MAXIMUM LIKELIHOOD ESTIMATION OF A GENERALIZED THRESHOLD MODEL

By Noelle I. Samia and Kung-Sik Chan

Northwestern University and University of Iowa

The open-loop Threshold Model, proposed by Tong [23], is a piecewise-linear stochastic regression model useful for modeling conditionally normal response time-series data. However, in many applications, the response variable is conditionally non-normal, e.g. Poisson or binomially distributed. We generalize the open-loop Threshold Model by introducing the Generalized Threshold Model (GTM). 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 introduce a likelihood-based estimation scheme for the GTM, and the consistency and limiting distribution of the maximum likelihood estimator are derived. A simulation study is conducted to illustrate the asymptotic results.

1. Introduction. The threshold autoregressive (TAR) model by Tong [22, 23] is perhaps the most popular nonlinear time-series models. Its extension that incorporates covariates is known as the open-loop threshold model (Tong [22]) which is a piecewise-linear stochastic regression model. While the model formulation of the threshold models does not impose the innovations to be normal, normality is generally the implicit assumption given that the threshold models specify a piecewise conditional mean structure.

However, in many applications including time-series response consisting of counts, the response variable is conditionally non-normal, e.g. Poisson or binomially distributed. Motivated by our recent works on modeling plague in Samia, Chan and Stenseth [17] and Samia *et al.* [18], we generalize the open-loop threshold model by introducing the Generalized Threshold Model (GTM). 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 through a known and invertible link function. On the other hand, the GTM is an extension of the generalized linear model (Nelder and Wedderburn [14], McCullagh and Nelder [13]), in which both non-normal re-

AMS 2000 subject classifications: Primary 62M10; secondary 62J12

Keywords and phrases: delay, exponential family, maximum likelihood estimation, stochastic regression, threshold

sponse distributions and piecewise linearity are accommodated. Hence, the link function is a natural device to remove any inherent constraints on the conditional mean function of a response variable, so that on the scale of the link function, the mean response is a piecewise-linear stochastic regression function. Note that if the link function is not the identity function, the conditional mean function of a GTM is generally piecewise nonlinear.

Threshold models may be estimated by various methods including conditional least squares and conditional maximum likelihood estimation. The conditional mean function of a threshold model is generally discontinuous, resulting in non-standard asymptotics for the estimators. Chan [3], Chan and Tsay [4], and Qian [16] established the asymptotic behavior of the threshold estimator in the threshold autoregressive models. Hansen [7] and Koul, Qian and Surgailis [9] studied the limiting behavior of the threshold estimator in the context of threshold regression models. We extend the previous asymptotic work to the GTM, where the vector of covariates may also contain lags of the response variable. However, because the conditional mean function of a GTM is generally piecewise nonlinear on the original scale, the ensuing complexity requires very different sets of regularity conditions and much innovations in the proof techniques than previous work for the threshold models.

The organization of the paper is as follows. Section 2 describes and formulates the model, namely the GTM. Section 3 presents the large-sample properties (i.e. consistency and limiting distribution) of the maximum likelihood estimator for the GTM. Section 4 conducts a simulation study that demonstrates the asymptotic theory for the GTM. Section 5 concludes by briefly discussing real motivating examples where the GTM is proven to be useful, and by discussing some related research problem. The proofs of the results stated in Section 3 are referred to Appendix A.

2. Model Formulation. Let $A = \{a_t, t = 1, \cdots, T\}$ be a positive process that may be the weights of the data cases, and let $X = \{x_t, t = 1, \cdots, T\}$ be a p-dimensional vector covariate process, where x_t consists of z_t and its lags, as well as some other covariates and their lags. The vector covariate x_t may also contain lags of the response variable. Denote by $\tilde{x}_t$ the part of x_t without the lags of the response variable. Define $\mathcal{F}_t$ as the σ-algebra generated by $y_{s-1}, x_s, s \le t$. Let y_t be a random variable whose conditional probability density function given A and $\mathcal{F}_t$, belongs to the (one-parameter canonical) exponential family, and takes the form

$$f(y_t; \gamma_t, a_t, \phi) = \exp\left[\frac{1}{\phi a_t}\{y_t\gamma_t - b(\gamma_t)\} + c\,(y_t; \phi a_t)\right], \tag{2.1}$$

where γ_t is the natural canonical parameter and ϕ is a dispersion parameter that is assumed to be known. The Generalized Threshold Model (GTM) specifies that conditional on A and $\mathcal{F}_t$, the conditional mean of y_t, denoted by μ_t, is linked to some piecewise-linear function

$$g(\mu_t) = \begin{cases} \beta_1^{'} x_t, & \text{if } z_{t-d} \leq r \\ \beta_2^{'} x_t, & \text{if } z_{t-d} > r; \end{cases} \tag{2.2}$$

$t = 1, \cdots, T$; and variances given by $\phi a_t v(\mu_t)$, where $v(.)$ is a specified variance function. The function g is a known invertible smooth link function with its inverse function denoted by g^{-1}. We assume that on the link scale, the model is discontinuous; i.e. the regression parameters are such that $\beta_1 \neq \beta_2$, β_1 and β_2 being $p \times 1$ vectors. The parameter r is known as the threshold and d is a non-negative integer referred to as the delay or threshold lag. For simplicity, we consider a two-regime model, but it can be easily extended to a multiple-regime model. The analysis of the above GTM is conditional on the observed a's, $\tilde{x}$'s, and $\mathcal{F}_1$. (We assume the initial values of y defining $\mathcal{F}_1$ are known.)

The parameter space Ω is $\Re^{2p} \times \Re \times \{0, 1, \cdots, D\}$, where D is a known upper bound of d, the delay parameter. A general parameter in the parameter space Ω is denoted by $\theta = (\beta_1', \beta_2', r, d)'$ and the true parameter $\theta_0 = (\beta_{1,0}', \beta_{2,0}', r_0, d_0)'$. The (conditional) log likelihood, in canonical form, is given by

$$l(\theta) = \sum_{t=1}^{T} \frac{1}{\phi a_t} \left\{ y_t \gamma_t - b(\gamma_t) \right\} + c\left(y_t; \phi a_t\right), \tag{2.3}$$

where $\dot{b}(\gamma_t) = \frac{\partial b(\gamma_t)}{\partial \gamma_t} = \mu_t$, and $\ddot{b}(\gamma_t) = \frac{\partial^2 b(\gamma_t)}{\partial \gamma_t^2} = v(\mu_t)$; see McCullagh and Nelder [13] and Firth [6]. Henceforth, $b(\gamma_t)$ is assumed to be a twice-differentiable function with positive second-order derivative, i.e. $b(\gamma_t)$ is strictly convex and $\dot{b}(\gamma_t)$ is a strictly monotone increasing function. In particular, since μ_t is a one-to-one function of γ_t, we can use μ_t as the parameter such that the log likelihood defined by (2.3) can be shown to equal

$$l(\theta) = \sum_{t=1}^{T} -\frac{1}{2\phi} d_t(y_t; \mu_t) + \ell_t(y_t), \tag{2.4}$$

where $d_t(y; \mu) = -2 \int_y^{\mu} \frac{y-u}{a_t v(u)} du$ is the deviance measure of fit, and $\ell_t(\mu_t)$ is the log likelihood for a single observation y_t given A and $\mathcal{F}_t$; see Breslow and Clayton [2].

Each distribution belonging to the exponential family has a unique canonical link function $\eta = \dot{b}^{-1}$ for which $\eta(\mu_t) = \gamma_t = \beta_1^{'} x_t I(z_{t-d} \leq r) + \beta_2^{'} x_t I(z_{t-d} > r)$, where $I(.)$ is the indicator function. Recall that as a result of the monotonicity of $\dot{b}$, the canonical parameter γ_t is a monotone function of μ_t. The canonical parameter space is generally either the real line, or a one-sided infinite interval, or an interval, depending on the distribution of the exponential family under consideration. In the case that the canonical parameter space is a proper subset of the real line, using the canonical link in the model is not attractive, in part because it puts restrictions on the parameter $\beta_i, i = 1, 2$. To avoid this issue, we shall assume that the link function (canonical or not) is such that the parameter $\beta_i, i = 1, 2$, is unconstrained and that $\gamma_t = w\{\beta_1^{'} x_t I(z_{t-d} \leq r) + \beta_2^{'} x_t I(z_{t-d} > r)\}$, where w is an increasing function. It is easy to check that $w = \eta \circ g^{-1}$, where η is the canonical link function and g is the link function considered in the model. Therefore, the log likelihood can be written as the sum of the log likelihoods of the two generalized linear submodels (in the lower and upper regimes) up to an additive constant, i.e.

$$l(\theta) = \sum_{t=1}^{T} M_{\beta_1}(y_t; a_t, x_t) I(z_{t-d} \leq r) + M_{\beta_2}(y_t; a_t, x_t) I(z_{t-d} > r) + c(y_t; \phi a_t),$$

where $M_{\beta_i}(y_t; a_t, x_t) = \frac{1}{\phi a_t}\{w(\beta_i^{'} x_t) y_t - b \circ w(\beta_i^{'} x_t)\}, i = 1, 2$.

Samia, Chan and Stenseth [17] studied the specific case where the non-negative discrete response variable equals zero in the lower regime; meaning that if the threshold is not met, the response is zero. While the latter model is of general applicability for analyzing epidemiological time series and other time-series data (Samia, Chan and Stenseth [17] and Stenseth *et al.* [20]), yet, we propose a more general form of the GTM that is obtained by (i) removing the restrictions on the positivity of the inverse link function and the discreteness and non-negativity of y_t, (ii) partitioning the sample space of (possibly vector-valued) z_{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, whenever $z_{t-d} \in R_i, i = 1, \cdots, m$.

3. Large-Sample Properties of the Estimator.

We first recall the notion of φ-mixing property. A stationary process $\{W_t\}$ is said to be φ-mixing if there exists a sequence of numbers $\{\varphi(k)\}$ with $\varphi(k) \to 0$ 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 $\{W_t, t \geq j + k\}$,

$$|P(E_1 \cap E_2) - P(E_1)P(E_2)| \leq \varphi(k)P(E_1). \tag{3.1}$$

See Billingsley [1, §20] and Doukhan [5, p. 3 and §1.3] for further discussion of φ-mixing.

In order to study the asymptotic properties of the estimator, the following set of assumptions will be required later. All expectations in the sequel are taken under the true model, unless stated otherwise.

(C1) The regression parameters are such that $\beta_{1,0} \neq \beta_{2,0}$. The cumulant function $b(\gamma_t)$ is strictly convex.

(C2) The process $\{(a_t, x_t^{'}, y_t)^{'}\}$ is stationary ergodic, having finite second moments.

(C3) The marginal probability density function of $\{x_t\}$ exists and is positive everywhere. Also, the process $Z = \{z_t\}$ 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 z, and $\pi(r_0) > 0$. The joint marginal probability density functions $\pi_{ij}(.,.)$ of $(z_i, z_j)^{'}$, for all $i \neq j$, are assumed to be positive everywhere and uniformly bounded. Also, a_t is uniformly bounded away from 0 and $+\infty$.

(C4) The parameter vector $\theta = (\beta_1^{'}, \beta_2^{'}, r, d)^{'}$ lies in a compact space $\Omega_1 \subseteq \Omega$, and Ω_1 contains the true parameter θ_0. The regression parameters $\beta_i, i = 1, 2$ are interior points of Ω_1.

(C5) Let $|\cdot|$ denote the absolute norm of the enclosed expression. Let $M_\beta(a_t, x_t, y_t) = \frac{1}{\phi a_t}\{\gamma_t y_t - b(\gamma_t)\}$ be the log likelihood for a single observation, namely y_t, where $\gamma_t = w(\beta^{'} x_t) = \eta \circ g^{-1}(\beta^{'} x_t)$, η being the canonical link function and g the link function considered in the model. There exist a square-integrable function $\tilde{w}$ and an integrable function $\tilde{m}$ such that $|w(\beta^{'} x) - w(\beta^{*'} x)| \leq \tilde{w}(x)|\beta - \beta^*|$ and $|b \circ w(\beta^{'} x) - b \circ w(\beta^{*'} x)| \leq \tilde{m}(x)|\beta - \beta^*|$, for every β, β^*, and x. Hence, if (C2) and (C3) hold, then there exists an integrable function $\Lambda(a_t, x_t, y_t)$ such that $|M_\beta(a_t, x_t, y_t) - M_{\beta^*}(a_t, x_t, y_t)| \leq \Lambda(a_t, x_t, y_t)|\beta - \beta^*|$, for every β, β^*, a_t, x_t, and y_t.

(C6) There exist $\tau > 0$ and $M > 0$ such that, for all $z_{t-d} \in [-\tau, \tau]$, $E\{\Lambda(a_t, x_t, y_t)^2 | z_{t-d}\} \leq M$.

(C7) There exists a $\Delta > 0$ such that the process $[\{\Lambda(a_t, x_t, y_t), z_{t-d} I(-\Delta \leq z_{t-d} \leq \Delta)\}^{'}]$ is φ-mixing with exponentially decaying mixing coefficients; i.e. for all $k \geq 0, |\varphi_k| \leq c\rho^k$ for some $c > 0$ and $0 \leq \rho < 1$.

(C8) The conditional distribution of $(a_t, x_t^{'})^{'}$ given $z_{t-d} = z$ is weakly continuous in z.

Remark 1. Without the assumption of $\left\{(a_t, x_t^{'}, y_t)^{'}\right\}$ being stationary ergodic, the consistency of the estimators may not be true as shown in Example 1 of Chan [3] which is a special case of a GTM with identity link and normal conditional distributions. The φ-mixing condition for the process $[\{\Lambda(a_t, x_t, y_t), z_{t-d}I(-\Delta \leq z_{t-d} \leq \Delta)\}']$ can be relaxed to a ρ-mixing condition; it is an essential condition for showing that the maximum likelihood estimator of the threshold has an $O_p(1/T)$ convergence rate. The assumption on the parameter space being compact can be removed or weakened in several cases of the GTM, as shown in the following Lemma 3.1. It is shown in Lemma 3.1, that the maximum likelihood estimator of the GTM is stochastically bounded in the case of a GTM with canonical link function and under a very mild condition on $\ddot{b}(.)$ that is generally true in all commonly used distributions of the exponential family. The proof of Lemma 3.1 is deferred to Appendix A.

LEMMA 3.1. *Assume that (C1)–(C3) hold. Assume, furthermore, that the link function considered in the model is the canonical link and the second derivative of the cumulant function* $b(.)$ *is such that either* $\ddot{b}(x+v) \geq \ddot{b}(x)$ *for all* x *and all* $v \geq 0$ *or* $\ddot{b}(x-v) \geq \ddot{b}(x)$ *for all* x *and all* $v \geq 0$. *Then, there exists* $\tau > 0$ *such that, for* T *sufficiently large, the maximum likelihood estimator* $\hat{\theta}_T$ *of the parameter vector* θ *lies in a compact space* $\Omega_1 = \{\theta \in \Omega : |\theta - \theta_0| \leq \tau\}$ *almost surely.*

The following Theorem 3.1 states the consistency of the maximum likelihood estimator $\hat{\theta}_T = (\hat{\beta}_1', \hat{\beta}_2', \hat{r}, \hat{d})'$, the proof of which is deferred to Appendix A.

THEOREM 3.1. *Assume that (C1)–(C5) hold. Then, the maximum likelihood estimator* $\hat{\theta}_T = (\hat{\beta}_1', \hat{\beta}_2', \hat{r}, \hat{d})'$ *is strongly consistent; that is,* $\hat{\theta}_T \to \theta_0$ *almost surely.*

Because of Theorem 3.1, it follows from the discreteness of the delay parameter that, for all sufficiently large T, $\hat{d} = d_0$ with probability 1. Thus, without loss of generality, we may and shall assume henceforth 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 3.2 that the maximum likelihood estimator of the threshold is T-consistent, whose proof is deferred to Appendix A. The $O_p(1/T)$ fast convergence rate is due to the discontinuity of the conditional mean function; see Chan [3], Chan and Tsay [4], and Hansen [7].

THEOREM 3.2. *Assume that (C1)–(C8) 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.*

Define $\delta = (\beta_1^{'}, \beta_2^{'})', \theta = (r, \delta^{'})'$. Let $l(\theta)$ be the log likelihood defined by (2.3), and let $\hat{\delta}_r = \arg\max_\delta l(\theta)$, for a fixed r. The log likelihood function of the GTM defined by (2.2), is given by

$$\begin{aligned} l(\theta) &= \sum_{t=1}^{T} \frac{1}{\phi a_t}\{w(\beta_1^{'}x_t)y_t - b \circ w(\beta_1^{'}x_t)\}I(z_{t-d} \le r) \\ &\quad + \frac{1}{\phi a_t}\{w(\beta_2^{'}x_t)y_t - b \circ w(\beta_2^{'}x_t)\}I(z_{t-d} > r) + c(y_t; \phi a_t) \\ &= \sum_{i=1}^{T} M_{\beta_1}(y_t; a_t, x_t)I(z_{t-d} \le r) + M_{\beta_2}(y_t; a_t, x_t)I(z_{t-d} > r) \\ &\quad + c(y_t; \phi a_t), \end{aligned}$$

where $M_{\beta_i}(y_t; a_t, x_t) = \frac{1}{\phi a_t}\{w(\beta_i^{'}x_t)y_t - b \circ w(\beta_i^{'}x_t)\}, i = 1, 2$.

Let $\psi_\delta(y_t; a_t, x_t) = \{\dot{M}_{\beta_1}^{'}(y_t; a_t, x_t)I(z_{t-d} \le r_0), \dot{M}_{\beta_2}^{'}(y_t; a_t, x_t)I(z_{t-d} > r_0)\}^{'}$,
where $\dot{M}_{\beta_i}(y_t; a_t, x_t) = \frac{\partial}{\partial \beta_i} M_{\beta_i}(y_t; a_t, x_t), i = 1, 2$. Define

(3.2)
$$\Psi_T(\delta) = \frac{1}{T}\sum_{i=1}^{T}\{\dot{M}_{\beta_1}^{'}(y_t; a_t, x)I(z_{t-d} \le \hat{r}), \dot{M}_{\beta_2}^{'}(y_t; a_t, x_t)I(z_{t-d} > \hat{r})\}^{'}.$$

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 GTM defined by (2.2) with known true threshold and delay, the maximum likelihood estimator equals $\hat{\delta}_{r_0}$ which is a root of the following estimating equation

$$\frac{1}{T}\sum_{i=1}^{T}\{\dot{M}_{\beta_1}^{'}(y_t; a_t, x_t)I(z_{t-d} \le r_0), \dot{M}_{\beta_2}^{'}(y_t; a_t, x_t)I(z_{t-d} > r_0)\}^{'} = 0. \tag{3.3}$$

The following conditions will be referred to in the sequel.

(D1) Let $\dot{M}_\beta(y; a, x) = \frac{\partial}{\partial \beta} M_\beta(y; a, x)$. The domain of δ is an open subset of the Euclidean space, in which $\beta_i \mapsto \dot{M}_{\beta_i}(y; a, x)$ is twice continuously differentiable for every $(y; a, x), i = 1, 2$.

(D2) Let $|\cdot|$ denote the absolute norm of the enclosed expression. Let $\ddot{M}_\beta(y; a, x) = \frac{\partial^2}{\partial \beta^2} M_\beta(y; a, x)$. For some neighborhood of $\beta_{i,0}$, say $V_i, i = 1, 2$, there exist two functions $m_1(y; a, x)$ and $m_2(y; a, x)$ such that $|\dot{M}_\beta(y; a, x)| \le m_1(y; a, x)$ and $|\ddot{M}_\beta(y; a, x)| \le m_2(y; a, x)$ for all y, a, x, and $\beta \in V_1 \cup V_2$. There exist $M_1, M_2 > 0$ and $\Delta > 0$ such that $E\{m_i(y_t; a_t, x_t)|z_{t-d}\} \le M_i$ for all $z_{t-d} \in [-\Delta, \Delta]$ and for $i = 1, 2$.

(D3) For some neighborhood of $\beta_{i,0}$, say $V_i, i = 1, 2$, the third-order partial derivatives of $M_\beta(y; a, x)$ with respect to β are dominated by a fixed integrable function $m_3(y; a, x)$ for every $\beta \in V_1 \cup V_2$.

(D4) $E\{\ddot{M}_{\beta_{1,0}}(y_t; a_t, x_t)I(z_{t-d} \le r_0)\}$ and $E\{\ddot{M}_{\beta_{2,0}}(y_t; a_t, x_t)I(z_{t-d} > r_0)\}$ exist and are nonsingular, where $\ddot{M}_{\beta_{i,0}}(y_t; a_t, x_t) = \frac{\partial^2}{\partial \beta_i^2} M_{\beta_i}(y_t; a_t, x_t)|_{\beta_i = \beta_{i,0}}, i = 1, 2$, and the expectation is taken under the true model.

Let $l(\theta)$ be the log likelihood of θ and let $l(., r)$ be globally maximized at $\hat{\delta}_r = (\hat{\beta}_{1,r}', \hat{\beta}_{2,r}')'$. The estimate of the threshold parameter r can be obtained by maximizing the profile log likelihood function $l(\hat{\delta}_r, r)$ of r. The optimization is conducted over the finite set of observed values of the threshold variable z_{t-d}. This is because, for a fixed delay d, the profile log likelihood function is constant between two consecutive sample percentiles of the threshold variable z_{t-d}. As a result of the strict convexity of $b(\gamma_t)$, the global maximum likelihood estimators $\hat{\beta}_{1,r}$ and $\hat{\beta}_{2,r}$ for a fixed threshold r (and a fixed delay d) can be attained by an exhaustive search with respect to the threshold variable z_{t-d}, subject to adequate number of data points in both regimes, e.g. number of data points in each regime is greater than $p + 1$, where p is the length of each of the regression coefficients β_1 and β_2.

We first state and prove the following two lemmas which are instrumental in the proof of the limiting distribution of the threshold estimator.

Lemma 3.2. *Assume that (C1)–(C8) and (D1)–(D4) hold. Then, for all* $K > 0$,

$$\sup_{|r - r_0| \le \frac{K}{T}} |\hat{\beta}_{i,r} - \hat{\beta}_{i,r_0}| = o_p(1/\sqrt{T}), \ i = 1, 2.$$

We now consider the limiting behavior of the normalized profile log likelihood. Define for $\kappa \in \Re$,

$$\tilde{l}(\kappa) = l(\hat{\delta}_{r_0 + \frac{\kappa}{T}}, r_0 + \kappa/T) - l(\hat{\delta}_{r_0}, r_0). \tag{3.4}$$

Lemma 3.3. *Assume that (C1)–(C8) and (D1)–(D4) hold. Then, for all* $K > 0$,

$$\sup_{|\kappa| \le K} |\tilde{l}(\kappa) - \{l(\delta_0, r_0 + \kappa/T) - l(\delta_0, r_0)\}| = o_p(1).$$

Next, we shall describe the limiting distribution of the threshold estimator $\hat{r}$. Consider two independent compound Poisson processes $\{\tilde{l}_1(\kappa), \kappa \ge 0\}$ and $\{\tilde{l}_2(\kappa), \kappa \ge 0\}$, both with rate $\pi(r_0)$, $\tilde{l}_1(0) = \tilde{l}_2(0) = 0$ a.s. and the distributions of jump being given by the conditional distribution of

$\xi_1 \doteq M_{\beta_{2,0}}(y_t; a_t, x_t) - M_{\beta_{1,0}}(y_t; a_t, x_t)$ given $z_{t-d} = r_0^-$ and the conditional distribution of $\xi_2 \doteq M_{\beta_{1,0}}(y_t; a_t, x_t) - M_{\beta_{2,0}}(y_t; a_t, x_t)$ given $z_{t-d} = r_0^+$, respectively. [We work with the left continuous version for $\tilde{l}_1(.)$ and the right continuous version for $\tilde{l}_2(.)$.] The former conditional distribution is the limiting conditional distribution of ξ_1 given $r_0 - \delta < z_{t-d} \le r_0$ as $\delta \downarrow 0$ and the latter that of ξ_2 given $r_0 < z_{t-d} \le r_0 + \delta$ as $\delta \downarrow 0$. We now state the following theorem.

THEOREM 3.3. *Assume that (C1)–(C8) and (D1)–(D4) hold. Then,* $(\{\tilde{l}(-\kappa), \kappa \ge 0\}, \{\tilde{l}(+\kappa), \kappa \ge 0\})$ *converges weakly to* $(\{\tilde{l}_1(\kappa), \kappa \ge 0\}, \{\tilde{l}_2(\kappa), \kappa \ge 0\})$ *in* $D[0,\infty) \times D[0,\infty)$, *the product space being equipped with the product Skorohod metric. Assume, furthermore, that* ξ_1 *and* ξ_2 *are continuous random variables. Then, the two random walks associated with the compound Poisson processes tend to* $-\infty$ *a.s. and hence,* $T(\hat{r} - r_0)$ *converges weakly to* M_- *where* $[M_-, M_+)$ *is the a.s. unique random interval of all* κ *at which* $\tilde{l}_1(-\kappa)I(\kappa < 0) + \tilde{l}_2(\kappa)I(\kappa \ge 0)$ *attains its global maximum.*

REMARK 2. We assume that ξ_1 and ξ_2 are continuous random variables to ensure that $\tilde{l}_1(-\kappa)I(\kappa < 0) + \tilde{l}_2(\kappa)I(\kappa \ge 0)$ attains its global maximum at the a.s. unique random interval $[M_-, M_+)$. In fact, this continuity assumption is generally true in many cases (e.g. the Poisson distribution.)

The super-consistency of the threshold parameter estimator, i.e. the $O_p(1/T)$ convergence rate, implies that under some regularity conditions, the threshold estimator is asymptotically independent of $\hat{\beta}_i, i = 1, 2$, which is the content of Theorem 3.4 below. Moreover, we show that $\hat{\beta}_1$ and $\hat{\beta}_2$ are $\sqrt{T}$-consistent and whose asymptotic joint distribution is identical to that for the case of known true delay and threshold, i.e. obtained from fitting the associated generalized linear model (GLM) defined by the equation $g(\mu_t) = \beta_1' x_t I(z_{t-d} \le r_0) + \beta_2' x_t I(z_{t-d} > r_0)$.

THEOREM 3.4. *Assume that (C1)–(C8) and (D1)–(D4) hold. Then,*

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

and the sequence $\sqrt{T}(\hat{\delta}_{\hat{r}} - \delta_0)$ *is asymptotically normal with mean zero and covariance matrix* $\Sigma = E(\dot{\psi}_{\delta_0})^{-1} E(\psi_{\delta_0} \psi_{\delta_0}') E(\dot{\psi}_{\delta_0})^{-1}$.

REMARK 3. As a result of Σ being a block diagonal matrix, the regression parameter estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are asymptotically independent of each other.

4. Simulation Study. We conduct a simulation study to illustrate the asymptotic results of the GTM defined by (2.2). Conditionally independent observations of y_t are generated from Poisson distributions with mean μ_t given by

$$\log(\mu_t) = \begin{cases} \beta_{10} + \beta_{11} x_t, & \text{if } z_{t-d} \leq r \\ \beta_{20} + \beta_{21} z_{t-1}, & \text{if } z_{t-d} > r; \end{cases} \tag{4.1}$$

$t = 1, \cdots, T$. The parameters d and r are taken to be 0 and 0.38, respectively. The regression coefficients are fixed at $\beta_{10} = 0.4, \beta_{11} = 1, \beta_{20} = 1.5$, and $\beta_{21} = 0.5$. The threshold variable z_t is generated as a series that follows an AR(2) process given by $z_t = \frac{w_t + 0.907}{2.37}$, where $w_t = 0.9255 w_{t-1} - 0.2736 w_{t-2} + \sqrt{0.02125}\, \eta_t$, and η_t denotes a series of uncorrelated normal random variables with zero mean and variance 1, truncated between -3 and 3. Note that z_t is bounded between 0 and 1. The covariate x_t is generated as a series of independent Uniform(0, 1) random variables. The sample sizes used are 50, 100, and 200, and for each sample size, the results are based on 1000 replications.

The estimators of the threshold parameter r and the delay parameter d are obtained by maximizing the log likelihood of the estimated GTM, with the delay being an integer between 0 and 2, and the search of the threshold done based on an exhaustive search with respect to z_{t-d}, where each regime has at least 4 data points. For given estimates of the threshold and the delay, the associated generalized linear submodels are estimated using the glm function in R; see Venables and Ripley [25].

Table 1 gives the percentage of times the threshold delay was estimated to be equal to the true value 0 and the percentage of times optimization failed. We also report in Table 1 the sample means, bias, and standard deviations of the estimates, and the empirical coverage probabilities of the regression parameters. All of the latter estimates and probabilities reported in Table 1 are based on fitting the GTM with the delay fixed at its true value 0. The empirical coverage probabilities are based on the 95% confidence intervals of the corresponding regression parameters.

In general, the percentage of times the threshold delay was estimated to be equal to 0, increases with larger sample size. The percentage of times optimization failed decreases with larger sample size. The standard deviation and the bias of the estimators generally become smaller with larger sample size, confirming the consistency results discussed previously. Moreover, the empirical coverage probabilities get generally closer to the nominal coverage probabilities with increasing sample sizes.

The Q-Q plots of the $\hat{\beta}$'s for sample sizes 100 and 200 confirm the asymptotic normality of the regression estimators in the associated generalized linear submodels, see Figure 1 where we show the results for $T = 100$, as the Q-Q plots for $T = 200$ are similar. For $T = 50$, the Q-Q plots show some departure from normality, which can be circumvented by restricting the search of the threshold to be between two predetermined percentiles of the threshold variable; say, between the 20$^{\text{th}}$ and 80$^{\text{th}}$ percentiles.

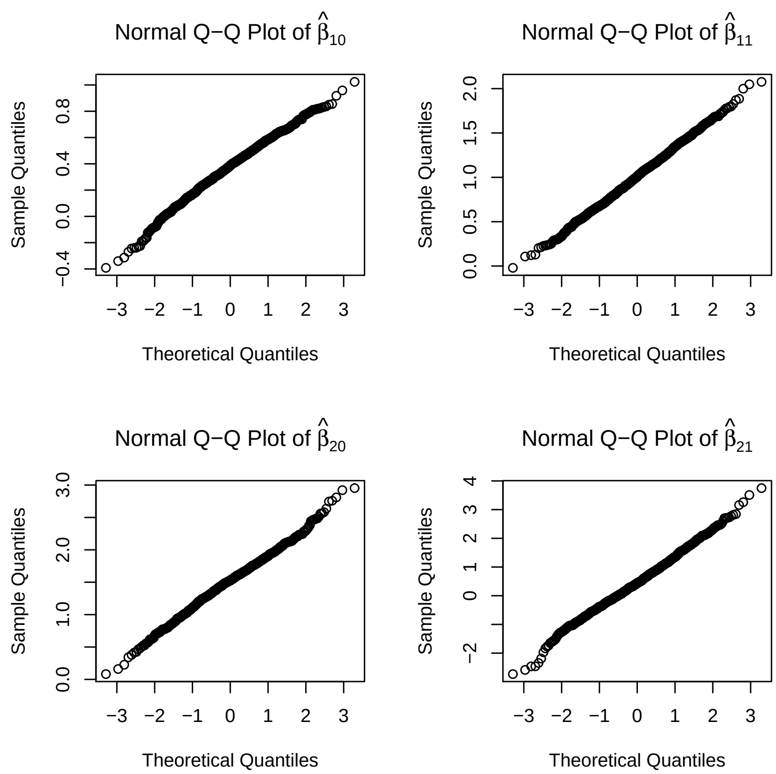

FIG 1. *Q-Q Plots for the Case when Sample Size = 100.*

5. Conclusion. The GTM is motivated by the need for modeling the dynamics of outbreaks of plague caused by the bacterium *Yersinia pestis* in humans in Kazakhstan. It is of interest to explain the sporadic occurrences of plague in humans using the information provided from systematic sampling of fleas and rodents during the study period. In particular, Samia *et al.* [18] showed that a sufficient number of viable fleas has to be achieved in order for the major human outbreaks to occur. However, below the critical

Sample Size	% of $\hat{d} = 0$ (in %)	% of Failures (in %)		Parameter Estimates $\hat{r}$	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	Coverage Probability of β_{10}	β_{11}	β_{20}	β_{21}
50	77.5	1.1		0.376	0.350	1.04	1.54	0.422	0.929	0.926	0.942	0.945
			sd	0.0200	0.522	0.752	0.572	1.31				
			bias	-0.00371	-0.0504	0.0386	0.0413	-0.0780				
100	97.7	0.8		0.379	0.376	1.0086	1.52	0.469	0.956	0.958	0.960	0.962
			sd	0.00830	0.209	0.327	0.411	0.906				
			bias	-0.00115	-0.0241	0.00862	0.0164	-0.0312				
200	99.9	0.5		0.379	0.392	1.003	1.51	0.484	0.948	0.949	0.948	0.952
			sd	0.00389	0.153	0.234	0.317	0.698				
			bias	-0.000678	-0.00830	0.00259	0.00824	-0.0164				
			True	0.38	0.40	1.0	1.5	0.50				

TABLE 1. *Results of the Simulation Study.*

threshold which is a proxy to the number of viable fleas in the area, sporadic cases of human plague outbreaks may occur. These findings are linked to the classical theory of the basic reproductive ratio. For further developments of this application, see Samia *et al.* [18].

In addition, the GTM is useful in modeling many other biological systems that undergo different dynamics; e.g. climate changes, Chitty hypothesis (Krebs [10]). The usefulness of the GTM can be widely adapted for use in diverse fields including natural sciences, marketing, economics, political science, and business.

An interesting future research problem is to allow the dispersion parameter ϕ to be regime-dependent, which introduces conditional heteroscedasticity in the GTM.

APPENDIX A: PROOFS

A.1. Proof of Lemma 3.1.

Proof. Without loss of generality, the delay parameter d is assumed to be known and $d = 0$. The proof can be easily extended to the general case where d is unknown and $0 \leq d \leq D$, D being a known upper bound of the delay parameter. The parameter vector becomes $\theta = (\beta_1^{'}, \beta_2^{'}, r)^{'}$, and the true parameter is denoted by $\theta_0 = (\beta_{1,0}^{'}, \beta_{2,0}^{'}, r_0)^{'}$. Let $l(\theta)$ be the log likelihood of θ. For ease of exposition, we first impose the restriction that $r \geq r_0$. The proof for the case $r \leq r_0$ is similar and hence is omitted. We have

$$\frac{l(\theta) - l(\theta_0)}{T} = R_{1,t} + R_{2,t} + R_{3,t}, \tag{A.1}$$

where

$$R_{1,t} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}\left\{(\beta_1 - \beta_{1,0})^{'} x_t y_t - b(\beta_1^{'} x_t) + b(\beta_{1,0}^{'} x_t)\right\} \times I(z_t \leq r_0), \tag{A.2}$$

$$R_{2,t} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}\left\{(\beta_1 - \beta_{2,0})^{'} x_t y_t - b(\beta_1^{'} x_t) + b(\beta_{2,0}^{'} x_t)\right\} \times I(r_0 < z_t \leq r), \tag{A.3}$$

$$R_{3,t} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}\left\{(\beta_2 - \beta_{2,0})^{'} x_t y_t - b(\beta_2^{'} x_t) + b(\beta_{2,0}^{'} x_t)\right\} \times I(z_t > r). \tag{A.4}$$

Our proof relies on verifying the following two claims.
Claim 1: There exists $M > 0$ such that, for T sufficiently large, $\hat{\theta}_T$ lies in

$$\mathcal{C}_1 = \{\theta \in \Omega : |\beta_1 - \beta_{1,0}| \le M, |\beta_2 - \beta_{2,0}| \le M\} \text{ a.s.} \tag{A.5}$$

Claim 2: There exists $\Delta > 0$ such that, for T sufficiently large, $\hat{\theta}_T$ lies in

$$\mathcal{C}_2 = \{\theta \in \mathcal{C}_1 : |r - r_0| \le \Delta\} \text{ a.s.} \tag{A.6}$$

Throughout the proof, the uniform law of large numbers will be applied a number of times, the validity of which can be routinely checked using Theorem 2 of Pollard [15, p. 8]. Although Pollard [15] assumes that the data are independent and identically distributed, this assumption can be relaxed to assuming a stationary ergodic process; see Pollard [15, p. 9]. A prototype of such checking is given at the end of the proof of Claim 2.

Verification of Claim 1: It suffices to show that for T sufficiently large and uniformly for $\theta \notin \mathcal{C}_1$, we have $\frac{l(\theta)-l(\theta_0)}{T} < 0$ almost surely. Without loss of generality, we consider the case that $r \ge r_0$ as the case $r \le r_0$ can be similarly dealt with. Note that $\{\theta \notin \mathcal{C}_1\}$ can be written as the union of $\mathcal{A}_1, \mathcal{A}_2$, and $\mathcal{A}_3$, where for some $M > 0$ and $\Delta > 0$ (to be determined), we have

$$\mathcal{A}_1 = \{\theta : |\beta_1| + |\beta_2| \ge M, r_0 \le r \le r_0 + \Delta\}; \tag{A.7}$$

$$\mathcal{A}_2 = \{\theta : |\beta_1| \ge M, r \ge r_0 + \Delta\}; \tag{A.8}$$

$$\mathcal{A}_3 = \{\theta : |\beta_1| \le M, |\beta_2| \ge M, r \ge r_0 + \Delta\}. \tag{A.9}$$

Consequently, the proof of Claim 1 is divided into the following three cases, namely $\theta \in \mathcal{A}_i, i = 1, 2, 3$.

Let $\theta \in \mathcal{A}_1 = \{\theta : |\beta_1| + |\beta_2| \ge M, r_0 \le r \le r_0 + \Delta\}$. It suffices to show that for T sufficiently large and uniformly for $\theta \in \mathcal{A}_1$, we have $\frac{l(\theta)-l(\theta_0)}{T} < 0$ almost surely. Let $\nu(\beta) = \frac{\beta}{|\beta|}$; i.e. $\nu(\beta)$ is on the boundary of the unit sphere centered at the origin. By expanding $R_{1,t}$ defined by (A.2) in a Taylor series expansion around $\beta_{1,0}^{'} x_t$, we have

$$\begin{aligned} R_{1,t} \le{}& |\beta_1 - \beta_{1,0}| \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\phi a_t} \nu(\beta_1 - \beta_{1,0})^{'} x_t \{y_t - \dot{b}(\beta_{1,0}^{'} x_t)\} I(z_t \le r_0) \\ & - |\beta_1 - \beta_{1,0}|^2 \frac{1}{T} \sum_{t=1}^{T} \frac{1}{2\phi a_t} \{\nu(\beta_1 - \beta_{1,0})^{'} x_t\}^2 \ddot{b}(\kappa_{1,t}) I(z_t \le r_0), \end{aligned}$$

for some $\kappa_{1,t}$ between $\beta_1^{'} x_t$ and $\beta_{1,0}^{'} x_t$. Without loss of generality, suppose that $\ddot{b}(\beta_{1,0}^{'} x_t + v) \ge \ddot{b}(\beta_{1,0}^{'} x_t)$ for all $v \ge 0$. Then, because $\ddot{b}(.)$ is a strictly positive function, we get

$$
\begin{aligned}
R_{1,t} \leq |\beta_1 - \beta_{1,0}| \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\phi a_t} \nu (\beta_1 - \beta_{1,0})' x_t \{ y_t - \dot{b}(\beta_{1,0}' x_t)\} I(z_t \leq r_0) \\
- |\beta_1 - \beta_{1,0}|^2 \frac{1}{T} \sum_{t=1}^{T} \frac{1}{2\phi a_t} \{\nu(\beta_1 - \beta_{1,0})' x_t\}^2 \ddot{b}(\beta_{1,0}' x_t) \\
I\{\nu(\beta_1 - \beta_{1,0})' x_t \geq 0\} I(z_t \leq r_0).
\end{aligned}
$$

Similarly, expand $R_{2,t}$ and $R_{3,t}$ in a Taylor series around $\beta_{2,0}' x_t$.

Then, by applying the uniform law of large numbers, for all $\epsilon > 0$, it holds almost surely that, for T sufficiently large and uniformly for $|\beta_1| + |\beta_2| \geq M, r_0 \leq r \leq r_0 + \Delta$, we have

$$
\begin{aligned}
\frac{l(\theta) - l(\theta_0)}{T} &\leq (|\beta_1 - \beta_{1,0}| + |\beta_1 - \beta_{2,0}| + |\beta_2 - \beta_{2,0}|)\, \epsilon \\
&\quad + \left(|\beta_1 - \beta_{1,0}|^2 + |\beta_1 - \beta_{2,0}|^2 + |\beta_2 - \beta_{2,0}|^2 \right) (\rho + \epsilon),
\end{aligned}
\tag{A.10}
$$

where

$$
\rho = \max_{\beta_1, \beta_2, r} J(\beta_1, \beta_2, r),
\tag{A.11}
$$

where the maximum is taken over all β_1, β_2, r such that $|\nu(\beta_1 - \beta_{1,0})| = |\nu(\beta_1 - \beta_{2,0})| = |\nu(\beta_2 - \beta_{2,0})| = 1$ and $r_0 \leq r \leq r_0 + \Delta$, and with

$$
\begin{aligned}
&J(\beta_1, \beta_2, r) = \\
&\Bigg(E\left[-\frac{1}{2\phi a_t} \{\nu(\beta_1 - \beta_{1,0})' x_t\}^2 \ddot{b}(\beta_{1,0}' x_t) I\{\nu(\beta_1 - \beta_{1,0})' x_t \geq 0\} I(z_t \leq r_0) \right], \\
&E\left[-\frac{1}{2\phi a_t} \{\nu(\beta_1 - \beta_{2,0})' x_t\}^2 \ddot{b}(\beta_{2,0}' x_t) I\{\nu(\beta_1 - \beta_{2,0})' x_t \geq 0\} I(r_0 < z_t \leq r) \right], \\
&E\left[-\frac{1}{2\phi a_t} \{\nu(\beta_2 - \beta_{2,0})' x_t\}^2 \ddot{b}(\beta_{2,0}' x_t) I\{\nu(\beta_2 - \beta_{2,0})' x_t \geq 0\} I(z_t > r) \right] \Bigg).
\end{aligned}
$$

Each component of $J(\beta_1, \beta_2, r)$ is continuous for every θ and attains its maximum on a compact set given by $\{|\nu(\beta_1 - \beta_{1,0})| = |\nu(\beta_1 - \beta_{2,0})| = |\nu(\beta_2 - \beta_{2,0})| = 1, r_0 \leq r \leq r_0 + \Delta\}$. Because the marginal probability density function of x_t is assumed to be positive everywhere, and since $\ddot{b}(.) > 0$, then ρ is strictly negative. We conclude that, for T sufficiently large and for sufficiently large M, and uniformly for $r_0 \leq r \leq r_0 + \Delta$, we have $\frac{l(\theta) - l(\theta_0)}{T} < 0$ almost surely. The latter result holds by fixing $\epsilon = \frac{-\rho}{8}$ in which case the right-hand side of the inequality in (A.10) is strictly negative.

This can be seen by letting $A = |\beta_1 - \beta_{1,0}| + |\beta_1 - \beta_{2,0}| + |\beta_2 - \beta_{2,0}|$ and $B = |\beta_1 - \beta_{1,0}|^2 + |\beta_1 - \beta_{2,0}|^2 + |\beta_2 - \beta_{2,0}|^2$ and by making use of Cauchy-Schwartz inequality where $A \leq \sqrt{3}\sqrt{B} \leq 3B$ for sufficiently large M, and after routine algebra.

Employing similar arguments and techniques when $\theta \in \mathcal{A}_2$ and $\theta \in \mathcal{A}_3$, it is readily checked that Claim 1 holds.

Verification of Claim 2: Let $\theta \in \mathcal{A}_4 = \{\theta : |\beta_1| + |\beta_2| \leq M, r \geq r_0 + \Delta, M > 0, \Delta > 0\}$. Applying the uniform law of large numbers to (A.1), we conclude that, it holds almost surely that, for all $\epsilon > 0$, for T sufficiently large, and uniformly for $r \geq r_0 + \Delta$ and $|\beta_1| + |\beta_2| \leq M$, we have

$$\begin{aligned}
&\frac{l(\theta) - l(\theta_0)}{T} \\
&\leq E\left[\frac{1}{\phi a_t}\left\{(\beta_1 - \beta_{1,0})' x_t y_t - b(\beta_1' x_t) + b(\beta_{1,0}' x_t)\right\} \times I(z_t \leq r_0)\right] \\
&\quad + E\left[\frac{1}{\phi a_t}\left\{(\beta_1 - \beta_{2,0})' x_t y_t - b(\beta_1' x_t) + b(\beta_{2,0}' x_t)\right\} \times I(r_0 < z_t \leq r)\right] \\
&\quad + E\left[\frac{1}{\phi a_t}\left\{(\beta_2 - \beta_{2,0})' x_t y_t - b(\beta_2' x_t) + b(\beta_{2,0}' x_t)\right\} \times I(z_t > r)\right] + \epsilon.
\end{aligned}$$

Thus, using the result of Kullback-Leibler divergence stated in Lemma 5.35 of van der Vaart [24], for all $\epsilon > 0$, it holds almost surely that, for T sufficiently large and for any fixed $\Delta > 0$, and uniformly for $r \geq r_0 + \Delta$ and $|\beta_1| + |\beta_2| \leq M$, we have $\frac{l(\theta)-l(\theta_0)}{T} \leq \kappa + \epsilon$, where

$$\kappa = \max_{|\beta_1|+|\beta_2| \leq M, r \geq r_0+\Delta} H(\beta_1, \beta_2, r)$$

and

$$\begin{aligned}
&H(\beta_1, \beta_2, r) \\
&= E\left[\frac{1}{\phi a_t}\left\{(\beta_1 - \beta_{1,0})' x_t y_t - b(\beta_1' x_t) + b(\beta_{1,0}' x_t)\right\} \times I(z_t \leq r_0)\right] \\
&\quad + E\left[\frac{1}{\phi a_t}\left\{(\beta_1 - \beta_{2,0})' x_t y_t - b(\beta_1' x_t) + b(\beta_{2,0}' x_t)\right\} \times I(r_0 < z_t \leq r)\right].
\end{aligned}$$

Note that $\kappa < 0$. To see this, observe that $H(\beta_1, \beta_2, r)$ can be extended continuously to $H(\beta_1, \beta_2, \infty) = \lim_{r \to \infty} H(\beta_1, \beta_2, r)$, by the dominated convergence theorem. (When $r = \infty$, the model becomes a generalized linear model with β_2 being absent from the model.) Hence, $\{|\beta_1| + |\beta_2| \leq M, r \geq r_0 + \Delta\}$ is a compact set of the extended parameter space $\Re^{2p} \times \overline{\Re}$, where $\overline{\Re} = \Re \cup \{-\infty, +\infty\}$ is equipped with the metric $d(x, y) = |\arctan(x) - \arctan(y)|$. Thus, $H(\beta_1, \beta_2, r)$ is continuous for every θ and attains its maximum on a compact set. Using the result of Kullback-Leibler divergence stated in Lemma 5.35 of van der Vaart [24], and because of the model assumption that $\beta_{1,0} \neq \beta_{2,0}$, we note that $H(\beta_1, \beta_2, r)$ and $H(\beta_1, \beta_2, \infty)$ are < 0; and hence, $\kappa < 0$. Choose $\epsilon = -\frac{\kappa}{2}$, so that it holds almost surely that, for T sufficiently large and for any fixed $\Delta > 0$, and uniformly for $r \geq r_0 + \Delta$, and $|\beta_1| + |\beta_2| \leq M$, $\frac{l(\theta) - l(\theta_0)}{T} \leq \frac{\kappa}{2} < 0$.

Finally, we check the approximating conditions of the uniform law of large numbers in Theorem 2 of Pollard [15, p. 8], applied in the proof of Claim 2. Although Pollard [15] assumes that the data are independent and identically distributed, this assumption can be relaxed to assuming a stationary ergodic process; see Pollard [15, p. 9]. A prototype of such checking is given below for the class $\mathcal{H}$ of functions of the form $h = h_{\beta_1,r}(a_t, x_t, y_t) = \frac{1}{\phi a_t}\left\{(\beta_1 - \beta_{2,0})' x_t y_t - b(\beta_1' x_t) + b(\beta_{2,0}' x_t)\right\} \times I(r_0 < z_t \leq r)$.

Let $\mathcal{F}$ be the collection of all indicator functions of the form $f = f_r(a_t, x_t, y_t) = VI(r_0 < z_t \leq r)$, where $V = V(a_t, x_t, y_t) \geq 1$ is a fixed function and $E(V) < \infty$, and with r ranging over $\overline{\Re}$. Note that $\frac{1}{T}\sum_{t=1}^{T} f = \frac{1}{T}\sum_{t=1}^{T} VI(r_0 < z_t \leq r)$ is a non-decreasing function of r; and for fixed r, $\frac{1}{T}\sum_{t=1}^{T} f$ converges to $E(f)$ wich is also a non-decreasing function of r. Also, $\lim_{r\to\infty} E(f) = E\{VI(z_t > r_0)\} \leq E(V)$. Therefore, we consider brackets of the form $[f_{\epsilon,L}, f_{\epsilon,U}]$ such that $f_{\epsilon,L} = VI(r_0 < z_t \leq r_{i-1}^*)$ and $f_{\epsilon,U} = VI(r_0 < z_t < r_i^*)$, for a grid of points $r_0 = r_0^* < r_1^* < \cdots < r_k^* = \infty$, with the property $E(f_{\epsilon,U} - f_{\epsilon,L}) = \frac{\epsilon}{2} < \epsilon$ for each i. Thus, the total number of brackets can be chosen to be $k = \frac{2E(V)}{\epsilon}$. These brackets have $L_1(P)$-size ϵ. Note that the collection of functions of the form $I(r_0 < z_t \leq r)$ is a special case of $\mathcal{F}$; and hence, it satisfies the finite bracketing condition.

On the other hand, let $\mathcal{G}$ be the collection of functions of the form $g = g_{\beta_1}(a_t, x_t, y_t) = \frac{1}{\phi a_t}\left\{(\beta_1 - \beta_{2,0})' x_t y_t - b(\beta_1' x_t) + b(\beta_{2,0}' x_t)\right\}$, where $|\beta_1| \leq M$. Because of the Lipschitz property of b stated in (C5), it is easy to check that $\mathcal{G}$ has an integrable envelope function G given by $\frac{1}{|\phi||a_t|}\{M|x_t||y_t| + |\beta_{2,0}' x_t||y_t| + \tilde{m}(x_t)(M + |\beta_{2,0}|)$, for $|\beta_1| \leq M$, $M > 0$, and for $\tilde{m}$ defined in (C5). Therefore, for each $\epsilon > 0$, it can be easily shown that there exists a finite class $\mathcal{G}_\epsilon$ containing lower and upper approximations to each $g = g_{\beta_1}(a_t, x_t, y_t) \in \mathcal{G}$, for which $g_{\epsilon,L} \leq g \leq g_{\epsilon,U}$ and $E(g_{\epsilon,U} - g_{\epsilon,L}) < \epsilon$.

Now, consider the class $\mathcal{H} = \mathcal{G}\mathcal{F}$ of functions of the form $h = gf$, where $g = \frac{1}{\phi a_t}\left\{(\beta_1 - \beta_{2,0})' x_t y_t - b(\beta_1' x_t) + b(\beta_{2,0}' x_t)\right\}$, and $f = I(r_0 < z_t \leq r)$. Note that $h = gf = (g^+ - g^-)f$, where $g^+ = \max(0, g)$ and $g^- = \max(0, -g)$. Because $\mathcal{G}$ is a class of functions that satisfies the finite bracketing conditions of Theorem 2 in Pollard [15, p. 8], it is easy to check that $\mathcal{G}^+ = \{g^+ : g^+ = \max(0, g)\}$ and $\mathcal{G}^- = \{g^- : g^- = \max(0, -g)\}$ also satisfy the finite bracketing conditions. Therefore, since $g = g^+ - g^-$, without loss of generality, we assume that $g \geq 0$. Then, for each $h \in \mathcal{H}$, we have $h_{\epsilon,L} \leq h \leq h_{\epsilon,U}$, where $h_{\epsilon,L} = g_{\epsilon,L} f_{\epsilon,L}$ and $h_{\epsilon,U} = g_{\epsilon,U} f_{\epsilon,U}$. Using the triangle inequality, it follows that for each $h \in \mathcal{H}$, we have

$$\begin{aligned} E\left(h_{\epsilon,U} - h_{\epsilon,L}\right) &= E\left(g_{\epsilon,U} f_{\epsilon,U} - g_{\epsilon,U} f_{\epsilon,L} + g_{\epsilon,U} f_{\epsilon,L} - g_{\epsilon,L} f_{\epsilon,L}\right) \\ &= E\left\{g_{\epsilon,U}(f_{\epsilon,U} - f_{\epsilon,L}) + (g_{\epsilon,U} - g_{\epsilon,L}) f_{\epsilon,L}\right\} \\ &\leq E\left\{G(f_{\epsilon,U} - f_{\epsilon,L})\right\} + E(g_{\epsilon,U} - g_{\epsilon,L}), \end{aligned}$$

where $G \geq 1$ is an integrable envelope function of g. Making use of the above findings, we conclude that, for each $\epsilon > 0$, there exists a finite class $\mathcal{H}_\epsilon$ containing lower and upper approximations to each $h \in \mathcal{H}$, for which $h_{\epsilon,L} \leq h \leq h_{\epsilon,U}$ and $E(h_{\epsilon,U} - h_{\epsilon,L}) < 2\epsilon$. Therefore, we can apply the uniform law of large numbers to $\frac{\sum_{t=1}^T h}{T}, h \in \mathcal{H}$.

□

A.2. Proof of Theorem 3.1.

PROOF. Let $l(\theta)$ be the log likelihood of $\theta = (\beta_1^{'}, \beta_2^{'}, r, d)^{'}$. The true parameter is denoted as $\theta_0 = (\beta_{1,0}^{'}, \beta_{2,0}^{'}, r_0, d_0)^{'}$. We first need to show that, as $T \to \infty$,

$$\sup_{\theta \in \Omega_1} \left| \frac{l(\theta)}{T} - E\left(\frac{l(\theta)}{T}\right) \right| \to 0, \text{ almost surely.}$$

The latter result holds if the approximating conditions of the uniform law of large numbers in Theorem 2 of Pollard [15, p. 8] are verified. Although Pollard [15] assumes that the data are independent and identically distributed, this assumption can be relaxed to assuming a stationary ergodic process; see Pollard [15, p. 9].

We have

$$\begin{aligned} \frac{l(\theta)}{T} &= \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\phi a_t} \left\{y_t \gamma_t - b(\gamma_t)\right\} + c\left(y_t, \phi a_t\right) \\ &= \frac{1}{T} \sum_{t=1}^{T} \left[\frac{1}{\phi a_t} \left\{w(\beta_1^{'} x_t) y_t - b \circ w(\beta_1^{'} x_t)\right\}\right] I(z_{t-d} \leq r) \\ &\qquad + \left[\frac{1}{\phi a_t} \left\{w(\beta_2^{'} x_t) y_t - b \circ w(\beta_2^{'} x_t)\right\}\right] I(z_{t-d} > r) + c\left(y_t, \phi a_t\right). \end{aligned}$$

Let $\mathcal{G}$ be the collection of functions of the form $g_{\beta_1}(a_t, x_t, y_t) = \frac{1}{\phi a_t} \Big\{w(\beta_1^{'} x_t)$ $y_t - b \circ w(\beta_1^{'} x_t)\Big\}$, where β_1 lies in a compact space. Because of the Lipschitz property of w and $b \circ w$ stated in (C5), it is easy to check that $\mathcal{G}$ has an integrable envelope function G given by $\frac{1}{\phi a_t}\left[\left\{|w(0)| + \tilde{w}(x_t)M\right\}|y_t| + |b\left\{w(0)\right\}| + \tilde{m}(x_t)M\right]$, for $|\beta_1| \leq M$, $M > 0$, and for $\tilde{w}$ and $\tilde{m}$ defined in (C5). Using a similar argument as in the proof of Lemma 3.1 where we check the validity of the uniform law of large numbers, we conclude that as $T \to \infty$, $\sup_{\theta \in \Omega_1}\left|\frac{l(\theta)}{T} - E\left(\frac{l(\theta)}{T}\right)\right| \to 0$ almost surely.

Using the result of Kullback-Leibler divergence stated in Lemma 5.35 of van der Vaart [24], and because $E\left(\frac{l(\theta)}{T}\right)$ is continuous for every $\theta \in \Omega_1$, a compact subset, then for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$\max_{\theta \in \Omega_1 : |\theta - \theta_0| \geq \epsilon} E\left(\frac{l(\theta)}{T}\right) + \delta < E\left(\frac{l(\theta_0)}{T}\right) - \delta. \tag{A.12}$$

Applying the uniform law of large numbers and by making use of (A.12), we conclude that, for all $\epsilon > 0$, there exists $\delta > 0$ such that, for T sufficiently large, and uniformly for $|\theta - \theta_0| \geq \epsilon$, $\frac{l(\theta)}{T} \leq E\left(\frac{l(\theta)}{T}\right) + \delta \leq \max_{\theta \in \Omega_1 : |\theta - \theta_0| \geq \epsilon} E\left(\frac{l(\theta)}{T}\right) + \delta < E\left(\frac{l(\theta_0)}{T}\right) - \delta < \frac{l(\theta_0)}{T}$ almost surely. Hence, for T sufficiently large, $\left|\hat{\theta}_T - \theta_0\right| \leq \epsilon$ almost surely. As $\epsilon > 0$ is arbitrary, $\hat{\theta}_T \to \theta_0$ almost surely. This completes the proof. ☐

A.3. Proof of Theorem 3.2.

PROOF. Without loss of generality, the delay parameter d is assumed to be known, and $d = 0$. Therefore, the parameter vector becomes $\theta = (\beta_1^{'}, \beta_2^{'}, r)^{'}$ and the parameter space Ω is modified accordingly. Since the maximum likelihood estimator $\hat{\theta}_T$ is strongly consistent, without loss of generality, the parameter space can be restricted to a neighborhood of θ_0, namely, $\Omega_1 = \left\{\theta \in \Omega : |\beta_i - \beta_{i,0}| < \Delta, i = 1, 2; |r - r_0| < \Delta\right\}$, for some $0 < \Delta < 1$ to be determined later. To simplify the notation, we assume that $r_0 = 0$. Then, it suffices to show that for all $\epsilon > 0$, there exists $K > 0$ such that, with probability greater than $1 - \epsilon$, $\theta \in \Omega_1$ and $|r| > \frac{K}{T}$ implies that $l(\beta_1, \beta_2, r) - l(\beta_1, \beta_2, 0) < 0$.

We first consider the case that $r > 0$. Then, we have

$$\begin{aligned}
&\frac{l(\beta_1,\beta_2,r)-l(\beta_1,\beta_2,0)}{T}\\
&=\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{\phi a_t}\left\{w(\beta_1^{'}x_t)y_t-b\circ w(\beta_1^{'}x_t)\right\}+c\left(y_t,\phi a_t\right)\right]I(z_t\le r)\\
&\qquad+\left[\frac{1}{\phi a_t}\left\{w(\beta_2^{'}x_t)y_t-b\circ w(\beta_2^{'}x_t)\right\}+c\left(y_t,\phi a_t\right)\right]I(z_t> r)\\
&\qquad-\left[\frac{1}{\phi a_t}\left\{w(\beta_1^{'}x_t)y_t-b\circ w(\beta_1^{'}x_t)\right\}+c\left(y_t,\phi a_t\right)\right]I(z_t\le 0)\\
&\qquad-\left[\frac{1}{\phi a_t}\left\{w(\beta_2^{'}x_t)y_t-b\circ w(\beta_2^{'}x_t)\right\}+c\left(y_t,\phi a_t\right)\right]I(z_t> 0)\\
&=\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}\left[\left\{w(\beta_1^{'}x_t)-w(\beta_2^{'}x_t)\right\}y_t-b\circ w(\beta_1^{'}x_t)+b\circ w(\beta_2^{'}x_t)\right]\\
&\qquad\qquad\times I(0<z_t\le r).
\end{aligned}$$

And hence,

$$\begin{aligned}
&\frac{l(\beta_1,\beta_2,r)-l(\beta_1,\beta_2,0)}{T}\\
&=\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}\left[\left\{w(\beta_1^{'}x_t)-w(\beta_{1,0}^{'}x_t)\right\}y_t-b\circ w(\beta_1^{'}x_t)+b\circ w(\beta_{1,0}^{'}x_t)\right]\\
&\qquad\qquad\times I(0<z_t\le r)\\
&+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}\left[\left\{w(\beta_{2,0}^{'}x_t)-w(\beta_2^{'}x_t)\right\}y_t-b\circ w(\beta_{2,0}^{'}x_t)+b\circ w(\beta_2^{'}x_t)\right]\\
&\qquad\qquad\times I(0<z_t\le r)\\
&+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}\left[\left\{w(\beta_{1,0}^{'}x_t)-w(\beta_{2,0}^{'}x_t)\right\}y_t-b\circ w(\beta_{1,0}^{'}x_t)+b\circ w(\beta_{2,0}^{'}x_t)\right]\\
&\qquad\qquad\times I(0<z_t\le r).
\end{aligned}$$

Define $Q(r) = E\left\{I(0<z_t\le r)\right\}$, for $0 < r \le \Delta$. Let $M_\beta(a_t,x_t,y_t) = \frac{1}{\phi a_t}\left\{\gamma_t y_t \; -b(\gamma_t)\right\}$, where $\gamma_t = w(\beta^{'}x_t)$. Recall that, by Assumption (C5), there exists an integrable function $\Lambda(a_t,x_t,y_t)$ such that $|M_\beta(a_t,x_t,y_t) - M_{\beta^*}(a_t,x_t,y_t)| \le \Lambda(a_t,x_t,y_t)\left|\beta \; -\beta^*\right|$, for every β, β^*, a_t, x_t, and y_t.

Thus, for $\Delta > 0$, we have

$$\begin{aligned}
&\frac{l(\beta_1,\beta_2,r)-l(\beta_1,\beta_2,0)}{TQ(r)} \\
&= \frac{1}{TQ(r)}\sum_{t=1}^{T}\left\{M_{\beta_1}(a_t,x_t,y_t)-M_{\beta_{1,0}}(a_t,x_t,y_t)\right\}I(0<z_t\le r) \\
&\qquad + \left\{M_{\beta_{2,0}}(a_t,x_t,y_t)-M_{\beta_2}(a_t,x_t,y_t)\right\}I(0<z_t\le r) \\
&\qquad + \left\{M_{\beta_{1,0}}(a_t,x_t,y_t)-M_{\beta_{2,0}}(a_t,x_t,y_t)\right\}I(0<z_t\le r) \\
&\le \left(|\beta_1-\beta_{1,0}|+|\beta_2-\beta_{2,0}|\right)\frac{1}{TQ(r)}\sum_{t=1}^{T}\Lambda(a_t,x_t,y_t)I(0<z_t\le r) \\
&+ \frac{1}{TQ(r)}\sum_{t=1}^{T}\left\{M_{\beta_{1,0}}(a_t,x_t,y_t)-M_{\beta_{2,0}}(a_t,x_t,y_t)\right\}I(0<z_t\le r).
\end{aligned}$$

Suppose that the following claim is valid; the verification of which is deferred to the end of this proof.

Claim I. Let M_t be a measurable function of $(a_t, x_t', y_t)'$. Assume that there exist $M > 0$ and $\Delta > 0$, such that $E\left(M_t^2|z_t = z\right) \le M$, for all $z \in [-\Delta, \Delta]$. Assume that the process $W = \left[\left\{M_t, z_t I\left(-\Delta \le z_t \le \Delta\right)\right\}'\right]$ is φ-mixing with exponentially decaying mixing coefficients. Then, for all $\epsilon > 0$, for all $\zeta > 0$, there exists $K > 0$ such that, for all T,

$$P\left(\sup_{\frac{K}{T}<r\le\Delta}\left|\sum\frac{I(0<z_t\le r)}{TQ(r)}-1\right|<\zeta\right)>1-\epsilon, \tag{A.13}$$

and

$$P\left(\sup_{\frac{K}{T}<r\le\Delta}\left|\sum\frac{M_tI(0<z_t\le r)-E\left\{M_tI(0<z_t\le r)\right\}}{TQ(r)}\right|<\zeta\right)>1-\epsilon. \tag{A.14}$$

It follows from Claim I that for all $\epsilon > 0, \zeta > 0$, there exist $K(\epsilon, \zeta) > 0$, such that with probability greater than $1-\epsilon$, $\frac{K}{T} < r \le \Delta$ implies that

$$\frac{l(\beta_1,\beta_2,r)-l(\beta_1,\beta_2,0)}{TQ(r)}<\left(|\beta_1-\beta_{1,0}|+|\beta_2-\beta_{2,0}|\right)(\zeta+M)+\zeta+\kappa, \tag{A.15}$$

where

$$\kappa=\frac{1}{TQ(r)}\sum_{t=1}^{T}E\left\{M_{\beta_{1,0}}(a_t,x_t,y_t)-M_{\beta_{2,0}}(a_t,x_t,y_t)|z_t\right\}I(0<z_t\le r). \tag{A.16}$$

Because of Assumptions (C1) and (C8), and using the result of Kullback-Leibler divergence stated in Lemma 5.35 of van der Vaart [24], we have that for each $z_t \in (0, \Delta]$, $E\left\{M_{\beta_{1,0}}(a_t, x_t, y_t) - M_{\beta_{2,0}}(a_t, x_t, y_t)|z_t\right\}$ is a continuous function and is negative; and hence, its maximum is $\leq -\chi$, for some $\chi > 0$. Hence, $\kappa \leq \frac{-\chi}{TQ(r)} \sum_{t=1}^{T} I(0 < z_t \leq r) \leq -\chi(1-\zeta)$, for some $\chi > 0$. Consequently, for all $\epsilon > 0, \zeta > 0$, there exist $K(\epsilon, \zeta) > 0, \chi > 0$, such that with probability greater than $1 - \epsilon$, $\frac{K}{T} < r \leq \Delta$ implies that $\frac{l(\beta_1,\beta_2,r)-l(\beta_1,\beta_2,0)}{TQ(r)} < 2\Delta(\zeta + M) + \zeta - \chi(1-\zeta)$. Now, choose $\Delta > 0$ and $\zeta > 0$ such that $2\Delta(\zeta + M) + \zeta - \chi(1-\zeta) < 0$; and hence the validity of Theorem 3.2 under the further condition that $r > 0$. Similar argument can be used to prove Theorem 3.2 for the case of $r < 0$.

We now verify Claim I. Define

$$Q_T(r) = \sum \frac{I(0 < z_t \leq r)}{T}, \tag{A.17}$$

$$R_T(r) = \sum \frac{M_t I(0 < z_t \leq r)}{T}, \tag{A.18}$$

$$\tilde{R}_T(r_1, r_2) = \sum \frac{M_t I(r_1 < z_t \leq r_2)}{T}. \tag{A.19}$$

By choosing Δ sufficiently small, it follows from Assumption (C3) that there exist $0 < m < M < \infty$, independent of T, such that for all r in $(0, \Delta)$,

$$mr \leq Q(r) \leq Mr. \tag{A.20}$$

Since $E\{I(0 < z_t \leq r)\} = E\{I(0 < z_t \leq r)^2\} = Q(r)$, then we have, for all r in $(0, \Delta)$, $\text{var}\{I(0 < z_t \leq r)\} = Q(r) - Q(r)^2 = Q(r)\{1 - Q(r)\} \leq Q(r)(1 - mr)$. And hence, for sufficiently small $\Delta > 0$, there exists $H > 0$, independent of T, such that for all r in $(0, \Delta)$,

$$\text{var}\{I(0 < z_t \leq r)\} \leq HQ(r). \tag{A.21}$$

Because $E\left(M_t^2|z_t\right)$ is assumed to be bounded above for all $z_t \in [-\Delta, \Delta]$, it is readily checked that there exists $H > 0$, independent of T, such that for all r_1, r_2 in $(0, \Delta)$,

$$E\{M_t I(r_1 < z_t \leq r_2)\} \leq H\{Q(r_2) - Q(r_1)\}. \tag{A.22}$$

Similarly, $\text{var}\{M_t I(r_1 < z_t \leq r_2)\} \leq E\{M_t^2 I(r_1 < z_t \leq r_2)\} \leq E\{E(M_t^2| z_t) I(r_1 < z_t \leq r_2)\}$; and hence,

$$\text{var}\{M_t I(r_1 < z_t \leq r_2)\} \leq H\{Q(r_2) - Q(r_1)\}. \tag{A.23}$$

Let $R_t = M_t I(r_1 < z_t \le r_2)$. Because the process $W = \left[\{M_t, z_t I(-\Delta \le z_t \le \Delta)\}'\right]$ is φ-mixing, then $|\mathrm{Cov}(R_t, R_s)| \le 2\varphi_{|t-s|}^{\frac{1}{p}} \{E(R_t^p)\}^{\frac{1}{p}} \{E(R_s^q)\}^{\frac{1}{q}}$, for any $p, q \ge 1$ and $\frac{1}{p} + \frac{1}{q} = 1$; see Doukhan [5, p. 9]. Choose $p = q = 2$. Because the φ-mixing coefficient is assumed to be exponentially decaying, and by making use of the stationarity assumption, we have $|\mathrm{Cov}(R_t, R_s)| \le 2c\rho^{\frac{|t-s|}{2}} E(R_t^2) \le 2c\rho^{\frac{|t-s|}{2}} H\{Q(r_2) - Q(r_1)\}$, for some $c > 0, 0 \le \rho < 1$, and $H > 0$. Hence, we make use of the latter inequality for the covariance of φ-mixing random variables to verify that for all $b > 0$, there exists $H > 0$ such that for all $r, r_1, r_2 \in [-b, b]$, for all T, we have

$$\mathrm{var}\{TQ_T(r)\} \le THQ(r), \tag{A.24}$$

$$\mathrm{var}\left\{T\tilde{R}_T(r_1, r_2)\right\} \le TH\{Q(r_2) - Q(r_1)\}, \tag{A.25}$$

$$\mathrm{var}\{TR_T(r)\} \le THQ(r). \tag{A.26}$$

Therefore, Claim I can be verified by making use of the inequalities (A.20)–(A.26), and by employing arguments as in Chan [3, p. 529].

□

A.4. Proof of Lemma 3.2.

PROOF. Let $l(\theta)$ be the log likelihood of $\theta = (\delta', r)'$, where $\delta = (\beta_1', \beta_2')'$. Let $l(., r)$ be globally maximized at $\hat{\delta}_r = (\hat{\beta}_{1,r}', \hat{\beta}_{2,r}')'$. Since the maximum likelihood estimator $\hat{\theta}_T$ is strongly consistent, without loss of generality, the parameter space can be restricted to some neighborhood of θ_0, say, $\Omega_1 = \{\theta \in \Omega : |\beta_i - \beta_{i,0}| < 1, i = 1, 2; |r - r_0| < 1\}$.

Let $\dot{l}(\hat{\delta}_{r_0}, r) = \frac{\partial}{\partial \delta} l(\delta, r)|_{\delta = \hat{\delta}_{r_0}}$ and $\ddot{l}(\hat{\delta}_{r_0}, r) = \frac{\partial^2}{\partial \delta^2} l(\delta, r)|_{\delta = \hat{\delta}_{r_0}}$. Using a Taylor's expansion about $\hat{\delta}_{r_0}$ carried out to the third-order terms, there exists $\tilde{\delta}$ between δ and $\hat{\delta}_{r_0}$ such that

$$\begin{aligned} & l(\delta, r) - l(\hat{\delta}_{r_0}, r) \\ & = (\delta - \hat{\delta}_{r_0})' \dot{l}(\hat{\delta}_{r_0}, r) + \frac{1}{2}(\delta - \hat{\delta}_{r_0})' \ddot{l}(\hat{\delta}_{r_0}, r)(\delta - \hat{\delta}_{r_0}) + R_T(\tilde{\delta}, \delta, \hat{\delta}_{r_0}), \end{aligned} \tag{A.27}$$

where the remainder term $R_T = R_T(\tilde{\delta}, \delta, \hat{\delta}_{r_0})$ satisfies

$$\lim_{T \to \infty} \sup_{|\delta - \hat{\delta}_{r_0}| \to 0} \frac{|R_T|}{T|\delta - \hat{\delta}_{r_0}|^2} = 0. \tag{A.28}$$

For simplicity, we shall prove this lemma for the case that $r \ge r_0$ and omit the case that $r < r_0$ as the proof is similar.

Since the score $\dot{l}(\delta, r_0)$ equals zero at $\delta = \hat{\delta}_{r_0}$, we have

$$\begin{aligned}\dot{l}(\hat{\delta}_{r_0}, r) &= \dot{l}(\hat{\delta}_{r_0}, r) - \dot{l}(\hat{\delta}_{r_0}, r_0)\\ &= \begin{bmatrix} \sum_{t=1}^{T} \dot{M}_{\hat{\beta}_{1,r_0}} I(r_0 < z_{t-d} \le r) \\ \sum_{t=1}^{T} -\dot{M}_{\hat{\beta}_{2,r_0}} I(r_0 < z_{t-d} \le r) \end{bmatrix}\end{aligned}$$

where $\dot{M}_{\hat{\beta}_{i,r_0}} = \frac{\partial}{\partial \beta_i} M_{\beta_i}|_{\beta_i = \hat{\beta}_{i,r_0}}$ and $M_{\beta_i} = M_{\beta_i}(y_t; a_t, x_t) = \frac{1}{\phi a_t}\{w(\beta_i^{'} x_t) y_t - b \circ w(\beta_i^{'} x_t)\}$, $i = 1, 2$. Let $\dot{l}^j(\delta, r)$ denote the j^{th} component of $\dot{l}(\delta, r)$. Let k be the dimension of δ. Denote the absolute norm of $\dot{l}(\hat{\delta}_{r_0}, r)$ by $|\dot{l}(\hat{\delta}_{r_0}, r)| = \sum_{j=1}^{k} |\dot{l}^j(\hat{\delta}_{r_0}, r)|$. Using (C3) and (D2), there exists a scalar $M_1 > 0$ such that for T sufficiently large, for all $K > 0$ and $|r - r_0| \le \frac{K}{T}$, we have $E(|\dot{l}(\hat{\delta}_{r_0}, r)|) = \sum_{i=1}^{2} \sum_{t=1}^{T} E\{|\dot{M}_{\hat{\beta}_{i,r_0}}| I(r_0 < z_{t-d} \le r)\} \le 2TM_1 P(r_0 < z_{t-d} \le r) = O(1)$. It follows readily from Markov's inequality that for T sufficiently large, for all $K > 0$ and $|r - r_0| \le \frac{K}{T}$, we have

$$|\dot{l}(\hat{\delta}_{r_0}, r)| = O_p(1). \tag{A.29}$$

On the other hand, the Hessian matrix $\ddot{l}(\hat{\delta}_{r_0}, r)$ can be written as

$$\begin{aligned}\ddot{l}(\hat{\delta}_{r_0}, r) &= \left\{\ddot{l}(\hat{\delta}_{r_0}, r) - \ddot{l}(\hat{\delta}_{r_0}, r_0)\right\} + \ddot{l}(\hat{\delta}_{r_0}, r_0)\\ &= \begin{bmatrix} \eta_1 + \xi_1 & 0 \\ 0 & -\eta_2 + \xi_2 \end{bmatrix},\end{aligned} \tag{A.30}$$

where $\eta_i = \sum_{t=1}^{T} \ddot{M}_{\hat{\beta}_{i,r_0}} I(r_0 < z_{t-d} \le r), i = 1, 2$, $\xi_1 = \sum_{t=1}^{T} \ddot{M}_{\hat{\beta}_{1,r_0}} I(z_{t-d} \le r_0)$, $\xi_2 = \sum_{t=1}^{T} \ddot{M}_{\hat{\beta}_{2,r_0}} I(z_{t-d} > r_0)$, and $\ddot{M}_{\hat{\beta}_{i,r_0}} = \frac{\partial^2}{\partial \beta_i^2} M_{\beta_i}\Big|_{\beta_i = \hat{\beta}_{i,r_0}}, i = 1, 2$.

By employing a similar argument as above, it can be shown that for T sufficiently large, for all $K > 0$ and $|r - r_0| \le \frac{K}{T}$, we have

$$|\eta_i| = O_p(1), i = 1, 2. \tag{A.31}$$

In reference to Example 19.8 of van der Vaart [24], it can be easily shown that the collection of functions $\{\ddot{M}_\beta, \beta \text{ in a fixed compact set}\}$ is Glivenko-Cantelli. Hence, using the argument of van der Vaart [24, p. 279], we have for T sufficiently large,

$$\frac{1}{T}\xi_1 = E\{\ddot{M}_{\beta_{1,0}} I(z_{t-d} \le r_0)\} + o_p(1), \tag{A.32}$$

$$\frac{1}{T}\xi_2 = E\{\ddot{M}_{\beta_{2,0}} I(z_{t-d} > r_0)\} + o_p(1); \tag{A.33}$$

where $E\{\ddot{M}_{\beta_{1,0}} I(z_{t-d} \le r_0)\}$ and $E\{\ddot{M}_{\beta_{2,0}} I(z_{t-d} > r_0)\}$ are negative-definite by (D4), and they essentially determine the curvature of the log likelihood.

Combining the results in (A.31)–(A.33) with the result in (A.30), and making use of the property that a negative-definite matrix has a maximum eigenvalue that is less than $-\lambda$, for some $\lambda > 0$, it follows that for T sufficiently large, for all $K > 0$ and $|r - r_0| \le \frac{K}{T}$, we have

$$\begin{aligned}
&\frac{1}{2}(\delta - \hat{\delta}_{r_0})' \ddot{l}(\hat{\delta}_{r_0}, r)(\delta - \hat{\delta}_{r_0}) \\
&\quad \le \frac{1}{2}\sum_{i=1}^{2}\{|\beta_i - \hat{\beta}_{i,r_0}|^2 |\eta_i| + (\beta_i - \hat{\beta}_{i,r_0})' \xi_i (\beta_i - \hat{\beta}_{i,r_0})\} \\
&\quad \le \frac{1}{2}\sum_{i=1}^{2} |\beta_i - \hat{\beta}_{i,r_0}|^2 [O_p(1) - T\{2\lambda - o_p(1)\}], \qquad \text{(A.34)}
\end{aligned}$$

for some scalar $\lambda > 0$.

Finally, we combine the results in (A.28), (A.29), and (A.34) with the result in (A.27). Then, for all $\epsilon > 0, a_T = o_p(T^\gamma) > 0$, where $-1 < \gamma < -\frac{1}{2}$, $|\delta - \hat{\delta}_{r_0}| < a_T$, $\forall K > 0$, and uniformly for $|r - r_0| \le \frac{K}{T}$, there exists T_0 such that with probability greater than $1 - \epsilon$, for any $T > T_0$, and for δ on the boundary of the open sphere N_{a_T} of radius a_T centered at $\hat{\delta}_{r_0}$, we have

$$\begin{aligned}
l(\delta, r) - l(\hat{\delta}_{r_0}, r) &\le a_T O_p(1) + \frac{1}{2} a_T^2 [O_p(1) - T\{2\lambda - o_p(1)\}] + T a_T^2 o_p(1) \\
&\le T a_T^2 \{-2\lambda + o_p(1)\}, \qquad \text{(A.35)}
\end{aligned}$$

where $-2\lambda + o_p(1) < 0$. Thus, $l(\delta, r)$ must attain a maximum at some point belonging to N_{a_T}. Because $l(\delta, r)$ is continuous for every $\theta \in \Omega_1$, a compact subset, then there exists a global maximum $\hat{\delta}_r = (\hat{\beta}'_{1,r}, \hat{\beta}'_{2,r})'$ such that for all $K > 0$,

$$\sup_{|r - r_0| \le \frac{K}{T}} |\hat{\beta}_{i,r} - \hat{\beta}_{i,r_0}| = o_p(1/\sqrt{T}),\ i = 1, 2.$$

This completes the proof. □

A.5. Proof of Lemma 3.3.

Proof. We use the same notations as in the proof of Lemma 3.2. For simplicity, we shall prove this lemma for the case that $\kappa \ge 0$ and omit the case that $\kappa < 0$ as the proof is similar. We have

$$\tilde{l}(\kappa) = \{l(\hat{\delta}_{r_0+\frac{\kappa}{T}}, r_0+\kappa/T) - l(\hat{\delta}_{r_0+\frac{\kappa}{T}}, r_0)\} + \{l(\hat{\delta}_{r_0+\frac{\kappa}{T}}, r_0) - l(\hat{\delta}_{r_0}, r_0)\}$$

$$= \sum_{t=1}^{T}(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}})I(r_0 < z_{t-d} \le r_0+\kappa/T) \tag{A.36}$$

$$+ (M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{1,r_0}})I(z_{t-d} \le r_0) \tag{A.37}$$

$$+ (M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{2,r_0}})I(z_{t-d} > r_0). \tag{A.38}$$

We first consider equation (A.36); we have

$$\sum_{t=1}^{T}(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}})I(r_0 < z_{t-d} \le r_0+\kappa/T)$$
$$= \sum_{t=1}^{T}(M_{\beta_{1,0}} - M_{\beta_{2,0}})I(r_0 < z_{t-d} \le r_0+\kappa/T)$$
$$+ \{(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\beta_{1,0}}) + (M_{\beta_{2,0}} - M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}})\}I(r_0 < z_{t-d} \le r_0+\kappa/T),$$

where $\sum_{t=1}^{T}(M_{\beta_{1,0}} - M_{\beta_{2,0}})I(r_0 < z_{t-d} \le r_0+\kappa/T) = l(\delta_0, r_0+\kappa/T) - l(\delta_0, r_0)$. Hence,

$$\left|\sum_{t=1}^{T}(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}})I(r_0 < z_{t-d} \le r_0+\kappa/T) - \{l(\delta_0, r_0+\kappa/T) - l(\delta_0, r_0)\}\right|$$
$$\le (|\hat{\beta}_{1,r_0+\frac{\kappa}{T}} - \beta_{1,0}| + |\hat{\beta}_{2,r_0+\frac{\kappa}{T}} - \beta_{2,0}|) \times \sum_{t=1}^{T}\Lambda(a_t, x_t, y_t)I(r_0 < z_{t-d} \le r_0+\kappa/T); \tag{A.39}$$

the latter inequality holds because of (C5). Because $E\{\Lambda(a_t,x_t,y_t)I(r_0 < z_{t-d} \le r_0+\kappa/T)\} = O(1/T)$, it follows that for T sufficiently large, $\sum_{t=1}^{T}\Lambda(a_t,x_t,y_t)I(r_0 < z_{t-d} \le r_0+\kappa/T) = O_p(1)$. On the other hand, for T sufficiently large, for all $K>0$ and uniformly for all $|\kappa| \le K$, it holds that $|\hat{\beta}_{i,r_0+\frac{\kappa}{T}} - \beta_{i,0}|, i=1,2$, is less than or equal to $|\hat{\beta}_{i,r_0+\frac{\kappa}{T}} - \hat{\beta}_{i,r_0}| + |\hat{\beta}_{i,r_0} - \beta_{i,0}| = o_p(1/\sqrt{T}) + O_p(1/\sqrt{T})$, using Lemma 3.2 and the property of the maximum likelihood estimator of the GTM with known true delay and threshold. Thus, for T sufficiently large, for all $K>0$ and uniformly for all $|\kappa| \le K$, the inequality in (A.39) entails that

$$\left|\sum_{t=1}^{T}(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}})I(r_0 < z_{t-d} \leq r_0 + \kappa/T) - \{l(\delta_0, r_0+\kappa/T) - l(\delta_0, r_0)\}\right| = o_p(1). \tag{A.40}$$

Next, we consider equation (A.37). Expand $M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}}$ and $M_{\hat{\beta}_{1,r_0}}$ in a Taylor series around $\beta_{1,0}$. We have

$$\begin{aligned}\sum_{t=1}^{T}&(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{1,r_0}})I(z_{t-d} \leq r_0)\\ &\leq (\hat{\beta}_{1,r_0+\frac{\kappa}{T}} - \hat{\beta}_{1,r_0})' \sum_{t=1}^{T} \dot{M}_{\beta_{1,0}} I(z_{t-d} \leq r_0)\\ &\quad + \frac{1}{2}(\hat{\beta}_{1,r_0+\frac{\kappa}{T}} - \hat{\beta}_{1,r_0})' \sum_{t=1}^{T} \ddot{M}_{\beta_{1,0}} I(z_{t-d} \leq r_0)\\ &\quad \times (\hat{\beta}_{1,r_0+\frac{\kappa}{T}} + \hat{\beta}_{1,r_0} - 2\beta_{1,0}) + r_T,\end{aligned} \tag{A.41}$$

where the remainder term r_T is such that for T sufficiently large, $r_T = o_p(1)$. The central limit theorem is applied to the martingale $\sum_{t=1}^{T}\sum_{j=1}^{p} c_j \dot{M}_{\beta_{1,0}}^{(j)} \times I(z_{t-d} \leq r_0)$, for all nonzero vectors of constants $c = (c_1, \cdots, c_p)$. Using Cramer-Wold device, it follows that for all T sufficiently large, $|\sum_{t=1}^{T} \dot{M}_{\beta_{1,0}} \times I(z_{t-d} \leq r_0)| = O_p(\sqrt{T})$. The latter indeed holds because $\{\dot{M}_{\beta_{1,0}}(y_t; a_t, x_t)\}$ is a martingale-difference sequence with respect to the σ-algebra $\mathcal{F}_t = \sigma(a_t, x_t, y_{t-k}, a_{t-k}, x_{t-k}, k \geq 1)$ and because $E\{\dot{M}_{\beta_{1,0}}\dot{M}_{\beta_{1,0}}' I(z_{t-d} \leq r_0)\} = -E\{\ddot{M}_{\beta_{1,0}} I(z_{t-d} \leq r_0)\}$ is finite; see Billingsley [1]. On the other hand, by the law of large numbers, for all T sufficiently large, $\frac{1}{T}\sum_{t=1}^{T} \ddot{M}_{\beta_{1,0}} I(z_{t-d} \leq r_0)$ converges to $E\{\ddot{M}_{\beta_{1,0}} I(z_{t-d} \leq r_0)\}$ in probability. Thus, for all T sufficiently large, for all $K > 0$ and uniformly for all $|\kappa| \leq K$, the inequality in (A.41) yields

$$\begin{aligned}&\left|\sum_{t=1}^{T}(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{1,r_0}})I(z_{t-d} \leq r_0)\right|\\ &\leq o_p(1/\sqrt{T})O_p(\sqrt{T}) + o_p(1/\sqrt{T})O_p(T)O_p(1/\sqrt{T}) + o_p(1) = o_p(1),\end{aligned} \tag{A.42}$$

using Lemma 3.2. Similarly, it can be shown that for all T sufficiently large, for all $K > 0$ and uniformly for all $|\kappa| \leq K$, we have

$$\left|\sum_{t=1}^{T}(M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{2,r_0}})I(z_{t-d} > r_0)\right| = o_p(1). \tag{A.43}$$

Combine the results in (A.40), (A.42), and (A.43) with the results in (A.36)–(A.38) to complete the proof.

□

A.6. Proof of Theorem 3.3.

Proof. Owing to Lemma 3.3, we shall proceed as if $\tilde{l}(\kappa) = l(\delta_0, r_0 + \kappa/T) - l(\delta_0, r_0)$. Without loss of generality, assume that $d = 0$ and $\kappa > 0$. Then,

$$\tilde{l}(\kappa) = \sum_{t=1}^{T} \{M_{\beta_{1,0}}(y_t; a_t, x_t) - M_{\beta_{2,0}}(y_t; a_t, x_t)\} I(r_0 < z_t \leq r_0 + \kappa/T).$$

Let A_i be the event that the sample path of $\tilde{l}(\kappa)$ possesses at least i discontinuities on the interval $(u, u+h], u \geq 0, h \geq 0, 0 \leq i \leq T$. Hence, by making use of (C3), it is easy to check that there exists $M > 0$ such that $P(A_2) \leq \sum_{t_1=1}^{T} \sum_{t_2=1,\ t_2 \neq t_1}^{T} P(r_0 + \frac{u}{T} < z_{t_1} \leq r_0 + \frac{u+h}{T}, r_0 + \frac{u}{T} < z_{t_2} \leq r_0 + \frac{u+h}{T}) \leq Mh^2$. Employing a similar argument as in the proof of Lemma 3.2 in Ibragimov and Has'minskii [8, p. 261], it can be readily checked that $(\{\tilde{l}(-\kappa), \kappa \geq 0\}, \{\tilde{l}(+\kappa), \kappa \geq 0\})$ is tight.

Let $\epsilon = \frac{1}{T} > 0$ and $\xi_t = (a_t, x_t^{'}, y_t)^{'}$. Define a piecewise-constant interpolation process, $x^\epsilon(.)$, indexed by ϵ with paths in $D[0,1]$, as follows

$$\begin{aligned} x^\epsilon(v) &= X^\epsilon_{[Tv]},\ 0 \leq v \leq 1, \\ X^\epsilon_0 &= 0,\ X^\epsilon_{t+1} = X^\epsilon_t + J^\epsilon_{t+1},\ t = 0, 1, 2, \cdots \\ J^\epsilon_t &= \{M_{\beta_{1,0}}(y_t; a_t, x_t) - M_{\beta_{2,0}}(y_t; a_t, x_t)\} I(r_0 < z_t \leq r_0 + \kappa\epsilon). \end{aligned} \tag{A.44}$$

Here, we denote by $[\cdot]$ the integer part of the expression inside the square bracket. Note that $x^\epsilon(1) = \tilde{l}(\kappa)$ and $\{x^\epsilon(v), 0 \leq v \leq 1\}$ is tight in $D[0,1]$. Furthermore, $x^\epsilon(v) = X^\epsilon_t$, for $v \in [t\epsilon, t\epsilon + \epsilon), t = 0, 1, \cdots, T$.

We now show that $\{x^\epsilon(v), 0 \leq v \leq 1\}$ converges weakly in $D[0,1]$ to $\{\rho(v), 0 \leq v \leq 1\}$, a compound Poisson process with rate $\pi(r_0)\kappa$ and the distribution of jump same as the conditional distribution of $M_{\beta_{1,0}}(y_t; a_t, x_t) - M_{\beta_{2,0}}(y_t; a_t, x_t)$ given $z_t = r_0^+$. We do this by making use of Theorem 1 of Kushner [12] via operator convergence. By employing truncation arguments as in Kushner [12], we can and will assume that x^ϵ are uniformly bounded. First, we define some notations and the operators. Let $\mathcal{F}_v$ denote an increasing sequence of σ-algebras to which $\{x^\epsilon(u), u \leq v\}$ are adapted, for all $\epsilon > 0$. Let $\mathscr{L}$ denote the progressively measurable functions with respect to $\mathcal{F}_v$. Define $\overline{\mathscr{L}}$ to be the subset of $\mathscr{L}$ for which $\sup_v E|f(v)| <$

∞. Let E_v^ϵ denote the conditional expectation given $\mathcal{F}_v^\epsilon$, which is the σ-algebra generated by $\{x^\epsilon(u), u \le v\}$. Note $\mathcal{F}_v^\epsilon$ is a subset of $\mathcal{F}_v$. For f and $f^\delta \in \overline{\mathscr{L}}$, define $p\text{-}\lim_{\delta\to 0} f^\delta = f$ if and only if $\sup_{v,\delta} E|f^\delta(v)| < \infty$ and $\lim_{\delta\to 0} E|f^\delta(v) - f(v)| = 0$ for every v. Define the p-infinitesimal operator $\hat{A}^\epsilon$ by $\hat{A}^\epsilon : \mathscr{D}(\hat{A}^\epsilon) \subseteq \overline{\mathscr{L}} \to \overline{\mathscr{L}}$ such that $f \in \mathscr{D}(\hat{A}^\epsilon)$ and $\hat{A}^\epsilon f = g$ if and only if for $f, g \in \overline{\mathscr{L}}$ and adapted to $\{\mathcal{F}_v^\epsilon\}$ and g being p-right continuous, we have $p\text{-}\lim_{\delta\to 0}[\frac{1}{\delta}\{E_v^\epsilon f(v+\delta) - f(v)\} - g(v)] = 0$. Let $\hat{\mathscr{C}}$ denote the space of continuous bounded real-valued functions which are zero at infinity and $\hat{\mathscr{C}}_0^2$ be the subset of $\hat{\mathscr{C}}$ with compact support and continuous second derivative. Define the operator A on $\hat{\mathscr{C}}_0^2$ by $Af(w) = \pi(r_0)\kappa \int \{f(w+y) - f(w)\} q(dy)$, where $q(dy)$ is the probability measure induced by the conditional distribution of $M_{\beta_{1,0}}(y_t; a_t, x_t) - M_{\beta_{2,0}}(y_t; a_t, x_t)$ given $z_t = r_0^+$.

Let $f(.) \in \hat{\mathscr{C}}_0^2$. For every $\tau_\epsilon > 0$, define $f^\epsilon(v) = \frac{1}{\tau_\epsilon}\int_0^{\tau_\epsilon} E_v^\epsilon\{f(x^\epsilon(v+s))\} ds$. Then, f^ϵ is in $\mathscr{D}(\hat{A}^\epsilon)$ with $\hat{A}^\epsilon f^\epsilon(v) = \frac{1}{\tau_\epsilon}[E_v^\epsilon\{f(x^\epsilon(v+\tau_\epsilon))\} - f(x^\epsilon(v))]$; see Kurtz [11, p. 625]. We next study the limiting behavior of $\hat{A}^\epsilon f^\epsilon$. We have

$$\begin{aligned}
\hat{A}^\epsilon f^\epsilon(v) &= \frac{1}{\tau_\epsilon}[E_v^\epsilon\{f(x^\epsilon(v+\tau_\epsilon))\} - f(x^\epsilon(v))] \\
&= \frac{1}{\tau_\epsilon} \sum_{k=[Tv]}^{[T(v+\tau_\epsilon)]-1} E_v^\epsilon\{f(X_{k+1}^\epsilon) - f(X_k^\epsilon)\} \\
&= \frac{1}{\tau_\epsilon} \sum_{k=0}^{[T(v+\tau_\epsilon)]-[Tv]-1} E_v^\epsilon\{f(X_{k+[Tv]}^\epsilon + J_{k+[Tv]+1}^\epsilon) - f(X_{k+[Tv]}^\epsilon)\}. \quad \text{(A.45)}
\end{aligned}$$

Because $\{x^\epsilon(v), 0 \le v \le 1\}$ is tight, any of its subsequence has a convergent subsequence. With no loss of generality, assume that $\{x^\epsilon(v), 0 \le v \le 1\}$ converges weakly to $\{x(v), 0 \le v \le 1\}$ and, indeed, by enlarging the probability space, the convergence may and will be assumed to be almost sure convergence. By making use of Theorem 15.3 in Billingsley [1, Equation (15.8)], we claim that

$$\begin{aligned}
\hat{A}^\epsilon f^\epsilon(v) = \frac{1}{\tau_\epsilon} &\sum_{k=0}^{[T(v+\tau_\epsilon)]-[Tv]-1} E_v^\epsilon\{f(x^\epsilon(v) + J_{k+[Tv]+1}^\epsilon) - f(x^\epsilon(v))\} \\
&+ o_p(1), \quad \text{(A.46)}
\end{aligned}$$

the verification of (A.46) is deferred to the end of the proof.

Let $m_\epsilon = [T(v+\tau_\epsilon)] - [Tv]$. Using the φ-mixing assumption in (C7) and the result in Theorem 2.2 of Serfling [19], we have, for any fixed X,

$$\frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1} E|E_v^\epsilon\{f(X+J^\epsilon_{k+[Tv]+1})-f(X)\} - E\{f(X+J^\epsilon_{k+[Tv]+1})-f(X)\}|$$
$$\leq \frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1} 2\sqrt{\varphi(k+1)}\sqrt{E[\{f(X+J^\epsilon_{k+[Tv]+1})-f(X)\}^2]}$$
$$\leq \frac{1}{\tau_\epsilon}K_1\sqrt{P(z_t\in(r_0,r_0+\kappa\epsilon])}\sum_{k=0}^{m_\epsilon-1}\sqrt{\varphi(k+1)}, \tag{A.47}$$

for some $K_1>0$; the last inequality is obtained by expanding f in a Taylor series about X and by making use of the compact support of f and (C6).

Choose a sequence $\{\tau_\epsilon\}$ such that $\lim_{\epsilon\to 0}\tau_\epsilon=0$, $\lim_{\epsilon\to 0}m_\epsilon=\infty$, and $\lim_{\epsilon\to 0}\sqrt{T}\tau_\epsilon=\infty$, which holds if, for example, $\tau_\epsilon=T^{-1/3}$. Then, (A.46) and (A.47) imply that $\hat{A}^\epsilon f^\epsilon(v)=Af(x^\epsilon(v))+o_p(1)$. Therefore, $\{x^\epsilon(v), 0\leq v\leq 1\}$ converges weakly to the compound Poisson process $\{\rho(v), 0\leq v\leq 1\}$ which is the unique solution to the martingale problem

$$f(x(t))-\int_0^t Af(x(s))ds \text{ is a martingale,} \tag{A.48}$$

for any function f with compact support and continuous second derivative, see Strook and Varadhan [21]. Consequently, $\tilde{l}(\kappa)$ converges weakly to $\tilde{l}_2(\kappa)$. Employing the Cramer-Wold device, similar arguments yield the convergence of finite-dimensional distributions of $(\{\tilde{l}(-\kappa),\kappa\geq 0\},\{\tilde{l}(+\kappa),\kappa\geq 0\})$ to those of $(\{\tilde{l}_1(\kappa),\kappa\geq 0\},\{\tilde{l}_2(\kappa),\kappa\geq 0\})$.

We complete the proof by verifying the claim in (A.46). By expanding f in Taylor series and by letting $\dot{f}(s)$ ($\ddot{f}(s)$) be the first (second) partial derivative of f with respect to s, we have, by repeated use of the mean value theorem,

$$\hat{A}^\epsilon f^\epsilon(v)-Af(x^\epsilon(v))$$
$$=\frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1}E_v^\epsilon\{f(x^\epsilon(v+k\epsilon)+J^\epsilon_{k+[Tv]+1})-f(x^\epsilon(v)+J^\epsilon_{k+[Tv]+1})\} - \frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1}E_v^\epsilon\{f(x^\epsilon(v+k\epsilon))-f(x^\epsilon(v))\}$$
$$=\frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1}E_v^\epsilon[\{\dot{f}(x^\epsilon(v+k\epsilon)+\chi_1 J^\epsilon_{k+[Tv]+1}) - \dot{f}(x^\epsilon(v)+\chi_1 J^\epsilon_{k+[Tv]+1})\}J^\epsilon_{k+[Tv]+1}], \tag{A.49}$$

for some χ_1 between 0 and 1. Therefore,

$$\begin{aligned}
&\hat{A}^\epsilon f^\epsilon(v) - Af(x^\epsilon(v)) \\
&= \frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1} E_v^\epsilon[\ddot{f}(x^\epsilon(v) + \chi_2(x^\epsilon(v+k\epsilon) - x^\epsilon(v)) + \chi_1 J^\epsilon_{k+[Tv]+1}) \\
&\qquad \times J^\epsilon_{k+[Tv]+1} \times \{x^\epsilon(v+k\epsilon) - x^\epsilon(v)\}],
\end{aligned} \tag{A.50}$$

for some χ_1 and χ_2 between 0 and 1.

Denote $\frac{1}{\tau_\epsilon}[\ddot{f}(x^\epsilon(v) + \chi_2(x^\epsilon(v+k\epsilon) - x^\epsilon(v)) + \chi_1 J^\epsilon_{k+[Tv]+1}) \times J^\epsilon_{k+[Tv]+1} \times \{x^\epsilon(v+k\epsilon) - x^\epsilon(v)\}]$ by $B_{v,k}$. Let $H_{\delta,\eta} = \{x : w_x[v, v+\delta) \le \eta\}$, for all $\eta > 0$, for $0 < \delta < 1$, and where w_x is the modulus of continuity of x defined by $w_x[v, v+\delta) = \sup_{0\le v\le 1-\delta} |x(v+\delta) - x(v)|$. The tightness of x^ϵ implies that (c.f. Billingsley [1, Theorem 15.3]) for all positive η and τ, there exists a δ such that for all ϵ sufficiently small, $P(x^\epsilon \notin H_{\delta,\eta}) \le \tau$. Let $I_1 = I(x^\epsilon \in H_{\delta,\eta})$ and $I_2 = 1 - I_1$ where $I(\cdot)$ is the indicator function. Hence, the last equation in (A.50) can be decomposed as

$$\sum_{k=0}^{m_\epsilon-1} E_v^\epsilon(B_{v,k} I_1) + \sum_{k=0}^{m_\epsilon-1} E_v^\epsilon(B_{v,k} I_2). \tag{A.51}$$

Note that the first sum is bounded by $\eta K_1 \frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1} E_v^\epsilon(J^\epsilon_{k+[Tv]+1})$ for some finite $K_1 > 0$. Using similar arguments as above, it can be checked that $\frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1} E_v^\epsilon(J^\epsilon_{k+[Tv]+1})$ is $O_p(1)$ for all sufficiently small ϵ. The fact that x^ϵ are uniformly bounded by truncation argument and using Cauchy-Schwartz inequality entail that, for some finite $K_2 > 0$, the square of the second sum in (A.51) is bounded by $K_2 E_v^\epsilon(I_2) \times \frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1} E_v^\epsilon\{(J^\epsilon_{k+[Tv]+1})^2\}$. Again the second term in the preceding product can be shown to be $O_p(1)$ for all sufficiently small ϵ. Because $E\{E_v^\epsilon(I_2)\} = E(I_2)$ which is smaller than τ for ϵ sufficiently small, $E_v^\epsilon(I_2) = \tau O_p(1)$. As η and τ can be chosen arbitrarily small, the claim follows. This completes the proof.

□

A.7. Proof of Theorem 3.4.

Proof. Because $\hat{r}$ is T-consistent and by Lemma 3.2, it follows that $\hat{r}$ and $\sqrt{T}\{(\hat{\beta}_1 - \beta_{1,0})', (\hat{\beta}_2 - \beta_{2,0})'\}'$ are asymptotically independent. Moreover, $\hat{\beta}_i = \hat{\beta}_{i,\hat{r}} = \hat{\beta}_{i,r_0} + o_p(1/\sqrt{T}), i = 1,2$; hence, $\hat{\beta}_i$ and $\hat{\beta}_{i,r_0}$ enjoy the same asymptotic distribution. But $\hat{\beta}_{i,r_0}$ is the maximum likelihood estimator of $\beta_{i,0}$ when the threshold parameter is known, for $i = 1,2$. Using Theorem 5.41 of van der Vaart [24], the sequence $\sqrt{T}(\hat{\delta}_{r_0} - \delta_0) =$

$\sqrt{T}\{(\hat{\beta}_{1,r_0} - \beta_{1,0})', (\hat{\beta}_{2,r_0} - \beta_{2,0})'\}'$ is asymptotically normal with mean zero and covariance matrix $E(\dot{\psi}_{\delta_0})^{-1}E(\psi_{\delta_0}\psi'_{\delta_0})E(\dot{\psi}_{\delta_0})^{-1}$. From this follows the result in Theorem 3.4.

□

REFERENCES

[1] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[2] Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88**, 9–25.

[3] Chan, K.S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21**, 520–533.

[4] Chan, K.S. and Tsay, R.S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika* **85**, 413–26.

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

[6] Firth, D. (1991). Generalized linear models. Chapter 3 of *Statistical Theory and Modelling,* D.V. Hinkley, N. Reid, E.J. Snell, eds. Chapman & Hall, London.

[7] Hansen, B.E. (2000). Sample splitting and threshold estimation. *Econometrica* **68**, 575–603.

[8] Ibragimov, I.A. and Has'minskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

[9] Koul, H.L., Qian, L. and Surgailis, D. (2003). Asymptotics of M-estimators in two-phase linear regression models. *Stochastic Process. Appl.* **103**, 123–154.

[10] Krebs, C.J. (1978). A review of the Chitty hypothesis of population regulation. *Canad. J. Zool.* **56**, 2463–2480.

[11] Kurtz, T.G. (1975). Semigroups of conditioned shifts and approximation of Markov processes. *Ann. Probab.* **3**, 618–642.

[12] Kushner, H.J. (1980). A martingale method for the convergence of a sequence of processes to a jump-diffusion process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **53**, 207–219.

[13] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall, London.

[14] Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135**, 370–384.

[15] Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

[16] Qian, L. (1998). On maximum likelihood estimators for a threshold autoregression. *J. Statist. Plann. Inference* **75**, 21–46.

[17] Samia, N.I., Chan, K.S. and Stenseth, N.C. (2007). A generalized threshold mixed model for analyzing nonnormal nonlinear time series, with application to plague in Kazakhstan. *Biometrika* **94**, 1, 101-118.

[18] Samia, N.I., Kausrud, K.L., Ageyev, V.S., Begon, M., Chan, K.S. and Stenseth, N.C. (2009). The dynamic structure of the full plague-wildlife-human system in Central Asia. *Submitted.*

[19] Serfling, R.J. (1968). Contributions to central limit theory for dependent variables. *Ann. of Math. Statist.* **39**, 1158–1175.

[20] Stenseth, N.C., Samia, N.I., Viljugrein, H., Kausrud, K.L., Begon, M., Davis, S., Leirs, H., Dubyanskiy, V.M., Esper, J., Ageyev, V.S., Klassovskiy, N.L., Pole, S.B. and Chan, K.S. (2006). Plague dynamics are driven by climate variation. *Proc. Natl. Acad. Sci. USA* **103**, 13110-13115.

[21] Strook, D.W. and Varadhan, S.R.S. (1971). Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.* **24**, 147–225.

[22] Tong, H. (1990). *Non-Linear Time Series. A Dynamical System Approach.* Oxford University Press, New York.

[23] Tong, H. (2007). Birth of the threshold time series model. *Statist. Sinica* **17**, (1), 8-14.

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

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

Noelle I. Samia
Department of Statistics
Northwestern University
2006 Sheridan Road
Evanston, Illinois 60208
U.S.A.
E-mail: n-samia@northwestern.edu

Kung-Sik Chan
Department of Statistics and Actuarial Science
University of Iowa
241 Schaeffer Hall
Iowa City, Iowa 52242
U.S.A.
E-mail: kung-sik-chan@uiowa.edu