_t are assumed to be a sequence of independent and identically distributed random variables with probability density function denoted by $f_\epsilon(.)$ that is indexed by some parameter vector ψ; the random effects are often assumed to be normally distributed with zero mean. Also, $\{\epsilon_t, t \in Z\}$ and $\left\{(N_t, Y_t')', t \in Z\right\}$ are assumed to be independent of each other. The parameter r is known as the threshold and d is a non-negative integer referred to as the delay or threshold lag. The analysis of the above threshold regression model is conditional on the observed Ns and Ys.

We now consider the general case that allows more general non-normal distribution including Poisson and negative binomial. Let $\{N_t, t = 1, 2, \cdots, T\}$ be a positive process and B_t be a non-negative discrete random variable whose conditional probability density function given N_t belongs to the exponential family and takes the form

$$f(B_t; \gamma_t, N_t, \nu) = \exp\left[\frac{N_t}{\nu}\{B_t\gamma_t - b(\gamma_t)\} + c(B_t, \frac{\nu}{N_t})\right], \tag{2}$$

where γ_t is the natural canonical parameter, N_t are weights, and ν a known scale parameter. It is well-known (McCullagh and Nelder, 1989) that, under some mild regularity conditions, the conditional mean $\mu_t = E(B_t) = \frac{\partial b(\gamma_t)}{\partial \gamma_t}$ is a one-to-one function of γ_t. Because of the non-negativity of B_t, $\mu_t = 0$ implies that B_t is degenerate and hence identically equals 0. We shall assume that the conditional distribution of B_t is degenerate if and only if $\mu_t = 0$. Hence, if $\mu_t > 0$, the (conditional) probability density function of B_t is strictly less than 1. Let ϵ be a latent process and Y be a covariate process that includes a univariate sub-process X. The Generalized Threshold Regression model (GTR) specifies that conditional on N, Y, and ϵ, B_t are independent random variables whose conditional means μ_t are given by

$$\mu_t = \begin{cases} 0, & \texttt{if } X_{t-d} < r \\ \mathtt{g}^{-1}\left(\beta' Y_t + \epsilon_t\right), & \texttt{if } X_{t-d} \geq r\texttt{;} \end{cases} \tag{3}$$

$t = 1, \cdots, T$. The function g is a known non-constant smooth link function such that its inverse function g^{-1} is a positive function. (As g^{-1} must be non-negative because of the non-negativity of B_t, this is a mild condition; it is required for the validity of Claim 1 below.) Clearly, the GTR model subsumes the specific model for the Kazakhstan monitoring data discussed earlier. Yet, a more general form of the GTR model obtains by (i) removing the restrictions on the positivity of the inverse link function and the discreteness and non-negativity of B_t, (ii) partitioning the sample space of (possibly vector-valued) X_{t-d} into a finite set of regions (often referred to as regimes), say $R_i, i = 1, \cdots, m$ and (iii) requiring that $g(\mu_t)$ equals a linear function, say $\ell_i(Y_t, \epsilon_t)$, whenever $X_{t-d} \in R_i$.

The GTR model bears resemblance to the so-called Open-Loop Threshold model. The Open-Loop Threshold model is essentially a piecewise linear stochastic regression model with the errors often assumed to be normal; see Tong (1990). In particular, the Open-loop Threshold model with normal errors is a special case of the GTR model with identity link and normal conditional distributions. The use of the link function removes any inherent constraints on the conditional mean function of a non-normal response so that on the scale of the link function, the mean response may be specified as some unconstrained linear or nonlinear stochastic regression function, being a piecewise linear stochastic regression function for the GTR model. From this perspective, the link function is a useful, natural device for extending useful normal time-series models to studying non-normal time-series data.

Henceforth, we focus on the GTR model defined by (3), i.e. of two regimes, with a positive inverse link function and of identically zero response in the lower regime. While the GTR model is motivated by the needs for analyzing the Kazakhstan plague monitoring data, it is of general applicability for analyzing epidemiological time series and other time-series data, e.g. consumer choice data, whose conditional distributions belong to the exponential family and that are subject to the zero-response condition below a threshold.

The rest of the paper is organized as follows. We introduce in § 2 a likelihood-based estimation scheme for the GTR model, and its large-sample properties are derived in § 3. The empirical performance of the proposed estimation scheme is studied through Monte Carlo in § 4. We illustrate the GTR model with an analysis of the Kazakhstan plague monitoring data in § 5. A brief conclusion is given in § 6. All proofs are deferred to several appendices.

2. Maximum Likelihood Estimation

Estimation of the Generalized Threshold Regression (GTR) model defined by (3) is straightforward were the delay parameter d and the threshold parameter r known, as then intuitively the regression parameter β and the nuisance parameter ψ can be estimated by fitting a mixed-effect generalized linear regression model with data cases whose $X_{t-d} \geq r$. This observation is the basis of the strategy of first estimating the threshold parameter given a fixed delay. We shall show below that, given the delay parameter, the profile log likelihood function of the threshold parameter is initially an increasing function and then drops to $-\infty$ at and beyond a point (the maximum likelihood estimator of the threshold) which can be determined by a simple sorting procedure. This property of the profile log likelihood simplifies the estimation of the threshold parameter. Let $\ell(\theta)$ be the log likelihood of θ, where $\theta = (d, r, \beta', \psi')'$. For any positive integer i, let $Z_i = (N_i, Y_i')'$. Because of the independence between $\{\epsilon_t, t \in Z\}$ and $\{Z_t, t \in Z\}$, it is readily checked that

$$\ell(\theta) = \sum_{i=1}^{T} \log \left[E\left\{ f(B_i|\epsilon_i, Z_i, \theta)|Z_i, \theta \right\} \right], \tag{4}$$

where $E(\cdot|\cdot)$ denotes the conditional expectation and the conditional probability density function $f(\cdot|\cdot)$ equals (2) with the canonical parameter γ_i being implicitly defined by the mean parameter μ_i specified by (3); c.f. Pinheiro and Bates (2002, p. 62). The summands in (4) can be simplified for data falling in the lower regime, i.e. $X_{i-d} < r$, in which case $\mu_i = 0$ and hence B_i identically equals 0 under the model indexed by θ, so that the corresponding summand equals 0 if $B_i = 0$ and $-\infty$ if $B_i > 0$. Hence,

$$\ell(\theta) = \sum_{i=1}^{T} \log \left[E\left\{ f(B_i|\epsilon_i, Z_i, \theta)|Z_i, \theta \right\} \right] I(X_{i-d} \geq r) + (-\infty) \times \sum_{i=1}^{T} I(X_{i-d} < r, B_i > 0), \tag{5}$$

where $I(.)$ equals 1 if and only if the enclosed expression is true and we adopt the convention that $0 \times (-\infty) = 0$ and a positive scalar times $-\infty$ equals $-\infty$. The first sum on the right side of (5) is clearly a finite number so that the log likelihood is $-\infty$ if and only if there exists some data case where $X_{i-d} < r$ but $B_i > 0$, i.e. the specified parameter values lead to an inconsistent observation for which the probability of success is zero and yet a success is observed. For a fixed d, the inconsistent cases can be readily identified by ordering the data cases by the values of Xs so that $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(T)}$. Define $\hat{t} = \hat{t}(d)$ such that $B_{(i)+d} = 0$ for all $1 \leq i < \hat{t}$ with $i + d \leq T$, and it is observed that $B_{(\hat{t})+d} > 0$. Then, it is readily checked that for $r > X_{(\hat{t})}$, the data pair $(X_{(\hat{t})}, B_{(\hat{t})+d})$ is inconsistent with the model. Thus, the profile log likelihood becomes $-\infty$, for all $r > X_{(\hat{t})}$. Consequently, for fixed d, the maximum likelihood estimator of r must be less than or equal to $X_{(\hat{t})}$. We show in Appendix 1 the validity of the following claim.

Claim 1: For fixed d and if the profile log likelihood of the threshold parameter is well-defined for $r \leq X_{(\hat{t})}$, then it is an increasing (step) function for $r \leq X_{(\hat{t})}$.

We note that the profile likelihood function may not be well-defined if, e.g. the sample size is too small. Hence, given d, the maximum likelihood estimator of r, denoted by $\hat{r}(d)$, equals $X_{(\hat{t})}$ which can be obtained by a simple sorting procedure as illustrated by the following example.

Example 1. Given a fixed threshold delay d, consider the following data points

$$\{(X_t, B_{t+d})\} = \{(0.5, 90), (0.8, 100), (0.2, 0), (0.49, 0), (0.7, 0), (0.15, 0)\}.$$

Then, by rearranging so that $\{X_t\}$ is in ascending order, we get the following

$$\left\{(X_{(j)}, B_{(j)+d})\right\} = \{(0.15, 0), (0.2, 0), (0.49, 0), (0.5, 90), (0.7, 0), (0.8, 100)\}.$$

In this example, $\hat{r}(d) = 0.5$.

Given $r = \hat{r}(d)$, the corresponding maximum likelihood estimator of the regression parameter β and that of the nuisance parameter ψ of the error distribution, denoted by $\hat{\beta}(d)$ and $\hat{\psi}(d)$, can be obtained by maximizing the first sum on the right side of (5), which is equivalent to fitting an associated mixed-effect generalized linear model for data falling in the upper regime, i.e. $X_{t-d} \geq \hat{r}(d)$. In practice, the true delay d_0 is unknown and needs to be estimated. Let $D > 0$ be some known integer upper bound of the unknown delay d. In the next section, we shall show that under suitable regularity conditions $\hat{r}(d_0) \geq r_0$, the true threshold, but there exists a $\delta > 0$ such that for $0 \leq d \leq D$ and $d \neq d_0$, $\hat{r}(d) < r_0 - \delta$ with probability $\to 1$ as $T \to \infty$. This observation motivates estimating the delay by the simple estimator $\tilde{d}$ defined as the smallest $0 \leq d \leq D$ with the largest $\hat{r}(d) = X_{(\hat{t}(d))}$. It is shown below in Appendix 2 that under some mild regularity conditions, $\tilde{d}$ is consistent. The other parameter estimators can then be simply defined as the maximum likelihood estimators given $d = \tilde{d}$, i.e. $\hat{r} = \hat{r}(\tilde{d})$, $\hat{\beta} = \hat{\beta}(\tilde{d})$ and $\hat{\psi} = \hat{\psi}(\tilde{d})$. Because of the discreteness of the delay parameter, consistency of $\tilde{d}$ implies that it is ultimately equal to the true delay d_0 almost surely.

The maximum likelihood estimator $\hat{d}$ provides a competitive alternative to $\tilde{d}$. However, the computation of $\hat{d}$ is more complex and its sampling properties depend on the covariate distribution and the specification of the random effects. Hence, we prefer the use of the simpler $\tilde{d}$ to $\hat{d}$ for exploratory data analysis. Nonetheless, $\hat{d}$ may be employed in a refined analysis, and the study of its sampling properties constitutes an interesting future research problem.

3. Large-Sample Properties of the Estimator

We first recall two notions of mixing properties. A stationary process $\{W_t\}$ is said to be α-mixing if there exists a sequence of numbers $\{\alpha(k)\}$ with $\alpha(k) \to 0$ as $k \to \infty$, and such that for any events E_1 in the σ-algebra generated by $\{W_t, t \leq j\}$ and E_2 in the σ-algebra generated by $\{W_t, t \geq j+k\}$,

$$|P(E_1 \cap E_2) - P(E_1)P(E_2)| \leq \alpha(k). \tag{6}$$

If the right side of inequality (6) is replaced by $\psi(k)P(E_1)$ with $\psi(k) \to 0$ as $k \to \infty$, then the process is said to be $\psi-$mixing. See Billingsley (1968, section 20) and Doukhan (1994, pp. 3 and 20) for further discussion of ψ-mixing. In order to study the asymptotic properties of the maximum likelihood estimator, the following set of assumptions will be required later.

A1. The process $Z = \{Z_t\}$ is stationary and α-mixing with exponentially decaying mixing coefficients, i.e. for all $k \geq 0, |\alpha_k| \leq c\rho^k$ for some $c > 0$ and $0 \leq \rho < 1$.

A2. There exists a $\delta > 0$ such that the process $\{X_t\ I(r_0 \leq X_t \leq r_0 + \delta)\}$ is $\psi-$mixing with exponentially decaying mixing coefficients.

A3. $\{X_t\}$ admits a marginal probability density function $\pi(.)$ that is continuous at the true threshold r_0 which is an interior point of the range of X, and $\pi(r_0) > 0$. Also, the joint marginal probability density functions $\pi_{ij}(.,.)$ of $(X_i, X_j)'$, for all $i \neq j$, are assumed to be positive everywhere and uniformly bounded.

A4. Conditional on $X_{t-d} = r$, Z_t has a probability density function that is weakly continuous at r_0; see Feller (1971, p. 243) for a discussion of weak continuity.

Remark 1. Recall that Z_t includes X_t so that A1 implies that $\{X_t\}$ is stationary and α-mixing. The α-mixing condition is needed so that the Law of Large Numbers and the Central Limit Theorem hold. It can be relaxed somewhat at the expense of requiring more complex conditions. The ψ-mixing condition for the process $\{X_t\ I(r_0 \leq X_t \leq r_0+\delta)\}$ holds if, e.g. $\{X_t\}$ is a Markov chain that is uniformly ergodic when it is restricted to the interval $[r_0, r_0 + \delta]$; see Nummelin (1984, section 5.6). The ψ-mixing condition serves as an essential condition entailing that for a fixed $\tau > 0$, the counting process $\sum_{i=1}^{T} I(r_0 \le X_t \le r_0 + \tau/T)$ has an asymptotic Poisson distribution. This asymptotic distribution forms a technical argument for showing that $\hat{r}$ has an $O_p(1/T)$ convergence rate. Assumptions A3 and A4 are mild regularity conditions.

Theorem 1 below states the consistency of the estimator $\tilde{d}$, the proof of which is deferred to Appendix 2.

Theorem 1. *Suppose assumptions A1 and A3 hold. Let* $\Omega = \{0, 1, \cdots, D\}$ *be the parameter space of* d*, where* $D > 0$ *is a known positive integer. Then, the estimator* $\tilde{d}$ *of the true parameter* $d_0 \in \Omega$ *is strongly consistent. It follows from the discreteness of the delay parameter that for all sufficiently large* $T, \tilde{d} = d_0$ *with probability 1.*

Because of Theorem 1, without loss of generality, we may and shall assume henceforth in this section that the delay parameter is known. Also, we write d for d_0. The parameter d is, furthermore, deleted from θ. We next show in Theorem 2 that the maximum likelihood estimator of the threshold is T-consistent, whose proof is deferred to Appendix 3.

Theorem 2. *Suppose assumptions A1–A4 hold. Then the maximum likelihood estimator of the threshold is such that* $\hat{r} = r_0 + O_p(1/T)$ *where* T *is the sample size. Moreover,* $T(\hat{r} - r_0)$ *has an asymptotic exponential distribution with mean equal to* $1/\{\pi(r_0)\, p_0\}$*, where*
$p_0 = 1 - P(B_t = 0|X_{t-d} = r_0) = 1 - E\left\{\int f(0|\epsilon_t, Z_t, \theta) f_\epsilon(\epsilon_t; \psi) d\epsilon_t \,|X_{t-d} = r_0\right\}.$

Remark 2. The $O_p(1/T)$ fast convergence rate is due to the discontinuity of the conditional mean function; see Chan (1993) and Chan and Tsay (1998). See also Hansen (2000). The asymptotic distribution of $\hat{r}$ can be expressed as $r_0 + \mathcal{E}/\{T\pi(r_0)p_0\}$ where $\mathcal{E}$ denotes the exponential distribution with unit mean. Let $r_{0.95,T}$ be the 95 percentile of $\mathcal{E}/\{T\pi(r_0)p_0\}$. Then $(\hat{r} - r_{0.95,T}, \hat{r}]$ is an asymptotic 95% confidence interval of r_0. In practice, $r_{0.95,T}$ needs to be estimated before computing an asymptotic confidence interval for r_0. However, the preceding consideration motivates the simpler alternative of bootstrap confidence intervals. Let $\hat{r}_1^*, \cdots, \hat{r}_B^*$ be B copies of independent parametric bootstrap threshold estimates and $r_{0.95}^*$ be the corresponding 95 percentile. Approximating the distribution of $\hat{r} - r_0$ by the bootstrap distribution of $\hat{r}^* - \hat{r}$, $(2\hat{r} - r_{0.95}^*, \hat{r}]$ is a bootstrap 95% confidence interval of r_0; c.f. section 2.4 of Hinkley and Davison (2003). The bootstrap method is illustrated in the real application.

The super-consistency of the threshold parameter, i.e. the $O_p(1/T)$ convergence rate, implies that under some regularity conditions, it is asymptotically independent of $\hat{\beta}$ and $\hat{\psi}$, which is the content of Theorem 3 below. Moreover, we show that $\hat{\beta}$ and $\hat{\psi}$ are $\sqrt{T}$-consistent and whose joint asymptotic distribution is identical to that obtained from fitting the associated mixed-effect generalized linear model with data falling in the upper regime of the GTR with known true threshold and delay. We now briefly outline this asymptotic equivalence result. First, we introduce some notations. Define $\delta = (\beta', \psi')', \theta = (r, \delta')'$. Recall that $Z_i = (N_i, Y_i')'$, and let $\hat{\delta}(r) = \arg\max_\delta \ell(\theta)$, for a fixed r. The log likelihood function of the associated mixed-effect generalized linear model is given by

$$\begin{aligned}\tilde{\ell}(\theta) &= \sum_{i=1}^{T} \log\left[E\left\{f(B_i|\epsilon_i, Z_i, \theta)|Z_i, \theta\right\}\right] I(X_{i-d} \geq r)\\ &= \sum_{i=1}^{T} l_\delta(B_i; Z_i) I(X_{i-d} \geq r),\end{aligned}$$

where $l_\delta(B_i; Z_i)$ is defined as the logarithmic expression on the first equation. Let $\varphi_\delta(Z_i) = \dot{l}_\delta(B_i; Z_i) I\left(X_{i-d} \geq r_0\right)$, where $\dot{l}_\delta(B_i; Z_i) = \frac{\partial}{\partial\delta} l_\delta(B_i; Z_i)$. Define

$$\Psi_T(\delta) = \frac{1}{T}\sum_{i=1}^{T} \dot{l}_\delta(B_i; Z_i) I\left(X_{i-d} \geq \hat{r}\right). \tag{7}$$

The maximum likelihood estimator $\hat{\delta} = \hat{\delta}(\hat{r})$ is a root of the estimating equation $\Psi_T(\delta) = 0$. On the other hand, for the associated mixed-effect generalized linear model with data from the true upper regime of the GTR model, the maximum likelihood estimator equals $\hat{\delta}(r_0)$ which is a root of the following estimating equation

$$\frac{1}{T}\sum_{i=1}^{T} \dot{l}_\delta(B_i; Z_i) I\left(X_{i-d} \geq r_0\right) = 0. \tag{8}$$

We show in Theorem 3 that the super-consistency of $\hat{r}$ and other regularity conditions imply that the two estimating equations (7) and (8) are asymptotically equivalent so that $\hat{\delta} = \hat{\delta}(\hat{r})$ and $\hat{\delta}(r_0)$ enjoy the same asymptotic distribution. The following set of assumptions will be required later.

- B1. The domain of δ is an open subset of the Euclidean space, in which $\delta \mapsto \dot{l}_\delta(b; z)$ is twice continuously differentiable for every (b, z).
- B2. $E\left|\dot{l}_{\delta_0}\right|^2 < \infty$, where $|\cdot|$ denotes the Euclidean norm of the enclosed expression and the expectation is taken under the true model.
- B3. $E(\ddot{l}_{\delta_0})$ exists and is nonsingular.

B4. The third-order partial derivatives of $l_\delta(b; z)$ with respect to δ are dominated by a fixed integrable function $g(b; z)$ for every δ in a neighborhood of δ_0.

We note that these assumptions are classical conditions for studying the asymptotic properties of maximum likelihood estimation; see Theorems 5.41 and 5.42 of van der Vaart (2000). Under assumptions A1, B1–B4, it can be shown using Theorem 5.41 in van der Vaart (2000) that the estimating equation (8) admits a root close to the true parameter value δ_0. These roots are designated as $\hat{\delta}(r_0)$ which can be shown to be $\sqrt{T}$-consistent, i.e.

$$\hat{\delta}(r_0) - \delta_0(r_0) = O_p(1/\sqrt{T}).$$

Moreover, by Theorem 5.42 of van der Vaart (2000), the sequence $\sqrt{T}\left\{\hat{\delta}(r_0) - \delta_0(r_0)\right\}$ is asymptotically normal with mean zero and covariance matrix $E(\dot{\varphi}_{\delta_0})^{-1}E(\varphi_{\delta_0}\varphi_{\delta_0}^{'})E(\dot{\varphi}_{\delta_0})^{-1}$. These results can be transferred to $\hat{\delta}$, as stated in Theorem 3 below.

Remark 3. We remark that Theorem 5.41 of van der Vaart (2000) assumes independent and identically distributed data. However, it can be easily extended to the case of α-mixing data with geometrically decaying mixing coefficients, and if $\{\varphi_{\delta_0}(Z_i)\}$ are uncorrelated. The latter condition holds because of the conditional independence of Bs given Z and ϵ.

Theorem 3. *Suppose assumptions A1–A4 and B1–B4 hold. Then,*

$$\hat{\delta}(\hat{r}) - \delta_0(r_0) = O_p(1/\sqrt{T}),$$

and the sequence $\sqrt{T}\left\{\hat{\delta}(\hat{r}) - \delta_0(r_0)\right\}$ *is asymptotically normal with mean zero and covariance matrix* $E(\dot{\varphi}_{\delta_0})^{-1}E(\varphi_{\delta_0}\varphi_{\delta_0}^{'})E(\dot{\varphi}_{\delta_0})^{-1}$.

See Appendix 4 for a proof of Theorem 3.

4. Simulation Study

We conduct a simulation study to investigate the empirical performance of the proposed estimation scheme for the GTR model defined by (3). Conditionally independent observations of N_tB_t are generated from binomial distributions with number of trials N_t equal to 1000 and conditional probabilities of success given by

$$P_t = \begin{cases} 0, & \texttt{if } X_{t-d} < r \\ \texttt{logit}^{-1}\left(\beta_0 + \beta_1 A_{t,1} + \beta_2 A_{t,2} + \beta_3 A_{t,2}A_{t,3} + \epsilon_t\right), & \texttt{if } X_{t-d} \geq r; \end{cases} \tag{9}$$

$t = 1, \cdots, T$. The parameters d and r are taken to be 1 and 0.4, respectively. The regression coefficients are fixed at $\beta' = (0, 5, -5, 0)$. The Xs are generated as a series that follows an AR(2) process given by $X_t = \frac{O_t + 0.907}{2.37}$, where $O_t = 0.9255 O_{t-1} - 0.2736 O_{t-2} + \sqrt{0.02125}\, \eta_t$, and η_t denotes a series of uncorrelated normal random variables with zero mean and variance 1, truncated between -3 and 3. Note that the parameters are chosen such that X_t is bounded between 0 and 1. The covariates $A_{t,1}, A_{t,2}$, and $A_{t,3}$ are generated as independent uniform(0, 1) random variables. The random effect sequence $\{\epsilon_t\}$ is generated as a series of independent $N(0, \sigma^2)$ random variables, where σ is taken to be 0.1, 0.2, and 0.5. The sample sizes used are 50, 100, and 200, and for each sample size, the results are based on 1,000 replications (excluding cases of failures, see below). The associated mixed-effect logistic regression model is fitted using the glmmPQL function in R, which implements approximate maximum likelihood estimation by linearization about the Best Linear Unbiased Predictors (BLUPs); see Breslow and Clayton (1993) and Venables and Ripley (2002, pp. 297–298).

Table 1 provides the mean number of data points in the upper regime, the percentage of times the delay was estimated to be equal to the true value 1, and the percentage of times an error has occurred while fitting the simulated model, namely the percentage of failures. The percentage of failures refers to the cases where glmmPQL gives an error (e.g. non-positive definite asymptotic covariance matrix) or a warning message (e.g. non-convergence and/or the Hauck-Donner phenomenon; see pp. 197–198 of Venables and Ripley, 2002). The percentage of failures generally decreased with larger sample size. We also report in Table 1 the sample means, biases, and standard deviations of the estimates, and the empirical coverage probabilities of the parameters in the associated mixed-effect logistic regression model. The empirical coverage probabilities are based on the asymptotic 95% confidence intervals of the corresponding parameters. In general, the standard deviation of the estimators generally became smaller with larger sample size, confirming the consistency results discussed previously. Moreover, the empirical coverage probabilities were generally closer to the nominal coverage probabilities with increasing sample sizes.

In all cases considered in this simulation study, the Q-Q plots (not shown) confirm the limiting exponential distribution of the threshold estimator and the asymptotic normality of the remaining estimators in the associated mixed-effect logistic regression model.

5. Application

In this section we illustrate the GTR model by studying which biotic and/or abiotic factors affect the prevalence of plague among great gerbil population in Kazakhstan. Plague exists in nature

Sample Size	Mean # of Data in Upper Regime	True σ	% of $\hat{d}=1$ (in %)		Parameter Estimates $\hat{r}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}$	Empirical Coverage Rate of β_0	β_1	β_2	β_3	σ	% of Failures (in %)
50	21	0.1	100		0.4049	0.0005	5.0004	-5.0017	-0.0043	0.0749	0.912	0.905	0.904	0.902	0.912	26.79
				sd	0.0052	0.0936	0.1477	0.2144	0.2804	0.0493						
				bias	0.0049	0.0005	0.0004	-0.0017	-0.0043	-0.0251						
50	21	0.2	100		0.4049	0.0029	4.9976	-5.0225	0.0296	0.1533	0.892	0.897	0.898	0.899	0.915	25.26
				sd	0.0049	0.1539	0.2185	0.3064	0.4204	0.0696						
				bias	0.0049	0.0029	-0.0024	-0.0225	0.0296	-0.0467						
50	21	0.5	100		0.4049	-0.0191	5.0019	-4.9743	0.0190	0.3841	0.900	0.887	0.906	0.909	0.849	28.93
				sd	0.0050	0.3480	0.5149	0.6939	0.9015	0.1448						
				bias	0.0049	-0.0191	0.0019	0.0257	0.0190	-0.1159						
100	43	0.1	100		0.4027	-0.0006	4.9958	-4.9961	0.0021	0.0826	0.912	0.924	0.916	0.924	0.936	15.11
				sd	0.0027	0.0593	0.0904	0.1155	0.1440	0.0409						
				bias	0.0027	-0.0006	-0.0042	0.0039	0.0021	-0.0174						
100	43	0.2	100		0.4024	-0.0042	4.9954	-4.9924	0.0089	0.1773	0.937	0.933	0.938	0.935	0.973	18.57
				sd	0.0025	0.0952	0.1387	0.1770	0.2350	0.0470						
				bias	0.0024	-0.0042	-0.0046	0.0076	0.0089	-0.0227						
100	43	0.5	100		0.4023	0.0012	4.9592	-4.9649	-0.0057	0.4373	0.942	0.924	0.925	0.946	0.928	28.37
				sd	0.0023	0.2153	0.3076	0.3937	0.4962	0.0941						
				bias	0.0023	0.0012	-0.0408	0.0351	-0.0057	-0.0627						
200	86	0.1	100		0.4013	0.0002	4.9938	-4.9953	-0.0010	0.0932	0.960	0.951	0.932	0.937	0.965	16.39
				sd	0.0012	0.0367	0.0576	0.0772	0.0969	0.0293						
				bias	0.0013	0.0002	-0.0062	0.0047	-0.0010	-0.0068						
200	86	0.2	100		0.4012	-0.0033	4.9911	-4.9881	-0.0015	0.1906	0.933	0.957	0.949	0.947	0.968	8.93
				sd	0.0012	0.0656	0.0901	0.1187	0.1557	0.0292						
				bias	0.0012	-0.0033	-0.0089	0.0119	-0.0015	-0.0094						
200	86	0.5	100		0.4012	-0.0086	4.9710	-4.9567	-0.0020	0.4657	0.938	0.93	0.947	0.954	0.952	21.14
				sd	0.0012	0.1505	0.2075	0.2616	0.3394	0.0587						
				bias	0.0012	-0.0086	-0.0290	0.0433	-0.0020	-0.0343						
				True	0.4	0.0	5.0	-5.0	0.0							

Table 1. Results of the simulation study.

as a disease of wild rodents caused by infection of the bacterium *Yersinia pestis.* The infection is maintained in natural foci of the disease, in wild rodent colonies through transmission between rodents by their flea ectoparasites. In desert and semidesert areas of Kazakhstan (and central Asia, in general), the great gerbil (*Rhombomys opimus*) and the fleas inhabiting their burrows (mainly of the genus *Xenopsylla*) are considered to be the main host and vectors of plague, respectively. The survey area is located south-east of Lake Balkhash in south-eastern Kazakhstan, being part of the PreBalkhash plague focus. The PreBalkhash focus is separated into specific landscape-epizootological regions. Kazakhstan was divided into 40×40 km^2 so called large squares (LSQ). Each LSQ comprises four 20×20 km^2 primary squares that are divided into four sectors. Within a given sector, data are maximally recorded twice a year, providing information on the results of bacteriological test (prevalence data) in addition to independent information on the burrow occupancy rate which is a proxy of the abundance of the great gerbils (and hence also for the contact rate between the great gerbils and the fleas). The sampling was done bi-annually during the spring and the fall from 1949 to 1995. The great gerbils were mainly caught between May and June in the spring and between September and October in the fall. Below we develop the GTR model that embodies the principle that plague outbreak occurs only if the occupancy is above a certain threshold, in which case the probability of an outbreak is higher with more favourable environmental conditions. Because the prevalence structure is likely to be seasonal, we extend the GTR model to a seasonal GTR model where both the threshold and the delay are seasonal and so is the prevalence function defined below. The estimation procedure introduced earlier can be readily modified to estimating a seasonal GTR model; the large-sample results can be similarly extended to the estimator of a seasonal GTR model. For the sake of illustration, we fit the GTR model for only one large square, namely LSQ 105. Potential covariates include a large set of climate variables, current occupancy, lag $\frac{1}{2}$ and lag 1 occupancies. We tried a number of subset seasonal GTR models. Because of the small sample size in the upper regime of each season and some covariates have missing data, some subset models cannot be fitted. Based on AIC and the significance of each covariate effect, we obtain the final fitted model for LSQ 105 that is given below. Recall that N_t is the number of great gerbils examined at time $t, t = 1, 2, \cdots, T$, and B_t is the (sample) proportion of great gerbils testing positive for plague under bacteriological tests at time t. Let $P_t = E(B_t)$ be the prevalence of plague at time t. We fitted models of the form specified by (1), but with a seasonal structure:

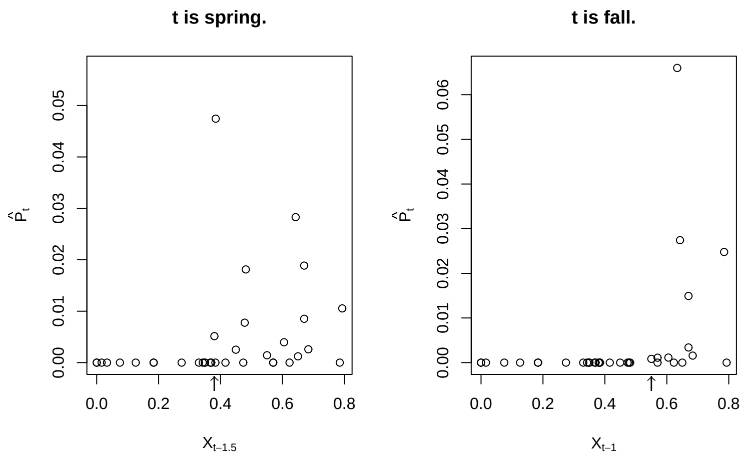
$$P_t = \begin{cases} \begin{cases} 0, & \texttt{if } X_{t-d_s} < r_s \texttt{ and } t \texttt{ is spring} \\ \texttt{logit}^{-1}\left(\beta_{0,s} + \beta_{1,s}X_{t-1} + \beta_{2,s}F_{wi,t} + \epsilon_t\right), & \texttt{if } X_{t-d_s} \geq r_s \texttt{ and } t \texttt{ is spring;} \end{cases} \\ \\ \begin{cases} 0, & \texttt{if } X_{t-d_f} < r_f \texttt{ and } t \texttt{ is fall} \\ \texttt{logit}^{-1}\Big(\beta_{0,f} + \beta_{1,f}X_{t-\frac{1}{2}} + \beta_{2,f}R_{su,t} & \\ \qquad +\beta_{3,f}R_{su,t} \times T_{su,t} + \epsilon_t\Big), & \texttt{if } X_{t-d_f} \geq r_f \texttt{ and } t \texttt{ is fall.} \end{cases} \end{cases} \tag{10}$$

Occupancy is the X-variable. It is important to note that we use lag $\frac{1}{2}$ occupancy, because data are collected twice per year. Spring climate covariate is the average monthly number of days with frost during the winter, namely $F_{wi,t}$. Fall climate covariates are the average monthly number of days with relative humidity less than 30% during the summer ($R_{su,t}$) and the average monthly summer temperature ($T_{su,t}$). Owing to possible overdispersion and some missing covariates such as the virulence of bacteria (infectivity variable), the latent variables ϵ_t are included in the model. Moreover, they are assumed to be independent $N(0, \sigma^2)$ random variables.

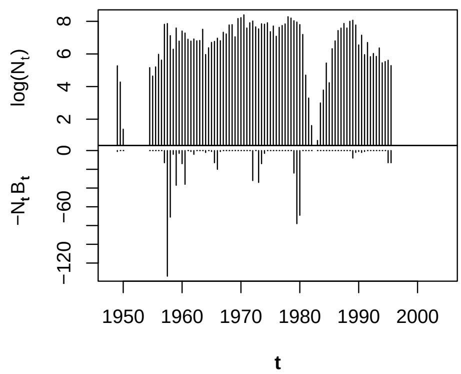
The seasonal delay parameters d_s and d_f are estimated to be 1.5 and 1, respectively, using the estimation procedure discussed in this paper. Figure 1(a) shows the seasonal scatter plots of the observed prevalence rate $\hat{P}_t = B_t$ versus the threshold variable, i.e. versus $X_{t-1.5}$ for the spring and X_{t-1} for the fall. The arrows in the plots of Figure 1(a) indicate the location of the maximum likelihood estimates of the seasonal thresholds which are estimated using the simple sorting procedure discussed earlier. Note that $\hat{P}_t$ is likely to be biased downwards, because the test sample of an infected rodent may not include bacteria, the efficacy of the test, etc. Fortunately, the bias can be absorbed by the intercept term; see Park *et al.* (2005).

The time-series plots of $\log(N_t)$ and $-N_tB_t$ in Figure 1(b) start from the spring of 1949 and end in the fall of 1995. These time-series plots show some missing observations and the existence of a large number of observations for which B_t is equal to zero. Figures 2(a) and (b) show the pairwise scatter plots for the dataset in the upper regime of the spring and fall seasons, respectively.

In Table 2, we report the maximum likelihood estimate of each of the parameters in the model defined by (10) with their asymptotic standard errors and asymptotic 95% confidence intervals. As in the simulation study, the associated mixed-effect logistic regression model was fitted using the glmmPQL function in R. Since the number of data points in the upper regime is 18 for the spring and 12 for the fall, it is prudent to alternatively calibrate the uncertainty of the parameter



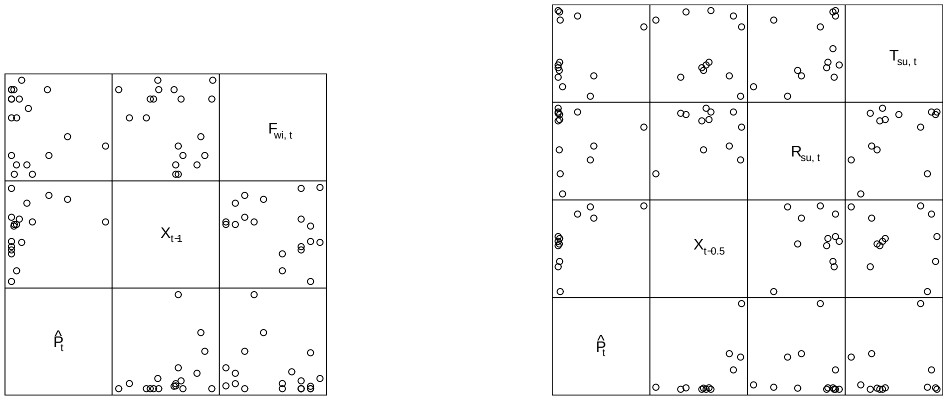
(a)



(b)

FIGURE 1. (a) Seasonal scatter plots of $\hat{P}_t$ vs. the threshold variable. (b) Time-series plots of $\log(N_t)$ and $-N_tB_t$.

estimates by using the parametric bootstrap with bootstrap size 1,000 and with the covariates fixed at their observed values and the seasonal delays fixed at $d_s = 1.5$ and $d_f = 1$; see Table 2. The bootstrap 95% confidence intervals for the seasonal threshold parameters are based on $r^*_{0.95}$, the 95 percentile of the bootstrap estimates. The end points of these 95% confidence intervals are $(2\hat{r} - r^*_{0.95}, \hat{r}]$; see Remark 2. The bootstrap 95% confidence intervals for the remaining parameters in the upper regime are obtained by the percentile method with the 2.5 and 97.5 percentiles of the



(a) (b)

FIGURE 2. Pairwise scatter plots for the data in the upper regime (a) Spring season; (b) Fall season.

bootstrap estimates being the end points of these 95% confidence intervals; see Efron and Tibshirani (1993, chapter 13). The bootstrap confidence intervals are generally wider than their asymptotic counterparts.

Standard diagnostic checks on the estimated random effects from our final fitted model show no evidence of any failure of the assumptions related to the random effect. Therefore, the fitted model seems adequate. Based on the asymptotic and parametric bootstrap confidence intervals, we conclude that (given everything else being equal) the occupancy of last spring season is positively correlated with the prevalence of plague during the spring. The average number of days with frost during the winter season is negatively correlated with the prevalence of plague during the spring; in other words, warm spells in the winter increase the chance of a plague outbreak during the spring season. On the other hand, the occupancy of last spring is positively correlated with the prevalence of plague during the fall season. Moreover, average summer temperature is positively correlated with the prevalence of plague during the fall. Average summer temperature ranged between 22.8 and 25.7, over the span of which the number of days with relative humidity less than 30% during the summer season is negatively correlated with the prevalence of plague during the fall. Hence, a dry summer decreases the chance of a plague outbreak during the fall season.

We have also fitted a simple logistic regression model to the plague data in LSQ 105 where the threshold phenomenon is discarded. The estimate of σ is found to be 8.43×10^{-5}, suggesting that the random effects can be dropped. Table 3 reports the simple logistic regression model fit without

Parameter	Estimated Value	Asymptotic Standard Error	Asymptotic 95% C.I.	Bootstrap 95% C.I.
r_s	0.380			$(0.311, \quad 0.380]$
r_f	0.550			$(0.530, \quad 0.550]$
$\beta_{0,s}$	3.48	11	$(-16.8, \quad 23.8)$	$(-15.4, \quad 28.8)$
$\beta_{1,s}$	12.3	5.5	$(2.37, \quad 22.3)$	$(3.35, \quad 27.3)$
$\beta_{2,s}$	-0.544	0.35	$(-1.18, \quad 0.0944)$	$(-1.48, \quad 0.0498)$
$\beta_{0,f}$	-11.6	3.1	$(-17.2, \quad -5.87)$	$(-22.8, \quad -7.05)$
$\beta_{1,f}$	23.2	6.2	$(11.9, \quad 34.5)$	$(12.2, \quad 47.0)$
$\beta_{2,f}$	-2.10	1.1	$(-4.16, \quad -0.0414)$	$(-5.71, \quad -0.220)$
$\beta_{3,f}$	0.0742	0.043	$(-0.00283, \quad 0.151)$	$(-5.69 \times 10^{-4}, \quad 0.201)$
σ	0.996		$(0.583, \quad 1.70)$	$(1.38 \times 10^{-7}, \quad 1.43)$

TABLE 2. Maximum Likelihood Estimates of the Parameters in the Plague Model.

Parameter	Estimated Value	Asymptotic Standard Error	Asymptotic 95% C.I.
$\beta_{0,s}$	-2.06	2.5	$(-6.85, \quad 2.91)$
$\beta_{1,s}$	13.1	1.2	$(10.7, \quad 15.6)$
$\beta_{2,s}$	-0.349	0.084	$(-0.519, \quad -0.188)$
$\beta_{0,f}$	-14.5	1.5	$(-17.6, \quad -11.9)$
$\beta_{1,f}$	24.6	1.9	$(21.1, \quad 28.7)$
$\beta_{2,f}$	-1.46	0.32	$(-2.10, \quad -0.848)$
$\beta_{3,f}$	0.0519	0.012	$(0.0293, \quad 0.0757)$

TABLE 3. Maximum Likelihood Estimates of the Parameters in the Plague Model Fitted without Random Effects and without the Threshold Phenomenon.

random effects. Except $\beta_{0,s}$, the signs of the estimates are same as their counterparts of the GTR model. But the standard errors of the estimates are much smaller when the threshold effect is ignored, primarily because of deceptively higher sample size. Incorporating the random effects in the logistic regression model did not change the other parameter estimates but their standard errors became larger although still much smaller than those of the GTR model. Both logistic regression model fits were deemed inadequate based on model diagnostics, e.g. the plot of residuals versus fitted values did not look random. We conclude that the GTR model provides a better fit to the data from LSQ 105.

6. Conclusion

The real application illustrates the potential usefulness of the GTR model in analyzing epidemiological time series subject to a threshold condition. The GTR model specifies zero response in the lower regime. While this specification has a sound epidemiological justification, it is of interest to study the more general model that the non-normal response follows a generalized piecewise-linear model. Another interesting problem is to study how to pool information across different large squares in the Kazakhstan plague monitoring data, and to incorporate the spatial correlation structure. Given the relatively small number of non-zero bacteriological positive cases, the latter problem is pivotal for extracting the biological signal from the data and its spatial variation.

Acknowledgements

We thank Hildegunn Viljugrein for sharing her biological insights on the Kazakhstan monitoring data and for her help with the organization of the data. We are grateful to Ming-Ching Li, Howell Tong, and Henghsiu Tsai for helpful comments on an earlier draft. This work was partially supported by the European Union as projects ISTC K-159 and STEPICA (INCO-COPERNICUS, ICA 2-CT2000-10048), and the National Science Foundation (DMS-0405267), as well as the authors' institutes.

Appendix 1

Proof of Claim 1

Because d is assumed to be fixed, d is deleted from θ; i.e. let $\theta = (r, \beta', \psi')'$, and let $\ell(\theta)$ be the log likelihood of θ given by (4). Let

$$l(r) = \max_{\beta,\psi} \ell(\theta)$$

be the profile log likelihood of r. For any positive integer i, let $r_i = X_{(i)}$, and

$$(\hat{\beta}_i', \hat{\psi}_i')' = \arg\max_{\beta,\psi} \ell(\theta), \texttt{for a fixed } r = r_i.$$

Moreover, let $\theta_i = (r_i, \hat{\beta}_i', \hat{\psi}_i')'$, and $Z_i = (N_i, Y_i')'$. For convenience of notations, let $d = 0$.

We now verify that $l(r)$ increases for $r \leq \hat{r} = X_{(\hat{t})}$. This is trivial if $\hat{t} = 1$, hence we consider the case $\hat{t} \geq 2$. First consider $l(r_1) - l(r_2)$ where $r_i = X_{(i)} \leq \hat{r} = X_{(\hat{t})}, i = 1, 2$. With no loss of generality, assume $r_1 < r_2$. We have

$$\begin{aligned}
&l(r_1) - l(r_2) \\
&= \sum_{i=1}^{T} \log\left[E\left\{f(B_i|\epsilon_i, Z_i, \theta)|Z_i, \theta = \theta_1\right\}\right] \\
&- \sum_{j=1}^{T} \log\left[E\left\{f(B_j|\epsilon_j, Z_j, \theta)|Z_j, \theta = \theta_2\right\}\right] \\
&= \sum_{i=1}^{T} \log\left[E\left\{f(B_{(i)}|\epsilon_{(i)}, Z_{(i)}, \theta)|Z_{(i)}, \theta = \theta_1\right\}\right] \\
&- \sum_{j=1}^{1} \log(1) - \sum_{l=2}^{T} \log\left[E\left\{f(B_{(l)}|\epsilon_{(l)}, Z_{(l)}, \theta)|Z_{(l)}, \theta = \theta_2\right\}\right].
\end{aligned}$$

Since $r_1 < r_2 \leq \hat{r}$, then $B_{(1)} = 0$. Thus,

$$\begin{aligned}
l(r_1) - l(r_2) &= \log\left[E\left\{f(0|\epsilon_{(1)}, Z_{(1)}, \theta)|Z_{(1)}, \theta = \theta_1\right\}\right] &\text{(A1)} \\
&+ \sum_{i=2}^{T} \log\left[E\left\{f(B_{(i)}|\epsilon_{(i)}, Z_{(i)}, \theta)|Z_{(i)}, \theta = \theta_1\right\}\right] &\text{(A2)} \\
&- \sum_{l=2}^{T} \log\left[E\left\{f(B_{(l)}|\epsilon_{(l)}, Z_{(l)}, \theta)|Z_{(l)}, \theta = \theta_2\right\}\right] &\text{(A3)}
\end{aligned}$$

Now note that the term in (A1) is always strictly < 0, because when $\theta = \theta_1, 0 \leq f(0|\epsilon_{(1)}, Z_{(1)}, \theta) < 1 \Rightarrow 0 \leq E\left\{f(0|\epsilon_{(1)}, Z_{(1)}, \theta)|Z_{(1)}, \theta = \theta_1\right\} < 1$, entailing that its logarithm is negative. Moreover, the term in (A2) is less than or equal to minus the term in (A3), because the maximum likelihood estimator $(\hat{\beta}_2', \hat{\psi}_2')'$ is the argument that maximizes the following function (with β, ψ as the arguments)

$$\sum_{l=2}^{T} \log\left[E\left\{f(B_{(l)}|\epsilon_{(l)}, Z_{(l)}, \theta)|Z_{(l)}, \theta = r_2, \beta', \psi')'\right\}\right].$$

Hence, $l(r_1) - l(r_2) < 0$. Similar approach can be used to show the above for any two arbitrary thresholds that are less than or equal to $\hat{r} = X_{(\hat{t})}$. This completes the proof.

Appendix 2

Consistency of $\tilde{d}$

Recall that $\theta = (d, r, \beta', \psi')'$ denotes the generic parameter and θ_0 denotes the true parameter value. Let I denote the interior of the range of $\{X_t\}$. We claim that for any fixed $r \in I$ and $d \neq d_0$, the log likelihood at θ equals $-\infty$ with probability $\to 1$ as $T \to \infty$. Then, because $r_0 \in I$, $\exists \delta > 0$, such that $r_0 - \delta \in I$. Hence, $\hat{r}(d) < r_0 - \delta$ with probability $\to 1$ as $T \to \infty$. The preceding claim can be proved by noting that the expression inside the expectation on the right side of log likelihood equation (5) is bounded above by 1, because $\mu_t > 0$ so the conditional distribution of B_t is non-degenerate. Hence the first sum on the right side of (5) is bounded away from ∞. The validity of the preceding claim follows if we can show that for any fixed r in the range of $\{X_t\}$, the second term on the right side of (5) equals $-\infty$, with probability $\to 1$ as $T \to \infty$. In turn, this is true if $\sum_{i=1}^{T} I(X_{i-d} < r, B_i > 0) > 0$, with probability $\to 1$ as $T \to \infty$. But, $\sum_{i=1}^{T} I(X_{i-d} < r, B_i > 0) = \sum_{i=1}^{T} I(X_{i-d} < r, X_{i-d_0} \geq r_0, B_i > 0)$. By assumption A1, $\{(X_{i-d}, X_{i-d_0}, B_i)\}$ can be readily checked to be α-mixing. It follows from the Law of Large Numbers for α-mixing process that $\sum_{i=1}^{T} I(X_{i-d} < r, X_{i-d_0} \geq r_0, B_i > 0)/T \to P\left(X_{i-d} \leq r, X_{i-d_0} > r_0, B_i > 0\right)$, which is greater than zero, by assumption A3. This completes the proof that for $d \neq d_0$, $\exists \delta > 0$ such that $\hat{r}(d) < r_0 - \delta$ with probability $\to 1$ as $T \to \infty$.

Next, we show that $\hat{r}(d_0) \geq r_0$ almost surely. Consider the log likelihood evaluated at θ_0. The first term on the right side of (5) is clearly a finite number but the second term vanishes as $\sum_{t=1}^{T} I(X_{i-d_0} < r_0, B_i > 0)$ clearly equals 0, by model assumption. Consequently, $\hat{r}(d_0) \geq r_0$ almost surely. Thus, $\tilde{d} = d_0$ with probability $\to 1$ as $T \to \infty$.

Appendix 3

Proof of Theorem 2 – Consistency of $\hat{r}$

Without loss of generality, assume $d = 0$, and that some Bs are positive. By model definition, it is clear that $\hat{r} \geq r_0$. Then, it suffices to show that $\forall \delta > 0$, there exists $\tau > 0$ such that $P(\hat{r} - r_0 > \frac{\tau}{T}) < \delta$, for all T sufficiently large. Consider $P(\hat{r} - r_0 > \frac{\tau}{T})$. Note that if $\hat{r} > r_0 + \frac{\tau}{T}$, then all Xs that are $\leq r_0 + \frac{\tau}{T}$ have their corresponding Bs being zero. This implies that for any fixed $\tau > 0$,

$P\left\{T(\hat{r} - r_0) \leq \tau\right\}$

$= P(\hat{r} \leq r_0 + \frac{\tau}{T}) = P(\texttt{one or more } X\texttt{s in } \left[r_0, r_0 + \frac{\tau}{T}\right] \texttt{ with corresponding } B > 0)$.

Now let $C_t^T = A_t^T \cap \{B_t > 0\}$, where $A_t^T = \left\{X_t \in \left[r_0, r_0 + \frac{\tau}{T}\right]\right\}, t = 1, \cdots, T$. Then, $\left\{C_t^T\right\}$ can be readily checked to be α-mixing using assumption A1. Moreover, using assumption A4 and because $\pi(.)$ is assumed to be continuous at r_0, then as $T \to \infty$, $T\ P(C_t^T) = T\ P(A_t^T)P(B_t > 0\,|A_t^T) \to \pi(r_0)\ \tau\ p_0$, with $p_0 = P(B_t > 0\,|X_t = r_0) = 1 - P(B_t = 0\,|X_t = r_0) = 1 - E\left\{\int f(0|\epsilon_t, Z_t, \theta) f_\epsilon(\epsilon_t;\psi)\right.$ $\left. d\epsilon_t\,|X_t = r_0\right\}$.

Now let $U_t = (X_t, B_t)'$. If the Us were independent and identically distributed, $\sum_{t=1}^T I(C_t^T)$ has an asymptotic Poisson distribution with mean equal to $a = \pi(r_0)\ \tau\ p_0$. For $\psi-$mixing processes, we can apply the result of Meyer (1973) to show the asymptotic Poisson result. Meyer (1973) used an expanding blocking scheme, indexed by m, that blocks the data $\{U_t, t = 1, \cdots, T\}$ into blocks of alternate block sizes p_m and q_m, where m depends on T. The essential idea is that the larger blocks of size p_m are asymptotically independent with the smaller blocks of size q_m being negligible asymptotically so that we are back to the independent case for proving the asymptotic Poisson distribution. Meyer (1973) required that the triangular array of events $\left\{C_t^T\right\}, T = 1, 2, \cdots$, be α-mixing with mixing coefficients given by

$$\alpha_T(k) = \sup_{D \in \left\{C_1^T, \cdots, C_t^T\right\}, E \in \left\{C_{t+k+1}^T, \cdots\right\}} |P(D \cap E) - P(D)P(E)| \to 0 \tag{A4}$$

as $k \to \infty$, and the validity of the following technical condition.

C1. Suppose that there exist two sequences of block sizes $\{p_m, m = 1, 2, \cdots\}$ and $\{q_m, m = 1, 2, \cdots\}$ such that:

(1) for any fixed $s > 0$, $m^s \alpha_{t_m}(q_m) \to 0$ as $m \to \infty$; $t_m = m(p_m + q_m)$.

(2) $\frac{q_m}{p_m} \to 0$ as $m \to \infty$.

(3) $p_{m+1} \sim p_m$; i.e. $\lim_{m\to\infty} \frac{p_{m+1}}{p_m} = 1$.

(4) For any fixed $\tau > 0$, $I_{p_m} = \sum_{i=1}^{p_m - 1}(p_m - i)P(C_{i+1}^{t_m} \cap C_1^{t_m}) = o(\frac{1}{m})$ as $m \to \infty$.

We remark that condition C1 holds under assumptions A1–A3 and if we set $p_m = [T^\beta]$ and $q_m = [T^\gamma]$ where $[\cdot]$ denotes the largest integer not greater than the enclosed expression and $0 < \gamma < \beta < 1$; see below.

Hence, owing to the result of Meyer (1973), $\sum_{t=1}^T I(C_t^T)$ has an asymptotic Poisson distribution with mean equal to $\pi(r_0)\tau p_0$. Therefore, $P\left\{T(\hat{r} - r_0) \le \tau\right\} = P\left\{\sum_{t=1}^T I(C_t^T) \ge 1\right\} = 1 - P\left\{\sum_{t=1}^T I(C_t^T) = 0\right\} \to 1 - \exp^{-\pi(r_0)\tau p_0}$. Consequently, $T(\hat{r} - r_0)$ has a limiting exponential distribution with mean equal to $\frac{1}{\pi(r_0)p_0}$.

Verification of C1

Recall that we assume $d = 0$. First, we need to show that because the process $Z = \{Z_t\}$ is assumed to be α-mixing, then the process $\left\{C_t^T\right\} = \left\{A_t^T \cap \{B_t > 0\}\right\}$ is also α-mixing. This we can verify by considering

$$
\begin{aligned}
& P(C_t^T \cap C_{t+k}^T) - P(C_t^T)P(C_{t+k}^T) \\
& = E\left[I(A_t^T)I(A_{t+k}^T)E\left\{I(B_t > 0, B_{t+k} > 0)|Z\right\}\right] \\
& \quad - E\left[I(A_t^T)E\left\{I(B_t > 0)|Z_t\right\}\right] E\left[I(A_{t+k}^T)E\left\{I(B_{t+k} > 0)|Z_{t+k}\right\}\right] \\
& = E\left[I(A_t^T)I(A_{t+k}^T)E\left\{I(B_t > 0|Z)\right\} E\left\{I(B_{t+k} > 0)|Z\right\}\right] \\
& \quad - E\left[I(A_t^T)E\left\{I(B_t > 0)|Z_t\right\}\right] E\left[I(A_{t+k}^T)E\left\{I(B_{t+k} > 0)|Z_{t+k}\right\}\right] && \text{(A5)} \\
& = E\left[I(A_t^T)I(A_{t+k}^T)E\left\{I(B_t > 0|Z_t)\right\} E\left\{I(B_{t+k} > 0)|Z_{t+k}\right\}\right] \\
& \quad - E\left[I(A_t^T)E\left\{I(B_t > 0)|Z_t\right\}\right] E\left[I(A_{t+k}^T)E\left\{I(B_{t+k} > 0)|Z_{t+k}\right\}\right] && \text{(A6)}
\end{aligned}
$$

Equality (A5) is true because $B_t|\left\{Z, \epsilon\right\}$ are independent random variables. Since $\{Z_t\}$ is assumed to be α-mixing and using the covariance inequality of Bosq (1998, page 7), we can conclude that $\left\{C_t^T\right\}$ is α-mixing.

For a fixed integer m, partition the set of positive integers $\{1, 2, \cdots, T\}$ into consecutive blocks of size $p_m = T^\beta$ and $q_m = T^\gamma$ alternately, where $0 < \gamma < \beta < 1$, beginning with the initial block $\{1, 2, \cdots, p_m\}$ of size p_m. (More rigorously, we should write $p_m = \left[T^\beta\right]$ and $q_m = [T^\gamma]$, where [.] denotes the integral part of the expression inside the square bracket.) Hence, $m \sim \frac{T}{T^\beta + T^\gamma} = T^{1-\beta}O(1) \to +\infty$ as $T \to +\infty$. Furthermore, recall that $\alpha_T(k) = \frac{c\rho^k}{T}$. Note that $T \to \infty$ if and only if $m \to \infty$. We now verify the requirements in Condition C1. We have

- for any fixed $s > 0$, $m^s \alpha_{t_m}(q_m) = T^{s(1-\beta)-1} c\rho^{T^\gamma} \{1 + o(1)\} \to 0$ as $T \to +\infty$, because ρ is between 0 and 1.
- $\frac{q_m}{p_m} = T^{\gamma-\beta} \to 0$ as $T \to +\infty$, because $0 < \gamma < \beta < 1$.
- $\frac{p_{m+1}}{p_m} = \left[\frac{(m+1)^{\beta-1}\{1+o(1)\}}{m^{\beta-1}\{1+o(1)\}}\right]^\beta = \frac{(m+1)^{\beta(\beta-1)}}{m^{\beta(\beta-1)}} \{1 + o(1)\} \to 1$, as $m \to +\infty$.

- For any fixed $\tau > 0$, and an L to be determined later,

$$
\begin{aligned}
I_{T^\beta} &= \sum_{i=1}^{T^\beta-1}(T^\beta - i)P(C_{i+1}^T \cap C_1^T) \\
&= \sum_{i=1}^{L-1}(T^\beta - i)P(C_{i+1}^T \cap C_1^T) + \sum_{j=L}^{T^\beta-1}(T^\beta - j)P(C_{j+1}^T \cap C_1^T) \\
&\leq \sum_{i=1}^{L-1}(T^\beta - i)P(C_{i+1}^T \cap C_1^T) + \sum_{j=L}^{T^\beta-1}(T^\beta - j)\left\{\frac{c\rho^j}{T} + P(C_{j+1}^T)P(C_1^T)\right\} \\
&\leq \sum_{i=1}^{L-1}(T^\beta - i)O\left(\frac{1}{T^2}\right) + \sum_{j=L}^{T^\beta-1}(T^\beta - j)\left\{\frac{c\rho^j}{T} + P(A_{j+1}^T)P(A_1^T)\right\},
\end{aligned}
$$

because $\pi_{i,j}(.,.)$ is uniformly bounded and using the ψ-mixing conditions. Hence,

$$
\begin{aligned}
I_{T^\beta} &\leq (L-1)(T^\beta - \frac{L}{2})O\left(\frac{1}{T^2}\right) + \sum_{j=L}^{T^\beta-1} T^\beta \frac{c\rho^j}{T} + \frac{(T^\beta - L)(T^\beta - L + 1)}{2}O\left(\frac{1}{T^2}\right) \\
&\leq (L-1)(T^\beta - \frac{L}{2})O\left(\frac{1}{T^2}\right) + T^{\beta-1}\frac{c\rho^L(1-\rho^{T^\beta - L})}{1-\rho} + \frac{(T^\beta - L)(T^\beta - L + 1)}{2}O\left(\frac{1}{T^2}\right).
\end{aligned}
$$

Now, $0 \leq T^{1-\beta}I_{T^\beta} \leq O\left(T^{\beta-1}\right) + \frac{c\rho^L}{1-\rho}$. Given $\epsilon > 0, M = \frac{c}{1-\rho}$, take $L = \left[\frac{\ln\left(\frac{\epsilon}{2M}\right)}{\ln(\rho)}\right] + 1$, where $[.]$ denotes the integral part of the expression inside the square bracket. Thus, $\forall \epsilon > 0, T^{1-\beta}I_{T^\beta} < \epsilon$, for all T sufficiently large.

Appendix 4

Proof of Theorem 3

We first state and prove a lemma which is instrumental in the proof of Theorem 3.

Lemma 1. *Suppose assumptions A1 and A3 hold. Then* $E\left\{I\left(r_0 \leq X_{t-d} < \hat{r}\right)\right\} = o(1)$.

Proof of Lemma 1. It has been shown in Theorem 2 that $\hat{r} = r_0 + O_p(1)$. Then, $\forall \epsilon > 0, \exists K$ such that for all T sufficiently large, $|\hat{r} - r_0| \leq \frac{K}{T}$ with probability $\geq 1 - \epsilon$. Let H_T be the event defined by $H_T = \left\{ r_0 \le \hat{r} \le r_0 + \frac{K}{T} \right\}$. Then, $\forall \epsilon > 0$, we have

$$\begin{aligned} P\left(r_0 \le X_{t-d} < \hat{r}\right) &= P\left(r_0 \le X_{t-d} < \hat{r}, H_T\right) + P\left(r_0 \le X_{t-d} < \hat{r}, H_T^C\right) \\ &\le P\left(r_0 \le X_{t-d} < r_0 + \frac{K}{T}\right) + P\left(r_0 \le X_{t-d} < \hat{r}, \hat{r} > r_0 + \frac{K}{T}\right) \\ &\le P\left(r_0 \le X_{t-d} < r_0 + \frac{K}{T}\right) + P\left(\hat{r} > r_0 + \frac{K}{T}\right) \\ &< P\left(r_0 \le X_{t-d} < r_0 + \frac{K}{T}\right) + \epsilon, \ \texttt{for all } T \texttt{ sufficiently large;} \\ &= \int_{r_0}^{r_0 + \frac{K}{T}} \pi(x) dx + \epsilon, \ \texttt{for all } T \texttt{ sufficiently large.} \end{aligned} \tag{A7}$$

Since $\pi(.)$ is continuous at r_0 by assumption A3, then $\exists M > 0$ such that $|\pi(x)| \le M$ for all x in a neighborhood of r_0. Consequently, (A7) implies that $\forall \epsilon > 0$,

$$P\left(r_0 < X_{t-d} \le \hat{r}\right) < \frac{MK}{T} + \epsilon < 2\epsilon, \tag{A8}$$

for all T sufficiently large. This completes the proof of Lemma 1. □

We now prove Theorem 3. First we prove the existence of a consistent estimator $\hat{\delta}(\hat{r})$ of δ_0. The proof is similar to the proof of Theorem 5.42 in van der Vaart (2000) with some modifications. Making use of Lemma 1, we show below that $\Psi_T(\delta) \to \Psi(\delta) = E(\varphi_\delta)$ as $T \to \infty$. The true parameter δ_0 satisfies the equation $\Psi(\delta) = 0$. Let $\Psi_T^j(\delta)$ and $\Psi^j(\delta)$ be the $j^{\texttt{th}}$ component of $\Psi_T(\delta)$ and $\Psi(\delta)$ respectively. Similarly, denote the $j^{\texttt{th}}$ component of l_δ by l_δ^j. Furthermore, write $\dot{\Psi}^j(\delta_0) = \frac{\partial}{\partial \delta} \Psi^j(\delta) |_{\delta=\delta_0}$, and $\ddot{\Psi}^j(\delta_0) = \frac{\partial^2}{\partial \delta^2} \Psi^j(\delta) |_{\delta=\delta_0}$. Similarly defined are $\dot{\Psi}_T^j(\delta_0)$ and $\ddot{\Psi}_T^j(\delta_0)$. Using Taylor's theorem applied on the $j^{\texttt{th}}$ component of φ_δ, $\exists \tilde{\delta}$ between δ and δ_0 such that

$$E(\varphi_\delta^j) = E(\varphi_{\delta_0}^j) + E(\dot{\varphi}_{\delta_0}^j)(\delta - \delta_0) + \frac{1}{2}(\delta - \delta_0)' E(\ddot{\varphi}_{\tilde{\delta}}^j)(\delta - \delta_0),$$

where $\dot{\varphi}_{\delta_0}^j = \frac{\partial}{\partial \delta} \varphi_\delta^j |_{\delta=\delta_0}$, and $\ddot{\varphi}_{\tilde{\delta}}^j = \frac{\partial^2}{\partial \delta^2} \varphi_\delta^j \big|_{\delta=\tilde{\delta}}$. Note that $\left| E(\ddot{\varphi}_{\tilde{\delta}}^j) \right| \le E \left| \ddot{\varphi}_{\tilde{\delta}}^j \right| \le E|g| < \infty$, if δ is sufficiently close to δ_0. Hence, $E(\varphi_\delta^j)$ is differentiable at δ_0 for every j. By the same argument, $E(\varphi_\delta^j)$ is differentiable throughout a small neighborhood of δ_0. On the other hand, using Taylor's theorem applied on $\dot{\varphi}_\delta^j$, $\exists \tilde{\delta}$ between δ and δ_0 such that $E(\dot{\varphi}_\delta^j) = E(\dot{\varphi}_{\delta_0}^j) + E(\ddot{\varphi}_{\tilde{\delta}}^j)(\delta - \delta_0)$. As discussed earlier, $\left| E(\ddot{\varphi}_{\tilde{\delta}}^j) \right| < \infty$, if δ is sufficiently close to δ_0. Thus, $E(\dot{\varphi}_\delta^j)$ is continuous at a small neighborhood of δ_0 for every j. By assumption, $E(\dot{\varphi}_{\delta_0})$ is nonsingular. Then, we can make the neighborhood still smaller to ensure that $E(\dot{\varphi}_\delta)$ is nonsingular throughout the neighborhood. Then, by the inverse function theorem, for every sufficiently small $\gamma > 0$, there exists an open neighborhood G_γ of δ_0, and a ball B_γ centered at the origin with radius γ, such that the map $\Psi : \overline{G}_\gamma \mapsto \overline{B}_\gamma$ is a homeomorphism.

By the mean-value theorem and because the norms of the derivatives $\{E(\dot{\varphi}_\delta)\}^{-1}$ are bounded in a neighborhood of δ_0, $\exists\tilde{\delta}$ between δ and δ_0 such that $\left|\Psi^j(\delta) - \Psi^j(\delta_0)\right| = \left|\Psi^j(\delta)\right| = \left|E(\dot{\varphi}^j_{\tilde{\delta}})\right| |\delta - \delta_0|$, where Ψ^j is the j^{th} component of Ψ. Thus, the diameter of $\overline{G}_\gamma$ is bounded by a multiple of γ. That is, $|\delta - \delta_0|$ is bounded by a multiple of γ, for $\delta \in \overline{G}_\gamma$.

Now consider $\Psi^j_T(\delta) - \Psi^j(\delta)$. Using Taylor's theorem, we have

$$\Psi^j_T(\delta) = \Psi^j_T(\delta_0) + \dot{\Psi}^j_T(\delta_0)(\delta - \delta_0) + \frac{1}{2}(\delta - \delta_0)'\ddot{\Psi}^j_T(\tilde{\delta})(\delta - \delta_0); \tag{A9}$$

$$\Psi^j(\delta) = \Psi^j(\delta_0) + \dot{\Psi}^j(\delta_0)(\delta - \delta_0) + \frac{1}{2}(\delta - \delta_0)'\ddot{\Psi}^j(\tilde{\delta})(\delta - \delta_0). \tag{A10}$$

Combining equations (A9) and (A10), we get the following

$$\Psi^j_T(\delta) - \Psi^j(\delta) = \Psi^j_T(\delta_0) - \Psi^j(\delta_0) \tag{A11}$$

$$+ \left\{\dot{\Psi}^j_T(\delta) - \dot{\Psi}^j(\delta)\right\}(\delta - \delta_0) \tag{A12}$$

$$+ \frac{1}{2}(\delta - \delta_0)'\left\{\ddot{\Psi}^j_T(\tilde{\delta}) - \ddot{\Psi}^j(\tilde{\delta})\right\}(\delta - \delta_0). \tag{A13}$$

Now, the right-hand side term in (A11) can be written as

$$\begin{aligned}
\Psi^j_T(\delta_0) - \Psi^j(\delta_0) &= \Psi^j_T(\delta_0) - 0 \\
&= \frac{1}{T}\sum_{i=1}^{T} \dot{l}^j_{\delta_0}(B_i; Z_i) I\left(X_{i-d} \geq \hat{r}\right) \\
&= \frac{1}{T}\sum_{i=1}^{T} \dot{l}^j_{\delta_0}(B_i; Z_i) I\left(X_{i-d} \geq r_0\right) + \frac{1}{T}\sum_{i=1}^{T} \dot{l}^j_{\delta_0}(B_i; Z_i) I\left(X_{i-d} \geq \hat{r}\right) \\
&\quad - \frac{1}{T}\sum_{i=1}^{T} \dot{l}^j_{\delta_0}(B_i; Z_i) I\left(X_{i-d} > r_0\right).
\end{aligned}$$

Because $\hat{r} \geq r_0$, we have

$$\begin{aligned}
&\Psi^j_T(\delta_0) - \Psi^j(\delta_0) \\
&= \frac{1}{T}\sum_{i=1}^{T} \dot{l}^j_{\delta_0}(B_i; Z_i) I\left(X_{i-d} \geq r_0\right) - \frac{1}{T}\sum_{i=1}^{T} \dot{l}^j_{\delta_0}(B_i; Z_i) I\left(r_0 < X_{i-d} \leq \hat{r}\right).
\end{aligned} \tag{A14}$$

By the law of large numbers, $\frac{1}{T}\sum_{i=1}^{T} \dot{l}^j_{\delta_0}(B_i; Z_i) I\left(X_{i-d} \geq r_0\right)$ converges to 0 in probability. On the other hand, by the Lebesgue's dominated convergence theorem, for every j and i,

$$E\left[\left|\dot{l}^j_{\delta_0}(B_i; Z_i)\right| \; I\left\{\left|\dot{l}^j_{\delta_0}(B_i; Z_i)\right| \leq M\right\}\right] \to E\left\{\left|\dot{l}^j_{\delta_0}(B_i; Z_i)\right|\right\} < \infty \texttt{ as } M \to \infty, \tag{A15}$$

because $E\left|\dot{l}_{\delta_0}\right|^2$ is assumed to be finite. Consequently, $\forall\epsilon > 0$, there exists M such that

$$\begin{aligned}&\left|E\left\{\dot{l}^j_{\delta_0}(B_i;Z_i)I\left(r_0 \le X_{i-d} < \hat{r}\right)\right\}\right| \\ &< \epsilon + E\left[\left|\dot{l}^j_{\delta_0}(B_i;Z_i)\right| \; I\left\{\left|\dot{l}^j_{\delta_0}(B_i;Z_i)\right| \le M\right\} I\left(r_0 \le X_{i-d} < \hat{r}\right)\right] \\ &\le \epsilon + ME\left\{I\left(r_0 \le X_{i-d} < \hat{r}\right)\right\}. \qquad \text{(A16)}\end{aligned}$$

It follows from Lemma 1 and the results in (A14) and (A16) that

$$\Psi^j_T(\delta_0) - \Psi^j(\delta_0) = o_p(1). \qquad \text{(A17)}$$

It can be similarly shown that

$$\left|\dot{\Psi}^j_T(\delta) - \dot{\Psi}^j(\delta)\right| = o_p(1). \qquad \text{(A18)}$$

Finally, because of the assumption that the third-order partial derivatives of $l_\delta(z)$ are dominated by a fixed integrable function $g(z)$ for every δ in a neighborhood of δ_0, then

$$\left|\ddot{\Psi}^j_T(\tilde{\delta}) - \ddot{\Psi}^j(\tilde{\delta})\right| = O_p(1). \qquad \text{(A19)}$$

Consequently, combining the results in (A17)–(A19) with the results in (A11)–(A13), we get

$$\begin{aligned}\left|\Psi^j_T(\delta) - \Psi^j(\delta)\right| &\le o_p(1) + \left|\delta - \delta_0\right| o_p(1) + \left|\delta - \delta_0\right|^2 O_p(1) \\ &\le o_p(1) + \gamma o_p(1) + \gamma^2 O_p(1), \qquad \text{(A20)}\end{aligned}$$

for $\delta \in \overline{G}_\gamma$. Now, letting k be the dimension of δ, we have

$$\begin{aligned}\left|\Psi_T(\delta) - \Psi(\delta)\right| = \sqrt{\sum_{j=1}^{k}\left|\Psi^j_T(\delta) - \Psi^j(\delta)\right|^2} &\le \sqrt{k}\left\{\sum_{i=1}^{k}\left|\Psi^i_T(\delta) - \Psi^i(\delta)\right|\right\} \\ \Rightarrow \sup_{\delta\in\overline{G}_\gamma}\left|\Psi_T(\delta) - \Psi(\delta)\right| &\le \sqrt{k}\left[k\left\{o_p(1) + \gamma o_p(1) + \gamma^2 O_p(1)\right\}\right]. \qquad \text{(A21)}\end{aligned}$$

Because for every $\gamma > 0, P\left\{o_p(1) + \gamma o_p(1) > \frac{1}{2}\gamma\right\} \to 0$ as $T \to \infty$, there exists $\gamma_T \downarrow 0$ such that $P\left\{o_p(1) + \gamma_T o_p(1) > \frac{1}{2}\gamma_T\right\} \to 0$ as $T \to \infty$.

Let $K_{T,\gamma}$ be the event such that $\sup_{\delta\in\overline{G}_\gamma}\left|\Psi_T(\delta) - \Psi(\delta)\right| < \frac{\gamma}{2}$. Thus, $P(K_{T,\gamma_T}) \to 1$ as $T \to \infty$.

Now consider the map $\eta \mapsto \eta - \Psi_T \circ \Psi^{-1}(\eta)$. On the event $K_{T,\gamma}$, this map maps the ball $\overline{B}_\gamma$ into itself, by the definition of $\overline{G}_\gamma$ and $K_{T,\gamma}$. Because the map is also continuous, and by Brouwer's fixed-point theorem, there exists a fixed point in $\overline{B}_\gamma$ such that

$$\eta - \Psi_T \circ \Psi^{-1}(\eta) = \eta \Rightarrow \Psi_T \circ \Psi^{-1}(\eta) = 0 \Rightarrow \exists\delta \texttt{ such that } \Psi_T(\delta) = 0, \delta \in \overline{G}_\gamma.$$

Therefore, the probability that the equations $\Psi_T(\delta) = 0$ has at least one root tends to 1, as $T \to \infty$, and there exists a sequence of roots $\hat{\delta}(\hat{r})$ such that $\hat{\delta}(\hat{r}) - \delta_0(r_0)$ converges to 0 in probability. Finally, by Theorem 5.41 in van der Vaart (2000) and Remark 3, we get that

$$\hat{\delta}(\hat{r}) - \delta_0(r_0) = O_p(\frac{1}{\sqrt{T}}),$$

and the sequence $\sqrt{T}\left\{\hat{\delta}(\hat{r}) - \delta_0(r_0)\right\}$ is asymptotically normal with mean zero and covariance matrix $E(\dot{\varphi}_{\delta_0})^{-1}E(\varphi_{\delta_0}\varphi_{\delta_0}^{'})E(\dot{\varphi}_{\delta_0})^{-1}$.

References

Anderson, R.M. and May, R.M. (1991). *Infectious Diseases of Humans.* Oxford: Oxford University Press.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, (2nd edition). New York: Springer-Verlag.

Breslow, N.E. and Clayton, D.G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association* **88**, 9–25.

Chan, K.S. (1993). Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model. *The Annals of Statistics* **21**, 520–533.

Chan, K.S. and Tsay, R.S. (1998). Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model. *Biometrika* **85**, 413–426.

Davis, D., Begon, M., De Bruyn, L., Ageyev, V., Viljugrein, H., Stenseth, N. and Leirs, H. (2004). Predictive Thresholds for Plague in Kazakhstan. *Science* **304**, 736–738.

Davison, A.C. and Hinkley, D.V. (2003). *Bootstrap Methods and their Application.* Cambridge: Cambridge University Press.

Dickmann, O. and Heesterbeek, J. A. P. (2000). *Mathematical Epidemiology of Infectious Disease: Model Building, Analysis and Interpretation.* Chichester: John Wiley.

Doukhan, P. (1994). *Mixing: Properties and Examples.* New York: Springer-Verlag.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications.* Volume 2. New York: Wiley.

Gage K.L. and Kosoy M.Y. (2005). Natural history of plague: Perspectives from more than a century of research. Annu. Rev. Entomol. **50**, 505-28.

Grenfell, B. T. and Dobson, A. P. (1995). *Ecology of Infectious Diseases in Natural Populations.* New York: Cambridge University Press.

Hansen B.E. (2000). Sample splitting and threshold estimation. Econometrica **68**, 575-603.

Hudson, P.J., Rizzoli, A., Grenfell, B.T., Heesterbeek, H. and Dobson, A.P. (2002). *The Ecology of Wildlife Diseases.* New York: Oxford University Press.

Keeling, M.J. and Gilligan, C.A. (2000). Bubonic plague: a metapopulation model of a zoonosis. *Proceedings of the Royal Society of London Series B-Biological Sciences* **267** (1458), 2219–2230.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* 2nd edition. London: Chapman and Hall.

Meyer, R.M. (1973). A Poisson-Type Limit Theorem for Mixing Sequences of Dependent 'Rare' Events. *The Annals of Probability* **1**, 480 – 483.

Nummelin, E. (1984). *General irreducible Markov chains and non-negative operators.* Cambridge: Cambridge University Press.

Park, S., Chan, K.S., Viljugrein, H., Nekrassova, L., Suleimenov, B., Ageyev, V.S., Klassovskiy, N.L., Pole, S.B. and Stenseth, N.C. (2005). Modelling Immunity Against Infectious Plague Bacteria in Great Gerbils in Kazakhstan: a Continuous-time Markov Chain Approach. Under review.

Pinheiro, J.C. and Bates, D.M. (2002). *Mixed-Effects Models in S and S-PLUS.* New York: Springer-Verlag.

Tong, H. (1990). *Non-linear Time Series. A Dynamical System Approach.* New York: Oxford University Press.

van der Vaart, A.W. (2000). *Asymptotic Statistics.* Cambridge: Cambridge University Press.

Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S.* 4th edition. New York: Springer-Verlag.