CONSISTENCY AND LIMITING DISTRIBUTION OF THE MAXIMUM LIKELIHOOD ESTIMATOR 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 [30], 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 [29, 30] is perhaps the most popular nonlinear time-series models. Its extension that incorporates covariates is known as the open-loop threshold model (Tong [29]) 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 [24] and Samia *et al.* [26], 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. The GTM is an extension of the generalized linear model (Nelder and Wedderburn [21],

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

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

McCullagh and Nelder [20]), in which both non-normal response 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 [5], Chan and Tsay [8], and Qian [23] established the asymptotic behavior of the threshold estimator in the threshold autoregressive models. Hansen [13] and Koul, Qian and Surgailis [15] 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 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. We conclude in Section 5. The proofs of the results stated in Section 3 are deferred to Appendix A.

2. Model Formulation.

Let $\{a_t, t = 1, \cdots, T\}$ be a positive process, and let y_t be a random variable whose conditional probability density function given a_t, belongs to the (one-parameter canonical) exponential family, and takes the form

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

where γ_t is the natural canonical parameter, a_t are known weights, ϕ is a known dispersion parameter, and c is a normalization constant. In practice, a_t model the weight of the data cases so generally equal to 1; it is assumed that they are uniformly bounded away from 0 and $+\infty$.

Let $X = \{x_t, t = 1, \cdots, T\}$ be a p-dimensional vector covariate process that includes a univariate subprocess $Z = \{z_t, t = 1, \cdots, T\}$. Let $\tilde{x}_t$ denote the part of x_t without the lags of the response variable y_t. Denote by $\mathcal{F}_t$, the

σ-algebra generated by $a_s, y_{s-1}, \tilde{x}_s, s \le t$. The Generalized Threshold Model (GTM) specifies that, conditional on $\mathcal{F}_t$, y_t belongs to the exponential family with conditional mean μ_t given by

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

$t = 1, \cdots, T$; and with conditional variance given by $\phi a_t v(\mu_t)$, where $v(\mu_t) = \ddot{b}(\gamma_t) = \frac{\partial^2 b(\gamma_t)}{\partial \gamma_t^2}$ is the variance function. The link function g is assumed to be a known invertible smooth function whose inverse is denoted by g^{-1}. On the scale of the link function g, the conditional mean of y_t is piecewise linear and the model is assumed to be discontinuous; i.e. the regression parameters are such that $\beta_1 \neq \beta_2$, β_1 and β_2 being $p \times 1$ vectors. (For some special cases (Chan and Tsay [8]), the regression function may be continuous despite its having distinct regression parameters in the two regimes, hence a further technical condition stated in (C1) below is needed to ensure discontinuity in the regression function.) The parameter d is a non-negative integer referred to as the delay or threshold lag, and r is the threshold. 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.)

Let $\theta = (\beta_1', \beta_2', r, d)'$ denote the parameters of the GTM. The parameter space for θ is $\Omega = \Re^{2p} \times \Re \times \{0, 1, \cdots, D\}$, where D is a known upper bound of d, the delay parameter. Let $\theta_0 = (\beta_{1,0}', \beta_{2,0}', r_0, d_0)'$ denote the true parameters. Conditional on the a's, the $\tilde{x}$'s and $\mathcal{F}_1$, the log likelihood, in canonical form, is given by

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

where $\dot{b}(\gamma_t) = \frac{\partial b(\gamma_t)}{\partial \gamma_t} = \mu_t$; see McCullagh and Nelder [20] and Firth [12]. 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 $\mathcal{F}_t$; see Breslow and Clayton [3].

Each distribution belonging to the exponential family has a unique canonical link function $\eta = \dot{b}^{-1}$ for which $\eta(\mu_t) = \gamma_t = \beta_1' x_t I(z_{t-d} \le r) + \beta_2' x_t I(z_{t-d} > r)$, 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 parameters $\beta_i, i = 1, 2$, are unconstrained and that $\gamma_t = w\{\beta_1' x_t I(z_{t-d} \le r) + \beta_2' x_t I(z_{t-d} > r)\}$, 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.

$$l(\theta) = \sum_{t=1}^{T} M_{\beta_1}(y_t; a_t, x_t) I(z_{t-d} \le r) + M_{\beta_2}(y_t; a_t, x_t) I(z_{t-d} > r) + c(y_t; \phi a_t), \tag{2.5}$$

$$= \sum_{t=1}^{T} l_t(\theta), \tag{2.6}$$

where $M_{\beta_i}(y_t; a_t, x_t) = \frac{1}{\phi a_t}\{w(\beta_i' x_t) y_t - b \circ w(\beta_i' x_t)\}, i = 1, 2$, and $l_t(\theta)$ is defined as the t-th summand of (2.5). The functions ℓ_t in (2.4) differ from l_t in that the latter are functions of θ whereas the former are functions of the mean parameter. Note that

$$l(\theta) \le \sum_{t=1}^{T} \ell_t(y_t). \tag{2.7}$$

Samia, Chan and Stenseth [24] studied the specific case where the nonnegative discrete response variable equals zero in the lower regime; meaning that if the threshold is not met, the response is zero, which restriction greatly simplifies the theoretical study of the model. While the latter model is of general applicability for analyzing epidemiological time series and other time-series data (Samia, Chan and Stenseth [24] and Stenseth *et al.* [27]), the proposed GTM relaxes the restriction of zero response in the lower regime, and provides full generality of piecewise linear stochastic regression on the link scale for continuous or discrete valued time series. Before we end this section, we give two examples illustrating the GTM.

Example 1. Let $\{y_t\}$ be positive-valued time series with conditional marginal exponential distribution with mean μ_t and the log-link function:

$$\log(\mu_t) = \begin{cases} \beta_{10} + \beta_{11}\log(y_{t-1}), & \text{if } \log(y_{t-1}) \le r \\ \beta_{20} + \beta_{21}\log(y_{t-1}), & \text{if } \log(y_{t-1}) > r, \end{cases} \tag{2.8}$$

hence $z_t = \log(y_{t-1})$ and $x_t = (1, z_t)$ with the true delay parameter $d = 0$. Also, assume $\phi \equiv 1$ and $a_t \equiv 1$. This is an observation driven model; see Cox [9]. The model can be readily generalized to include higher lags of the response variable, as well as exogeneous covariates. Lawrance and Lewis [19] proposed the NEAR model for analyzing non-negative time series with stationary exponential distribution. One of their motivation for the use of the exponential marginal distribution is to handle data with high number of zeroes, e.g. in wind speed data. For data with zeroes, the above GTM can be modified by replacing $\log(y_{t-1})$ by $\log(y_{t-1}+1)$. Note that the canonical parameter space consists of the set of non-positive numbers, and hence we use the non-canonical log-link function that ensures that the regression coefficients are non-constrained.

It follows from the conditional exponential distribution that $y_t = \mu_t e_t$ where e_t is independent of the past y's and distributed as an exponential distribution of unit mean. Consequently,

$$\log(y_t) = \begin{cases} \beta_{10} + \beta_{11}\log(y_{t-1}) + \log(e_t), & \text{if } \log(y_{t-1}) \le r \\ \beta_{20} + \beta_{21}\log(y_{t-1}) + \log(e_t), & \text{if } \log(y_{t-1}) > r, \end{cases} \tag{2.9}$$

showing that $z_t = \log(y_{t-1})$ is a two-regime first-order threshold autoregressive (TAR) process. It can be readily checked that $-\log(e_t)$ is Gumbel distributed with location parameter equal to 0 and scale parameter equal to 1. The probability density function (pdf) of $\log(e_t)$ equals $h(x) = \exp(x - \exp(x)), x \in \Re$, and hence all its moments are finite. Chan et al. [6] obtained the necessary and sufficient conditions for the stationarity and ergodicity for the first-order TAR models. In particular, the above GTM is geometrically ergodic if $\beta_{11} < 1, \beta_{21} < 1, \beta_{11}\beta_{21} < 1$. These conditions will be assumed throughout this example. Indeed, geometric ergodicity of $\{z_t\}$ entails that it admits a unique stationary pdf, denoted by π, such that the

L^1-norm of $p^k(z,\cdot)$, the conditional probability density of z_{t+k} given $z_t = z$, from the stationary density π be bounded by $H(z)\rho^k$ for some $0 < \rho < 1$ and a π-integrable function H, i.e.

$$\int |p^k(z,\tilde{z}) - \pi(\tilde{z})|d\tilde{z} \le H(z)\rho^k. \tag{2.10}$$

Furthermore, Chan et al. [6] showed that, for this example, there exists some constant $K > 0$ such that $H(z) \le K + K|z|$. Moreover, the stationary density, π, of $z_t = \log(y_{t-1})$ satisfies the invariant equation:

$$\pi(z) = \int h(z - \{\beta_{10} + \beta_{11}y\}I(y \le r) - \{\beta_{20} + \beta_{21}y\}I(y > r))\pi(y)dy \tag{2.11}$$

from which it can be checked that π is positive everywhere, bounded and infinitely differentiable. The 1-step ahead conditional probability density function of $z_{t+1} = \tilde{z}$ given $z_t = z$ equals $p(z,\tilde{z}) = h(\tilde{z} - \{\beta_{10} + \beta_{11}z\}I(z \le r) - \{\beta_{20} + \beta_{21}z\}I(z > r))$ so that the joint stationary density of (z_t, z_{t+1}) at $(z,\tilde{z})$ equals $\pi(z)p(z,\tilde{z})$, is bounded and positive everywhere. It can be similarly checked that the joint density of (z_i, z_j) are everywhere positive and uniformly bounded in i and j.

For the purpose of verifying some regularity conditions stated in the next section, we claim that under the stationary distribution $\exp(M|z_t|)$ is an integrable function for any positive constant M. (Hence, H in (2.10) is π-square integrable.) This claim can be verified by routine analysis showing that the function $\exp(M|z|)$ satisfies the drift condition stated for the function g in Theorem 1 of Chan [4]. Hence, $\{y_t = \exp(z_{t+1})\}$ is stationary, and has finite moments of all orders.

To compute the log likelihood, note that the pdf of y_t can be written as $\exp(\gamma_t y_t + \log(-\gamma_t))$ where $\gamma_t = -1/\mu_t$ and $b(\gamma_t) = -\log(-\gamma_t)$. It can be checked that the log-link function implies that $w(x) = -\exp(-x)$ and hence $b \circ w$ equals the identity function; these results will be instrumental in studying the consistency of the maximum likelihood estimator.

Example 2. Consider time series of counts $\{y_t\}$ where the covariate process $\{x_t\}$ is exogeneous, i.e. it does not involve the y's. Furthermore, assume that given the covariate process, the conditional distribution of the response is Poisson and we use the log canonical link. Also, assume $\phi \equiv 1$ and $a_t \equiv 1$. It is then clear that $\{\log(\mu_t)\}$ is a stationary ergodic process and so is $\{y_t\}$ whenever $\{x_t\}$ is stationary ergodic. For this example, $b(\gamma_t) = \exp(\gamma_t)$, w is the identity function and $b \circ w = b$.

3. Large-Sample Properties of the Estimator. The following assumptions will be used to establish the asymptotic properties of the maximum likelihood estimator. All expectations in the sequel are taken under the true model, unless stated otherwise.

(C1) The regression parameters are such that $\beta_{1,0} \neq \beta_{2,0}$ and that $P\{(\beta_{1,0} - \beta_{2,0})' x_t \neq 0 | z_{t-d_0} = r_0\} > 0$. 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 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 $i \neq j$, with $0 \leq |i-j| \leq D$, are assumed to be positive everywhere and uniformly bounded. Also, there exists $\epsilon > 0$ such that for all $0 \leq j \leq D$ and β with $|\beta| = 1$,

$$P(|\beta' x_t| > \epsilon, z_{t-d_0} \in A, z_{t-j} \in B) > 0, \tag{3.1}$$

for any A of the form $(-\infty, r_0]$ or its complement, and B of the form $(-\infty, r]$ or its complement, for any real constant r.

Some explanations on these assumptions follow. Even though the regression parameters in the two regimes are distinct, the regression function may still be a continuous function. For example, this is the case with real-valued x_t and mean specification: $\mu_t = \beta_1 x_t I(x_t \leq 0) + \beta_2 x_t I(x_t > 0)$, which is a continuous function for any β_1 and β_2. Hence, (C1) imposes the further restriction that the two associated linear submodels are conditional distinct given the threshold variable at the threshold, with which we exclude the special case of a 'continuous' threshold model; see Chan and Tsay [8]. Discontinuity is an essential condition for showing that the maximum likelihood estimator of the threshold parameter r is T-consistent. Strict convexity of the cumulant function guarantees the strict concavity of the log likelihood function in (2.3) and, hence for known threshold and delay and with sufficiently large sample size, at most one global maximum likelihood estimator of the regression parameters $\beta_i, i = 1, 2$, exists.

Without the assumption (C2) 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 [5] which is a special case of a GTM with identity link and normal conditional distributions. Therefore, in the case of a threshold autoregressive (TAR) model (which is a special case of the GTM), Chan and Tong [7] have shown that under some conditions on the model parameters, the TAR model is stationary and geometric ergodic. In particular, under the condition that the sum of the magnitude of the autoregressive coefficients in each regime is less than 1 and under mild moment and density conditions of the error process, a TAR model is stationary and geometric ergodic; see Tong [29, p. 464, Example A1.2].

Assumption (C3) also requires the threshold variable to have a continuous marginal distribution satisfying some mild regularity conditions. The condition on x_t rules out redundancy in x_t, e.g. x_t cannot be linearly dependent. Note that (3.1) required by (C3) holds under very general conditions, e.g. it holds if $P(|\beta^{'} x_t| > \epsilon | z_{t-i}, z_{t-j}) > 0$ a.s. w.r.t. the joint distribution of (z_{t-i}, z_{t-j}) for $0 \le i, j \le D$. Conditions (C1–C3) ensure model identifiability of the GTM. We now state additional assumptions.

(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, an interior point of Ω_1 in the relative topology of the Euclidean space.

While the assumption of a compact parameter space, stated in Assumption (C4), may appear restrictive, it can be removed or considerably weakened in several cases of the GTM, as demonstrated in the following two lemmas. Indeed, Lemma 3.1 shows that the maximum likelihood estimator of a GTM with canonical link function is stochastically bounded under some mild regularity conditions. We now state Lemma 3.1. (Proofs of all results in this section are deferred to the Appendix.)

LEMMA 3.1. *Suppose that (C1)–(C3) hold. Assume, furthermore, that (i) the link function considered in the model is the canonical link with the canonical parameter space equals* $\Re^1$, *and (ii)* $E(|l_t(\theta_0)|) < \infty$ *and* $E(\ell_t^+(y_t)) < \infty$ *where* $\ell_t^+(y_t) = \max(\ell_t(y_t), 0)$. *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 the compact set* $\Omega_1 = \{\theta \in \Omega : |\theta - \theta_0| \le \tau\}$ *almost surely.*

The condition on the canonical link function stated in Lemma 3.1 holds for many distributions including binomial, Poisson and normal (with known variance), and it ensures that the piecewise linear stochastic regression function is unconstrained on the link scale. The condition $E(|l_t(\theta_0)|) < \infty$ holds under (C2) and if the b function is Lipschitz continuous in the sense as stated in (C5) below. (Using the canonical link is equivalent to requiring the function w to be the identity function.) Recall the notation $\ell_t(y_t)$ is the maximum log likelihood for the t-th data case with the mean $\mu_t = y_t$. For discrete response variables, the probability mass function is bounded above by 1, so $\ell_t^+(y_t) \leq 0$, hence the condition $E(\ell_t^+(y_t)) < \infty$ clearly holds for such cases. It also holds for normal distributions with known positive variance. For responses that are conditionally Gamma with k being the shape parameter, it can be checked that

$$\ell_t(y_t) = -\log(y_t) - k + k\log(k).$$

Consequently, $\ell_t^+(y_t) \leq 0$ holds if $E\{-\log(y_t)I(y_t \leq 1)\} < \infty$. As remarked earlier, for some distributions including the Gamma distribution, the canonical parameter space is a proper subset of $\Re^1$, thereby necessitating the use of a non-canonical link function to ensure that the regression function is unconstrained. However, the preceding lemma can be extended to the case that the link function is such that $b \circ w$ is the identity function and w is a strictly concave function, which is valid for some distributions, e.g. Gamma distributions. Specifically, we have the following result.

LEMMA 3.2. *Suppose that (C1)–(C3) hold. Assume, furthermore, that (i) the link function considered in the model is such that $b \circ w$ equals the identity function and w is strictly concave, (ii) $E(|l_t(\theta_0)|) < \infty$, (iii) $E(\ell_t^+(y_t)) < \infty$ and (iv) $y_t/\mu_t(\theta_0)$ has finite variance, where $\mu_t(\theta_0)$ is the true conditional mean of y_t. Then, there exists $\tau > 0$ such that, for T sufficiently large, the maximum likelihood estimator $\hat{\theta}_T$ lies in the compact set $\Omega_1 = \{\theta \in \Omega : |\theta - \theta_0| \leq \tau\}$ almost surely.*

Let $|\cdot|$ denote the absolute norm of the enclosed expression. Let

$$M_\beta(a_t, x_t, y_t) = \frac{1}{\phi a_t}\{\gamma_t y_t - b(\gamma_t)\} \tag{3.2}$$

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. The following Assumption (C5) requires the functions w and b to be Lipschitz continuous.

(C5) 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^* in a compact set, and for all x (in the support of x_t).

It follows from (C2) and (C5) that 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 all β, β^*, in a compact set, and all a_t, x_t, and y_t. The following two Assumptions (C6) and (C7) are imposed on the function $\Lambda(a_t, x_t, y_t)$.

(C6) There exist $\tau > 0$ and $M > 0$ such that, for all $z_{t-d_0} \in [r_0 - \tau, r_0 + \tau]$, $E\{\Lambda(a_t, x_t, y_t)^2 | z_{t-d_0}\} \leq M$.

(C7) There exists a $\Delta > 0$ such that the process $[\{\Lambda(a_t, x_t, y_t)I(-\Delta \leq z_{t-d_0} - r_0 \leq \Delta), z_{t-d_0}I(-\Delta \leq z_{t-d_0} - r_0 \leq \Delta)\}']$ is ρ-mixing with summable mixing coefficients.

Assumption (C6) is a conditional second-moment bound on the function $\Lambda(a_t, x_t, y_t)$. We now briefly summarize the ρ-mixing property of a stationary process, say, $\{W_t\}$. Let $\mathcal{A}$ be the σ-algebra generated by $\{W_t, t \leq j\}$ and let $\mathcal{B}$ be the σ-algebra generated by $\{W_t, t \geq j + k\}$. The process $\{W_t\}$ is said to be ρ-mixing if there exists a sequence of numbers $\{\rho(k)\}$ with $\rho(k) \to 0$ as $k \to \infty$, and such that for any square-integrable random variables f and g such that f is $\mathcal{A}$-measurable and g is $\mathcal{B}$-measurable,

$$|\text{corr}(f, g)| \leq \rho(k). \tag{3.3}$$

See Billingsley [2, §19] and Doukhan [10, p. 3 and §1.3] for further discussion of ρ-mixing. The ρ-mixing assumption stated in Assumption (C7) "controls the degree of time series dependence", as noted by Hansen [13]. For geometrically ergodic Markov processes, the ρ-mixing condition required in (C7) often holds in view of (2.10) and the fact that the process defined in (C7) confines z_{t-d_0} to lie in a compact set; this approach is taken when we demonstrate the validity of (C7) for Example 1 at the end of this section.

The following Assumption (C8) is a mild regularity condition; see Feller [11] for a discussion of weak continuity.

(C8) The conditional distribution of $(a_t, x_t')'$ given $z_{t-d_0} = z$ is weakly continuous at $z = r_0$, i.e. the conditional distribution as a function of z is continuous at r_0 in the topology of weak convergence.

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

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. The $O_p(1/T)$ fast convergence rate is due to the discontinuity of the conditional mean function; see Chan [5], Chan and Tsay [8], and Hansen [13].

THEOREM 3.2. *Suppose that $\hat{\theta}_T$ is consistent and assume that (C5)–(C8) hold. Then the maximum likelihood estimator of the threshold is T-consistent, i.e., $\hat{r} = r_0 + O_p(1/T)$, 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

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

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

(3.4)
$$\Psi_T(\delta) = \frac{1}{T}\sum_{i=1}^{T}\{\dot{M}_{\beta_1}^{'}(y_t;a_t,x_t)I(z_{t-d}\le\hat{r}), \dot{M}_{\beta_2}^{'}(y_t;a_t,x_t)I(z_{t-d}>\hat{r})\}^{'}.$$

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

$$\frac{1}{T}\sum_{i=1}^{T}\{\dot{M}_{\beta_1}^{'}(y_t;a_t,x_t)I(z_{t-d}\le r_0), \dot{M}_{\beta_2}^{'}(y_t;a_t,x_t)I(z_{t-d}>r_0)\}^{'} = 0. \tag{3.5}$$

The following classical conditions will be used to study the limiting distribution of the maximum likelihood estimator. Assumptions (D1)–(D3) require the existence of third derivatives of $M_\beta(y;a,x)$ with some (conditional or unconditional) moment bounds. Assumption (D4) essentially determines the curvature of the log likelihood function.

(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 $\beta_i \mapsto \dot{M}_\beta(y;a,x)$ is twice continuously differentiable for every $(y;a,x)$.

(D2) 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 exists function $m(y;a,x)$ such that $|\dot{M}_\beta(y;a,x)| \le m(y;a,x)$ for all y, a, x, and $\beta \in V_1 \cup V_2$. There exists $M > 0$ and $\Delta > 0$ such that $E\{m(y_t;a_t,x_t)|z_{t-d}\} \le M$ for all $z_{t-d} \in [r_0 - \Delta, r_0 + \Delta]$.

(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 the following two lemmas which are instrumental in the proof of the limiting distribution of the threshold estimator.

LEMMA 3.3. *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 $\kappa \in \Re$,

$$\tilde{l}(\kappa) = l(\hat{\delta}_{r_0+\frac{\kappa}{T}}, r_0 + \kappa/T) - l(\hat{\delta}_{r_0}, r_0). \tag{3.6}$$

LEMMA 3.4. *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 $\delta \downarrow 0$. The following theorem states that $T(\hat{r} - r_0)$ converges in distribution to some functional of the compound Poisson processes.

THEOREM 3.3. *Assume that (C1)–(C8) and (D1)–(D4) hold. Moreover, assume that in (C3), the condition imposed on the marginal density of* $(z_i, z_j)'$ *holds for all* i, j. *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 1. 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 $g(\mu_t) = \beta_1^{'} x_t I(z_{t-d} \leq r_0) + \beta_2^{'} x_t I(z_{t-d} > r_0)$.

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 2. 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.

We now revisit the two examples at the end of Section 2, and discuss when conditions (C1–C8) and (D1–D4) hold for these examples.

Example 1 Revisited. The reader may want to re-read Example 1 as we shall make heavy use of the notations and results introduced there. The cumulant $b(\gamma_t) = -\log(-\gamma_t)$ is strictly convex as its second derivative equals $\gamma_t^{-2} > 0$. Condition (C1) holds if $\beta_{10} + \beta_{11} r \neq \beta_{20} + \beta_{21} r$. Note that $a_t \equiv 1$ and $x_t = (1, z_t)^{'}$ with $z_t = \log(y_{t-1})$ so the process $\{(a_t, x_t^{'}, y_t)^{'}\}$ is stationary ergodic under the condition $\beta_{11} < 1, \beta_{21} < 1, \beta_{11}\beta_{21} < 1$. So, condition (C2) holds. The validity of (C3) follows from the positivity of the joint density derived in Example 1.

To verify (C5), consider the following expression where β, β^* lie in a compact set:

$$\begin{aligned} |w(\beta^{'} x) - w(\beta^{*'} x)| &= |\exp(-\beta^{'} x) - \exp(-\beta^{*'} x)| \\ &= \exp(-\tilde{\beta}^{'} x)|(\beta - \beta^*)^{'} x| \\ &\leq M|x|\exp(M|x|) \end{aligned}$$

for some constant $M > 0$, where $\tilde{\beta}$ is some vector in the line segment between β and β^*, and the last inequality follows from the Cauchy-Schwartz inequality. Since $b \circ w$ is the identity function, (C5) holds because $\exp(K|x_t|) = \exp(K + K|z_t|)$ has finite expectation under the stationary distribution, for all $K > 0$, implying that $|x|\exp(M|x|)$ is $\pi-$integrable. To verify (C6), we take $\Lambda = (y_t + 1)|x_t|\exp(M|x_t|)$. Now, $E(\Lambda^2|z_t) = \{\exp(2\{\beta_{10} + \beta_{11} z_t\} I(z_t \leq$

$r) + 2\{\beta_{20} + \beta_{21} z_t\} I(z_t > r)) + 1\}|z_t|^2 \exp(2M|x_t|)$, so (C6) follows. To verify (C7), notice that $\Lambda = (\mu_t e_t + 1)|x_t| \exp(M|x_t|)$, where e_t is independent of $z_{t-j}, j \geq 0$ and has an exponential distribution with unit mean, and $\mu_t = \exp(2\{\beta_{10} + \beta_{11} z_t\} I(z_t \leq r) + 2\{\beta_{20} + \beta_{21} z_t\} I(z_t > r))$. The ρ-mixing condition in (C7) then follows from the additional fact that $\{z_t\}$ is a geometrically ergodic Markov chain and in view of (2.10); see the Appendix for the proof. (C8) holds trivially because $a_t \equiv 1$ and $x_t = (1, z_t)$ so that given $z_t = z$, the conditional distribution of $(a_t, x_t^{'})^{'}$ is the probability measure concentrated on $(1, 1, z)^{'}$ which is continuous in terms of the topology of weak convergence. Note that (C4) holds under the conditions (C1–C3), owing to Lemma 3.2.

It can be shown that

$$M_\beta(y_t; a_t, x_t) = -y_t \times \exp(-\beta^{'} x_t) - \beta^{'} x_t$$

so

$$\ddot{M}_\beta(y_t; a_t, x_t) = -y_t \times \exp(-\beta^{'} x_t) x_t x_t^{'}.$$

Based on these results and routine analysis, (D1–D4) can be verified. Hence, the maximum likelihood estimator $\hat{\theta}_T$ is consistent, and the large-sample distribution results developed in this section hold for this example.

Example 2 Revisited. We shall assume that the joint pdf of $(z_t, z_{t-1}, .., z_{t-D})$ and the components in x_t other than the preceding z's is continuous and positive everywhere, which implies (C3) and (C8). Condition (C1) holds if $\beta_{1,0} \neq \beta_{2,0}$. Assume that $\{x_t\}$ is stationary and ρ-mixing with summable mixing coefficients. Note that $E(y_t^2) = E\{\exp(2\beta_1^{'} x_t I(z_{t-d} \leq r) + 2\beta_2^{'} x_t I(z_{t-d} > r))\}$, which is finite if $E\{\exp(K|x_t|)\}$ is finite for all $K > 0$, which shall be assumed; hence (C2) holds. Recall $b(\gamma_t) = \exp(\gamma_t)$, w is the identity function and $b \circ w = b$. Thus, (C5) holds because $\exp(K|x_t|)$ has finite moment for any $K > 0$. The conditions of Lemma 3.1 are then valid, so (C4) holds. We can take $\Lambda(a_t, x_t, y_t) = |y_t||x_t| + \exp(K|x_t|)|x_t|$, for some $K > 0$. After some algebra, it can be shown that condition (C6) holds if, for all $z_{t-d} \in [r_0 - \tau, r_0 + \tau]$, $E\{\exp(K|x_t|)|z_{t-d}\}$ is finite for any $K > 0$. Condition (C7) clearly holds if $\{x_t\}$ is ρ-mixing with summable mixing coefficients.

It can be shown that

$$M_\beta(y_t; a_t, x_t) = y_t \beta^{'} x_t - \exp(\beta^{'} x_t)$$

so

$$\ddot{M}_\beta(y_t; a_t, x_t) = -\exp(\beta^{'} x_t) x_t x_t^{'}.$$

Consequently, (D1–D4) can be verified by routine analysis. Therefore, the maximum likelihood estimator $\hat{\theta}_T$ is consistent, and the large-sample asymptotics developed in this section hold for this example.

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

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

$t = 1, \cdots, T$. 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. (This particular AR(2) generating mechanism follows an example used in Samia, Chan and Stenseth [24].) 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 [32].

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^{th} and 80^{th} percentiles.

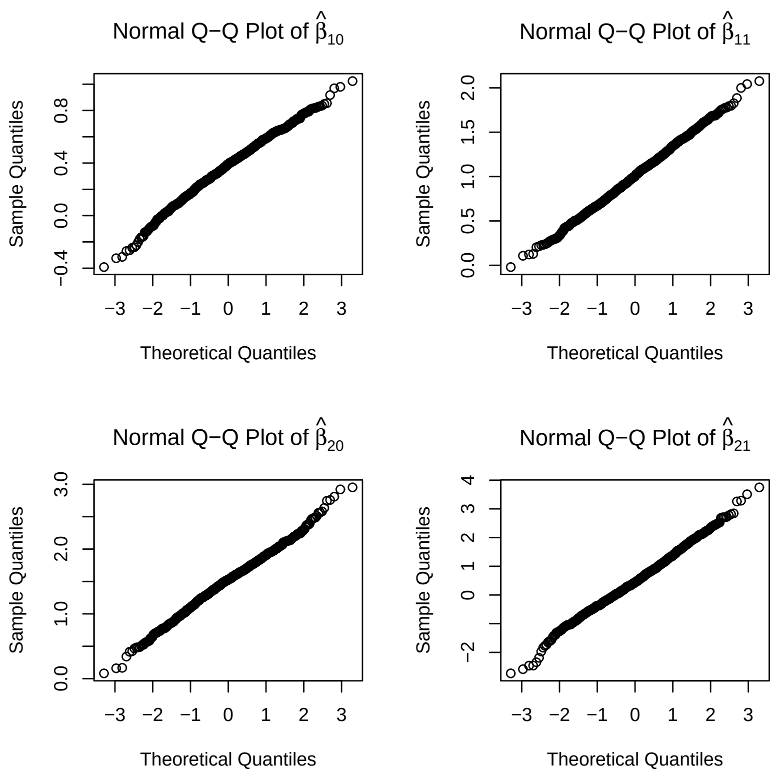

FIG 1. *Q-Q Plots for the Case when Sample Size = 100.*

5. Conclusion. The GTM is useful in modeling many epidemiological time series; e.g. the dynamics of bubonic plague in humans in Kazakhstan

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.*

[26]. Using data obtained from systematic sampling of fleas and rodents during the study period, Samia *et al.* [26] used the GTM to explain the sporadic occurrences of bubonic plague in humans. In particular, they showed that a sufficient number of viable fleas has to be achieved in order for the major human outbreaks to occur. Otherwise, if the critical threshold is not met, sporadic minor cases of human plague may occur.

In addition, the GTM is useful in modeling many other biological systems that undergo different dynamics; e.g. climate changes, Chitty hypothesis (Krebs [16]). 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 true delay parameter d is assumed to equal 0, with the integer delay parameter searched over the range $0 \leq d \leq D$, D being a known upper bound of the delay parameter. We have

$$\frac{l(\theta) - l(\theta_0)}{T} = R_{1,t} + R_{2,t} + R_{3,t} + R_{4,t}, \tag{A.1}$$

where

$$\begin{aligned} R_{1,t} = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\phi a_t} &\Big\{ (\beta_1 - \beta_{1,0})' x_t y_t - b(\beta_1' x_t) + b(\beta_{1,0}' x_t) \Big\} \\ &\times I(z_t \leq r_0, z_{t-d} \leq r)), \end{aligned} \tag{A.2}$$

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

$$\begin{aligned} R_{3,t} = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\phi a_t} &\Big\{ (\beta_2 - \beta_{2,0})' x_t y_t - b(\beta_2' x_t) + b(\beta_{2,0}' x_t) \Big\} \\ &\times I(r_0 < z_t, z_{t-d} > r), \end{aligned} \tag{A.4}$$

$$\begin{aligned} R_{4,t} = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\phi a_t} &\Big\{ (\beta_2 - \beta_{1,0})' x_t y_t - b(\beta_2' x_t) + b(\beta_{1,0}' x_t) \Big\} \\ &\times I(z_t \leq r_0, z_{t-d} > r). \end{aligned} \tag{A.5}$$

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

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

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

$$\mathcal{C}_2 = \{\theta \in \mathcal{C}_1 \, |\beta_1 - \beta_{1,0}| \leq M, |\beta_2 - \beta_{2,0}| \leq M\} \text{ a.s.} \tag{A.7}$$

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 [22, p. 8]. Although Pollard [22] assumes that the data are independent and identically distributed, this assumption can be relaxed to assuming a stationary ergodic process; see Pollard [22, p. 9]. A prototype of such checking is given in Samia and Chan [25].

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. First, consider the case that $r \geq r_0 + \Delta$. We shall determine $\Delta > \Delta_0$ below, where (C3) implies the existence of $\Delta_0 > 0$ such that $P(|z_{t-d} - r_0| > \Delta_0 | z_t = z) > 0$ for all z in some neighborhood of r_0, for all $0 \leq d \leq D$. Here, we make use of the result that the strict convexity of b implies that for all positive ϵ, there exists a positive, bounded measurable function $\eta(x)$ such that for all real numbers x, y,

$$b(y) - b(x) - \dot{b}(x)(y - x) \geq \eta(x)|y - x| I(|y - x| > \epsilon). \tag{A.8}$$

A proof of this result will be deferred after verifying both claims. For any non-zero β, we define its direction by the unit vector $\nu(\beta) = \beta/|\beta|$. For zero β, we adopt the convention that its direction $\nu(\beta) = 0$.

Let $\epsilon > 0$ be as in condition (C1). We shall bound $R_{i,t}, 1 \leq i \leq 4$ as follows. First,

$$\begin{aligned}
R_{1,t} &= \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\phi a_t}[\{(\beta_1 - \beta_{1,0})' x_t y_t - \dot{b}(\beta_{1,0}' x_t)(\beta_1 - \beta_{1,0})' x_t\} \\
&\quad - \{b(\beta_1' x_t) - b(\beta_{1,0}' x_t) - \dot{b}(\beta_{1,0}' x_t)(\beta_1 - \beta_{1,0})' x_t\}] I(z_t \leq r_0, z_{t-d} \leq r) \\
&\leq \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\phi a_t}\{(\beta_1 - \beta_{1,0})' x_t y_t - \dot{b}(\beta_{1,0}' x_t)(\beta_1 - \beta_{1,0})' x_t\} \\
&\qquad \times I(z_t \leq r_0, z_{t-d} \leq r) \\
&\quad - \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\phi a_t}\eta(\beta_{1,0}' x_t)|(\beta_1 - \beta_{1,0})' x_t| \\
&\qquad \times I(|(\beta_1 - \beta_{1,0})' x_t| > \epsilon, z_t \leq r_0, z_{t-d} \leq r),
\end{aligned}$$

$$
\begin{aligned}
R_{1,t} \leq |\beta_1 - \beta_{1,0}| \Bigg[& \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\phi a_t} \nu(\beta_1 - \beta_{1,0})' x_t \{y_t - \dot{b}(\beta_{1,0}' x_t)\} \\
& \times I(z_t \leq r_0, z_{t-d} \leq r) \\
& - \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\phi a_t} \eta(\beta_{1,0}' x_t) |\nu(\beta_1 - \beta_{1,0})' x_t| \\
& \times I(|\nu(\beta_1 - \beta_{1,0})' x_t| > \epsilon/|\beta_1 - \beta_{1,0}|, z_t \leq r_0, z_{t-d} \leq r) \Bigg],
\end{aligned}
$$

where the ratio $\epsilon/|\beta_1 - \beta_{1,0}|$ is set to be $+\infty$ if $\beta_1 = \beta_{1,0}$. Note that it can be checked that (C2) implies that, for any β with unit norm, $\frac{1}{\phi a_t}\beta' x_t\{y_t - \dot{b}(\beta_{1,0}' x_t)\} I(z_t \leq r_0, z_{t-d} \leq r)$ has zero mean and finite absolute first moment.

Similarly, $R_{2,t}$ admits the following bound:

$$
\begin{aligned}
R_{2,t} \leq & \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\phi a_t} \{(\beta_1 - \beta_{2,0})' x_t y_t - \dot{b}(\beta_{2,0}' x_t)(\beta_1 - \beta_{2,0})' x_t\} \\
& \quad \times I(z_t > r_0, z_{t-d} \leq r) \\
& - \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\phi a_t} \eta(\beta_{2,0}' x_t) |(\beta_1 - \beta_{2,0})' x_t| \\
& \quad \times I(|(\beta_1 - \beta_{2,0})' x_t| > \epsilon, z_t > r_0, z_{t-d} \leq r), \\
& \leq |\beta_1 - \beta_{2,0}| \Bigg[\frac{1}{T}\sum_{t=1}^{T} \frac{1}{\phi a_t} \nu(\beta_1 - \beta_{2,0})' x_t \{y_t - \dot{b}(\beta_{2,0}' x_t)\} \\
& \quad \times I(z_t > r_0, z_{t-d} \leq r) \\
& \quad - \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\phi a_t} \eta(\beta_{2,0}' x_t) |\nu(\beta_1 - \beta_{2,0})' x_t| \\
& \quad \times I(|\nu(\beta_1 - \beta_{2,0})' x_t| > \epsilon/|\beta_1 - \beta_{2,0}|, z_t > r_0, z_{t-d} \leq r) \Bigg].
\end{aligned}
$$

$R_{3,t}$ and $R_{4,t}$ admit similar inequalities, but for verifying Claim 1, a different approach for bounding them is needed. Indeed, the assumptions that $E(\ell_t^+(y_t)) < \infty$ and $E|l_t(\theta_0)| < \infty$ imply that for all positive $H > 0$, there exists $\Delta > 0$ such that for all $\theta \in \mathcal{B} = \{\theta \in \Omega : r > r_0 + \Delta, 0 \leq d \leq D\}$, $R_{i,t} \leq \frac{1}{T}\sum_{t=1}^{T}\{\ell_t^+(y_t) + |l_t(\theta_0)|\} I(z_{t-d} > r_0 + \Delta),\ i = 3, 4$, with the mean of the latter bound being less than H. Therefore, uniformly for all such θ,

$$
R_{3,t} + R_{4,t} \leq H + o_p(1), \tag{A.9}
$$

where $o_p(1)$ holds uniformly for $\theta \in \mathcal{B}$.

Since $\beta_{1,0} \neq \beta_{2,0}$, by re-scaling the covariate if necessary, we may and shall assume that $|\beta_{1,0} - \beta_{2,0}| \geq 2$, resulting in three cases: (I) $|\beta_1 - \beta_{1,0}| < 1, |\beta_1 - \beta_{2,0}| \geq 1$; (II) $|\beta_1 - \beta_{1,0}| \geq 1, |\beta_1 - \beta_{2,0}| < 1$; (III) $|\beta_1 - \beta_{1,0}| \geq 1, |\beta_1 - \beta_{2,0}| \geq 1$. For case (I),

$$\begin{aligned}
&R_{1,t} + R_{2,t} \\
&\leq \frac{|\beta_1 - \beta_{1,0}|}{T} \sum_{t=1}^{T} \frac{1}{\phi a_t} \nu(\beta_1 - \beta_{1,0})^{'} x_t \{y_t - \dot{b}(\beta_{1,0}^{'} x_t)\} I(z_t \leq r_0, z_{t-d} \leq r) \\
&+ \frac{|\beta_1 - \beta_{2,0}|}{T} \sum_{t=1}^{T} \frac{1}{\phi a_t} \nu(\beta_1 - \beta_{2,0})^{'} x_t \{y_t - \dot{b}(\beta_{1,0}^{'} x_t)\} I(z_t > r_0, z_{t-d} \leq r) \\
&- \frac{|\beta_1 - \beta_{2,0}|}{T} \sum_{t=1}^{T} \frac{1}{\phi a_t} \eta(\beta_{2,0}^{'} x_t) |\nu(\beta_1 - \beta_{2,0})^{'} x_t| \\
&\qquad \times I(|\nu(\beta_1 - \beta_{2,0})^{'} x_t| > \epsilon, z_t > r_0, z_{t-d} \leq r_0 + \Delta).
\end{aligned}$$

Applying the uniform law of large numbers to the three sample means on the right side of the preceding inequality, we get

$$R_{1,t} + R_{2,t} \leq o_p(1) + |\beta_1 - \beta_{2,0}| \{o_p(1) - \tau\},$$

where $\tau = \inf_{|\beta|=1} E\{\frac{1}{\phi a_t} \eta(\beta_{2,0}^{'} x_t) |\beta^{'} x_t| I(|\beta^{'} x_t| > \epsilon, z_t > r_0, z_{t-d} \leq r_0 + \Delta_0)\}$ which is positive by condition (C3); $o_p(1)$ denotes a term that converges to 0, uniformly for $\theta \in \mathcal{B}$ and such that $|\beta_1 - \beta_{1,0}| < 1$ and $|\beta_1 - \beta_{2,0}| \geq 1$. In view of (A.9), we have

$$\frac{l(\theta) - l(\theta_0)}{T} \leq o_p(1) + |\beta_1 - \beta_{2,0}| \{o_p(1) - \tau\} + H,$$

which is negative with probability approaching 1 as $T \to \infty$ if H is chosen to be less than τ. The proof for cases (II) and (III) are similar and hence omitted, which completes the proof of Claim 1.

Verification of Claim 2: We shall show that, for $M > 1$ sufficiently large, $\hat{\theta}_T \notin \mathcal{C}_2$ for T sufficiently large a.s. The complement of $\mathcal{C}_2$ consists of two cases: (I) $|\beta_1 - \beta_{1,0}| > M, |\beta_1 - \beta_{1,0}| \geq |\beta_2 - \beta_{2,0}|$, and (II) $|\beta_2 - \beta_{2,0}| > M, |\beta_2 - \beta_{2,0}| \geq |\beta_1 - \beta_{1,0}|$. Consider case (I). Recall that $R_{3,t}$ and $R_{4,t}$ admit similar inequality as for $R_{1,t}$ and $R_{2,t}$, in which case we have, uniformly for

$\theta \in \mathcal{C}_1$,

$$
\begin{aligned}
&\frac{l(\theta)-l(\theta_0)}{T}\\
&= R_{1,t} + R_{2,t} + R_{3,t} + R_{4,t}\\
&\le (|\beta_1-\beta_{1,0}| + |\beta_1-\beta_{2,0}| + |\beta_2-\beta_{2,0}| + |\beta_2-\beta_{1,0}|)\times o_p(1)\\
&- \frac{|\beta_1-\beta_{1,0}|}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}\eta(\beta_{1,0}^{'}x_t)|\nu(\beta_1-\beta_{1,0})^{'}x_t|\\
&\qquad \times I(|\nu(\beta_1-\beta_{1,0})^{'}x_t| > \epsilon/|\beta_1-\beta_{1,0}|, z_t \le r_0, z_{t-d} \le r)\\
&- 2(|\beta_1-\beta_{1,0}| + |\beta_{1,0}-\beta_{2,0}|)\times o_p(1) - |\beta_1-\beta_{1,0}|\times(\gamma + o_p(1)),
\end{aligned}
$$

where γ equals $\inf_{|\beta|=1} E\{\frac{1}{\phi a_t}\eta(\beta_{1,0}^{'}x_t)|\beta^{'}x_t|I(|\beta^{'}x_t| > \epsilon, z_t \le r_0, z_{t-d} \le r_0-\Delta)\}$, which is positive by condition (C3) and the fact that an examination of the proof of Claim 1 shows that Δ can be chosen so that the events $\{z_t \le r_0, z_{t-d} \le r_0-\Delta\}$ and $\{z_t > r_0, z_{t-d} > r_0+\Delta\}$ have positive probability for each $0 \le d \le D$. A similar inequality holds for case II, hence for M sufficiently large and uniformly for $\theta \notin \mathcal{C}_2$, $\frac{l(\theta)-l(\theta_0)}{T}$ is negative with probability approaching 1 with increasing sample size. This completes the proof of Claim 2.

It remains to demonstrate (A.8). Define

$$g(y,x) = \frac{b(y)-b(x)-\dot{b}(x)(y-x)}{y-x}.$$

For fixed x,

$$\frac{\partial g}{\partial y} = \frac{-\{b(y)-b(x)-\dot{b}(y)(y-x)\}}{(y-x)^2},$$

which is negative for $y<x$ and positive for $y>x$. Thus, for a fixed positive $\epsilon > 0$, there exists $1 \ge \eta = \eta(x) > 0$ such $g(y,x) > \eta$ for $y > x+\epsilon$ and $g(y,x) < -\eta$ for $y < x-\epsilon$, hence (A.8), after some algebra. □

A.2. Proof of Lemma 3.1.

Proof. Let $\epsilon > 0$ be as in condition (C3). Observe that the strict concavity of w, (A.8) and the condition that $b \circ w$ is the identity function imply the existence of a positive, bounded function η such that $R_{1,t}$ admits the following bound:

$$
\begin{aligned}
R_{1,t} &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}[\{w(\beta_1^{'}x_t) - w(\beta_{1,0}^{'}x_t)\}y_t - (\beta_1 - \beta_{1,0})^{'}x_t] \\
&\qquad \times I(z_t \le r_0, z_{t-d} \le r) \\
&= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}\{w(\beta_1^{'}x_t) - w(\beta_{1,0}^{'}x_t) - \dot{w}(\beta_{1,0}^{'}x_t)(\beta_1 - \beta_{1,0})^{'}x_t\}y_t \\
&\qquad \times I(z_t \le r_0, z_{t-d} \le r) \\
&+ \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}(\beta_1 - \beta_{1,0})^{'}x_t\{\dot{w}(\beta_{1,0}^{'}x_t)y_t - 1\}I(z_t \le r_0, z_{t-d} \le r) \\
&\le -\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}\eta(\beta_{1,0}^{'}x_t)|(\beta_1 - \beta_{1,0})^{'}x_t| \\
&\qquad \times I(|(\beta_1 - \beta_{1,0})^{'}x_t| > \epsilon, z_t \le r_0, z_{t-d} \le r) \\
&+ \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\phi a_t}(\beta_1 - \beta_{1,0})^{'}x_t\{\dot{w}(\beta_{1,0}^{'}x_t)y_t - 1\}I(z_t \le r_0, z_{t-d} \le r).
\end{aligned}
$$

Because $b \circ w$ is the identity function, $\dot{w}(\beta_{1,0}^{'}x_t) = 1/\dot{w}(\beta_{1,0}^{'}x_t)$, hence given $z_t \le r_0$, $\dot{w}(\beta_{1,0}^{'}x_t)y_t - 1 = y_t/\mu_t(\theta_0) - 1$. Other $R_{i,t}$'s admit similar inequalities so that we can proceed as in the proof of Lemma 3.1. □

A.3. 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 [22, p. 8] are verified; see also Pollard [22, p. 9].

We have

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

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 $\frac{1}{\phi a_t}\left[\{|w(0)| + \tilde{w}(x_t)M\}|y_t| + |b\{w(0)\}| + \tilde{m}(x_t)M\right]$, for $|\beta_1|\le M$, $M>0$, and for $\tilde{w}$ and $\tilde{m}$ defined in (C5). Using a similar argument as in Samia and Chan [25] 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.

Lemma 5.35 of van der Vaart [31], 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

$$\max_{\theta\in\Omega_1:|\theta-\theta_0|\ge\epsilon} E\left(\frac{l(\theta)}{T}\right) + \delta < E\left(\frac{l(\theta_0)}{T}\right) - \delta. \tag{A.10}$$

Applying the uniform law of large numbers and by making use of (A.10), we conclude that, for all $\epsilon>0$, there exists $\delta>0$ such that, for T sufficiently large, and uniformly for $|\theta-\theta_0|\ge\epsilon$, $\frac{l(\theta)}{T}\le E\left(\frac{l(\theta)}{T}\right)+\delta \le \max_{\theta\in\Omega_1:|\theta-\theta_0|\ge\epsilon}$ $E\left(\frac{l(\theta)}{T}\right)+\delta < E\left(\frac{l(\theta_0)}{T}\right) - \delta < \frac{l(\theta_0)}{T}$ almost surely. Hence, for T sufficiently large, $\left|\hat{\theta}_T-\theta_0\right|\le\epsilon$ almost surely. As $\epsilon>0$ is arbitrary, $\hat{\theta}_T\to\theta_0$ almost surely. This completes the proof. □

A.4. 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, $\Omega_1 = \{\theta\in\Omega : |\beta_i-\beta_{i,0}|<\Delta, i=1,2; |r-r_0|<\Delta\}$, 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 $|M_\beta(a_t, x_t, y_t) - M_{\beta^*}(a_t, x_t, y_t)| \le \Lambda(a_t, x_t, y_t)\,|\beta \ - \beta^*|$, 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 $W=\left[\{M_tI\left(-\Delta\le z_t\le\Delta\right),z_tI\left(-\Delta\le z_t\le\Delta\right)\}'\right]$ is ρ-mixing with summable ρ-mixing coefficients. Then, for all $\epsilon>0$, for all $\zeta>0$, there exists $K>0$ such that, for all T,

$$P\left(\sup_{\frac{K}{T}<r\le\Delta}\left|\sum\frac{I(0<z_t\le r)}{TQ(r)}-1\right|<\zeta\right)>1-\epsilon, \tag{A.11}$$

and

$$P\left(\sup_{\frac{K}{T}<r\le\Delta}\left|\sum\frac{M_tI(0<z_t\le r)-E\left\{M_tI(0<z_t\le r)\right\}}{TQ(r)}\right|<\zeta\right)>1-\epsilon. \tag{A.12}$$

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

$$\frac{l(\beta_1,\beta_2,r)-l(\beta_1,\beta_2,0)}{TQ(r)}<\left(|\beta_1-\beta_{1,0}|+|\beta_2-\beta_{2,0}|\right)(\zeta+M)+\zeta+\kappa, \tag{A.13}$$

where

$$\kappa=\frac{1}{TQ(r)}\sum_{t=1}^{T}E\left\{M_{\beta_{1,0}}(a_t,x_t,y_t)-M_{\beta_{2,0}}(a_t,x_t,y_t)|z_t\right\}I(0<z_t\le r). \tag{A.14}$$

Because of Assumptions (C1) and (C8), and Lemma 5.35 of van der Vaart [31], we have that for each $z_t \in (0, \Delta]$, $E\{M_{\beta_{1,0}}(a_t, x_t, y_t) - M_{\beta_{2,0}}(a_t, x_t, y_t) |z_t\}$ 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

$$Q_T(r) = \sum \frac{I(0 < z_t \leq r)}{T}, \tag{A.15}$$

$$R_T(r) = \sum \frac{M_t I(0 < z_t \leq r)}{T}, \tag{A.16}$$

$$\tilde{R}_T(r_1, r_2) = \sum \frac{M_t I(r_1 < z_t \leq r_2)}{T}. \tag{A.17}$$

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 $(0, \Delta)$,

$$mr \leq Q(r) \leq Mr. \tag{A.18}$$

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

$$\operatorname{var}\left\{I(0 < z_t \leq r)\right\} \leq HQ(r). \tag{A.19}$$

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 $(0, \Delta)$,

$$E\left\{M_t I(r_1 < z_t \leq r_2)\right\} \leq H\left\{Q(r_2) - Q(r_1)\right\}. \tag{A.20}$$

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

$$\operatorname{var}\left\{M_t I(r_1 < z_t \leq r_2)\right\} \leq H\left\{Q(r_2) - Q(r_1)\right\}. \tag{A.21}$$

Let $R_t = M_t I(r_1 < z_t \leq r_2)$. Because the process $W = \left[\left\{M_t I\left(-\Delta \leq z_t \leq \Delta\right), z_t I\left(-\Delta \leq z_t \leq \Delta\right)\right\}'\right]$ is ρ-mixing, then $|\operatorname{Cov}(R_t, R_s)| \leq \rho(|t - s|) \times$

$\{E(R_t^2)\}^{\frac{1}{2}}\{E(R_s^2)\}^{\frac{1}{2}}$; see Doukhan [10, p. 9]. Because the ρ-mixing coefficient is assumed to be summable, and by making use of the stationarity assumption, we have $|\mathrm{Cov}(R_t, R_s)| \le \rho(|t-s|)E(R_t^2) \le \rho(|t-s|)H\{Q(r_2) - Q(r_1)\}$, for some $H > 0$. Hence, we make use of the latter inequality for the covariance of ρ-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

$$\mathrm{var}\{TQ_T(r)\} \le THQ(r), \tag{A.22}$$

$$\mathrm{var}\left\{T\tilde{R}_T(r_1, r_2)\right\} \le TH\{Q(r_2) - Q(r_1)\}, \tag{A.23}$$

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

Therefore, Claim I can be verified by making use of the inequalities (A.18)–(A.24), and by employing arguments as in Chan [5, p. 529].

□

A.5. Proof of Lemma 3.3.

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

$$\begin{aligned} & l(\delta, r) - l(\hat{\delta}_{r_0}, r) \\ &= (\delta - \hat{\delta}_{r_0})'\dot{l}(\hat{\delta}_{r_0}, r) + \frac{1}{2}(\delta - \hat{\delta}_{r_0})'\ddot{l}(\hat{\delta}_{r_0}, r)(\delta - \hat{\delta}_{r_0}) + R_T(\tilde{\delta}, \delta, \hat{\delta}_{r_0}), \end{aligned} \tag{A.25}$$

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

$$\lim_{T\to\infty} \sup_{|\delta - \hat{\delta}_{r_0}|\to 0} \frac{|R_T|}{T|\delta - \hat{\delta}_{r_0}|^2} = 0. \tag{A.26}$$

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 $|r - r_0| \leq \frac{K}{T}$, we have $E(|\dot{l}(\hat{\delta}_{r_0}, r)|) = \sum_{i=1}^{2} \sum_{t=1}^{T} E\{|\dot{M}_{\hat{\beta}_{i,r_0}}| I(r_0 < z_{t-d} \leq r)\} \leq 2TM_1 P(r_0 < z_{t-d} \leq r) = O(1)$. It follows readily from Markov's inequality that for T sufficiently large, for all $K > 0$ and $|r - r_0| \leq \frac{K}{T}$, we have

$$|\dot{l}(\hat{\delta}_{r_0}, r)| = O_p(1). \tag{A.27}$$

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

$$\begin{aligned} \ddot{l}(\hat{\delta}_{r_0}, r) &= \left\{\ddot{l}(\hat{\delta}_{r_0}, r) - \ddot{l}(\hat{\delta}_{r_0}, r_0)\right\} + \ddot{l}(\hat{\delta}_{r_0}, r_0) \\ &= \begin{bmatrix} \eta_1 + \xi_1 & 0 \\ 0 & -\eta_2 + \xi_2 \end{bmatrix}, \end{aligned} \tag{A.28}$$

where $\eta_i = \sum_{t=1}^{T} \ddot{M}_{\hat{\beta}_{i,r_0}} I(r_0 < z_{t-d} \leq r), i = 1, 2$, $\xi_1 = \sum_{t=1}^{T} \ddot{M}_{\hat{\beta}_{1,r_0}} I(z_{t-d} \leq r_0)$, $\xi_2 = \sum_{t=1}^{T} \ddot{M}_{\hat{\beta}_{2,r_0}} I(z_{t-d} > r_0)$, 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 $|r - r_0| \leq \frac{K}{T}$, we have

$$|\eta_i| = O_p(1), i = 1, 2. \tag{A.29}$$

In reference to Example 19.8 of van der Vaart [31], it can be easily shown that the collection of functions $\{\ddot{M}_\beta, \beta$ in a fixed compact set$\}$ is Glivenko-Cantelli. Hence, using the argument of van der Vaart [31, p. 279], we have for T sufficiently large,

$$\frac{1}{T}\xi_1 = E\{\ddot{M}_{\beta_{1,0}} I(z_{t-d} \leq r_0)\} + o_p(1), \tag{A.30}$$

$$\frac{1}{T}\xi_2 = E\{\ddot{M}_{\beta_{2,0}} I(z_{t-d} > r_0)\} + o_p(1); \tag{A.31}$$

where $E\{\ddot{M}_{\beta_{1,0}} I(z_{t-d} \leq r_0)\}$ 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.29)–(A.31) with the result in (A.28), 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

$$\begin{aligned}
&\frac{1}{2}(\delta - \hat{\delta}_{r_0})' \ddot{l}(\hat{\delta}_{r_0}, r)(\delta - \hat{\delta}_{r_0}) \\
&\le \frac{1}{2}\sum_{i=1}^{2}\{|\beta_i - \hat{\beta}_{i,r_0}|^2|\eta_i| + (\beta_i - \hat{\beta}_{i,r_0})'\xi_i(\beta_i - \hat{\beta}_{i,r_0})\} \\
&\le \frac{1}{2}\sum_{i=1}^{2}|\beta_i - \hat{\beta}_{i,r_0}|^2[O_p(1) - T\{2\lambda - o_p(1)\}], \qquad \text{(A.32)}
\end{aligned}$$

for some scalar $\lambda > 0$.

Finally, we combine the results in (A.26), (A.27), and (A.32) with the result in (A.25). 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

$$\begin{aligned}
l(\delta, r) - l(\hat{\delta}_{r_0}, r) &\le a_T O_p(1) + \frac{1}{2}a_T^2[O_p(1) - T\{2\lambda - o_p(1)\}] + Ta_T^2 o_p(1) \\
&\le Ta_T^2\{-2\lambda + o_p(1)\}, \qquad \text{(A.33)}
\end{aligned}$$

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.6. Proof of Lemma 3.4.

Proof. We use the same notations as in the proof of Lemma 3.3. 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

$$\begin{aligned}
\tilde{l}(\kappa) &= \{l(\hat{\delta}_{r_0+\frac{\kappa}{T}}, r_0 + \kappa/T) - l(\hat{\delta}_{r_0+\frac{\kappa}{T}}, r_0)\} + \{l(\hat{\delta}_{r_0+\frac{\kappa}{T}}, r_0) - l(\hat{\delta}_{r_0}, r_0)\} \\
&= \sum_{t=1}^{T}(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}})I(r_0 < z_{t-d} \le r_0 + \kappa/T) \qquad \text{(A.34)} \\
&\quad + (M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{1,r_0}})I(z_{t-d} \le r_0) \qquad \text{(A.35)} \\
&\quad + (M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{2,r_0}})I(z_{t-d} > r_0). \qquad \text{(A.36)}
\end{aligned}$$

We first consider equation (A.34); we have

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

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

$$\left|\sum_{t=1}^{T}(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}})I(r_0 < z_{t-d} \leq r_0 + \kappa/T) - \{l(\delta_0, r_0 + \kappa/T) - l(\delta_0, r_0)\}\right|$$
$$\leq (|\hat{\beta}_{1,r_0+\frac{\kappa}{T}} - \beta_{1,0}| + |\hat{\beta}_{2,r_0+\frac{\kappa}{T}} - \beta_{2,0}|) \times \sum_{t=1}^{T}\Lambda(a_t, x_t, y_t)I(r_0 < z_{t-d} \leq r_0 + \kappa/T); \quad \text{(A.37)}$$

the latter inequality holds because of (C5). Because $E\{\Lambda(a_t, x_t, y_t)I(r_0 < z_{t-d} \leq r_0 + \kappa/T)\} = O(1/T)$, it follows that for T sufficiently large, $\sum_{t=1}^{T}\Lambda(a_t, x_t, y_t)I(r_0 < z_{t-d} \leq r_0 + \kappa/T) = O_p(1)$. On the other hand, for T sufficiently large, for all $K > 0$ and uniformly for all $|\kappa| \leq K$, 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 3.3 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 $|\kappa| \leq K$, the inequality in (A.37) entails that

$$\left|\sum_{t=1}^{T}(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}})I(r_0 < z_{t-d} \leq r_0 + \kappa/T) - \{l(\delta_0, r_0 + \kappa/T) - l(\delta_0, r_0)\}\right| = o_p(1). \quad \text{(A.38)}$$

Next, we consider equation (A.35). 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

$$
\begin{aligned}
&\sum_{t=1}^{T}(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{1,r_0}})I(z_{t-d} \le r_0) \\
&\quad \le (\hat{\beta}_{1,r_0+\frac{\kappa}{T}} - \hat{\beta}_{1,r_0})' \sum_{t=1}^{T} \dot{M}_{\beta_{1,0}} I(z_{t-d} \le r_0) \\
&\qquad + \frac{1}{2}(\hat{\beta}_{1,r_0+\frac{\kappa}{T}} - \hat{\beta}_{1,r_0})' \sum_{t=1}^{T} \ddot{M}_{\beta_{1,0}} I(z_{t-d} \le r_0) \\
&\qquad\qquad \times (\hat{\beta}_{1,r_0+\frac{\kappa}{T}} + \hat{\beta}_{1,r_0} - 2\beta_{1,0}) + r_T,
\end{aligned}
\tag{A.39}
$$

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 $\sum_{t=1}^{T}\sum_{j=1}^{p} c_j \dot{M}_{\beta_{1,0}}^{(j)} \times I(z_{t-d} \le r_0)$, for all nonzero vectors of constants $c = (c_1, \cdots, c_p)$. Using Cramer-Wold device, it follows that for all T sufficiently large, $|\sum_{t=1}^{T} \dot{M}_{\beta_{1,0}} \times I(z_{t-d} \le r_0)| = O_p(\sqrt{T})$. 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 $\mathcal{F}_t = \sigma(a_t, x_t, y_{t-k}, a_{t-k}, x_{t-k}, k \ge 1)$ and because $E\{\dot{M}_{\beta_{1,0}}\dot{M}_{\beta_{1,0}}' I(z_{t-d} \le r_0)\} = -E\{\ddot{M}_{\beta_{1,0}} I(z_{t-d} \le r_0)\}$ is finite; see Billingsley [1]. On the other hand, by the law of large numbers, for all T sufficiently large, $\frac{1}{T}\sum_{t=1}^{T} \ddot{M}_{\beta_{1,0}} I(z_{t-d} \le r_0)$ converges to $E\{\ddot{M}_{\beta_{1,0}} I(z_{t-d} \le r_0)\}$ in probability. Thus, for all T sufficiently large, for all $K > 0$ and uniformly for all $|\kappa| \le K$, the inequality in (A.39) yields

$$
\begin{aligned}
&\left|\sum_{t=1}^{T}(M_{\hat{\beta}_{1,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{1,r_0}})I(z_{t-d} \le r_0)\right| \\
&\le o_p(1/\sqrt{T})O_p(\sqrt{T}) + o_p(1/\sqrt{T})O_p(T)O_p(1/\sqrt{T}) + o_p(1) = o_p(1),
\end{aligned}
\tag{A.40}
$$

using Lemma 3.3. Similarly, it can be shown that for all T sufficiently large, for all $K > 0$ and uniformly for all $|\kappa| \le K$, we have

$$
\left|\sum_{t=1}^{T}(M_{\hat{\beta}_{2,r_0+\frac{\kappa}{T}}} - M_{\hat{\beta}_{2,r_0}})I(z_{t-d} > r_0)\right| = o_p(1). \tag{A.41}
$$

Combine the results in (A.38), (A.40), and (A.41) with the results in (A.34)–(A.36) to complete the proof.

□

A.7. Proof of Theorem 3.3.

PROOF. Owing to Lemma 3.4, 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 [14, 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

$$\begin{aligned} x^\epsilon(v) &= X^\epsilon_{[Tv]},\ 0 \leq v \leq 1, \\ X^\epsilon_0 &= 0,\ X^\epsilon_{t+1} = X^\epsilon_t + J^\epsilon_{t+1},\ t = 0, 1, 2, \cdots \\ J^\epsilon_t &= \{M_{\beta_{1,0}}(y_t; a_t, x_t) - M_{\beta_{2,0}}(y_t; a_t, x_t)\}I(r_0 < z_t \leq r_0 + \kappa\epsilon). \end{aligned} \tag{A.42}$$

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 $\{\mathcal{C}(v), 0 \leq v \leq 1\}$, 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 [18] via operator convergence. By employing truncation arguments as in Kushner [18], we can and will assume that x^ϵ are uniformly bounded. (We omit the technical details showing that the limiting result obtained below preserves for the original, non-truncated data by passing to the limit with increasingly negligible truncation; see Kushner [18].) 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 $\sup_v E|f(v)| < \infty$. Let E^ϵ_v denote the conditional expectation given $\mathcal{F}^\epsilon_v$, which is the σ-algebra generated by $\{x^\epsilon(u), u \leq v\}$.

Note $\mathcal{F}_v^\epsilon$ is a subset of $\mathcal{F}_v$. For f and $f^\delta \in \overline{\mathscr{L}}$, define p-$\lim_{\delta\to 0} f^\delta = f$ 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-$\lim_{\delta\to 0}[\frac{1}{\delta}\{E_v^\epsilon f(v+\delta) - f(v)\} - g(v)] = 0$. 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 [17, p. 625]. We next study the limiting behavior of $\hat{A}^\epsilon f^\epsilon$. We have

$$\begin{aligned}
\hat{A}^\epsilon f^\epsilon(v) &= \frac{1}{\tau_\epsilon}[E_v^\epsilon\{f(x^\epsilon(v+\tau_\epsilon))\} - f(x^\epsilon(v))] \\
&= \frac{1}{\tau_\epsilon} \sum_{k=[Tv]}^{[T(v+\tau_\epsilon)]-1} E_v^\epsilon\{f(X_{k+1}^\epsilon) - f(X_k^\epsilon)\} \\
&= \frac{1}{\tau_\epsilon} \sum_{k=0}^{[T(v+\tau_\epsilon)]-[Tv]-1} E_v^\epsilon\{f(X_{k+[Tv]}^\epsilon + J_{k+[Tv]+1}^\epsilon) - f(X_{k+[Tv]}^\epsilon)\}. \qquad \text{(A.43)}
\end{aligned}$$

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

$$\begin{aligned}
\hat{A}^\epsilon f^\epsilon(v) = \frac{1}{\tau_\epsilon} &\sum_{k=0}^{[T(v+\tau_\epsilon)]-[Tv]-1} E_v^\epsilon\{f(x^\epsilon(v) + J_{k+[Tv]+1}^\epsilon) - f(x^\epsilon(v))\} \\
&+ o_p(1), \qquad \text{(A.44)}
\end{aligned}$$

the verification of (A.44) is deferred to the end of the proof.

Let $m_\epsilon = [T(v+\tau_\epsilon)] - [Tv]$. Using the ρ-mixing assumption in (C7) and the result of Billingsley [2, p. 261, Equation (48)], we have, for any fixed X,

$$\begin{aligned}
&\frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1} E|E_v^\epsilon\{f(X+J^\epsilon_{k+[Tv]+1})-f(X)\} \\
&\qquad\qquad - E\{f(X+J^\epsilon_{k+[Tv]+1})-f(X)\}| \\
&\leq \frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1}\rho(k+1)\sqrt{E[\{f(X+J^\epsilon_{k+[Tv]+1})-f(X)\}^2]} \\
&\leq \frac{1}{\tau_\epsilon}K_1\sqrt{P(z_t\in(r_0,r_0+\kappa\epsilon])}\sum_{k=0}^{m_\epsilon-1}\rho(k+1), \qquad \text{(A.45)}
\end{aligned}$$

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.44) and (A.45) 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 $\{\mathcal{C}(v), 0\leq v\leq 1\}$ which is the unique solution to the martingale problem

$$f(x(t)) - \int_0^t Af(x(s))ds \text{ is a martingale}, \qquad \text{(A.46)}$$

for any function f with compact support and continuous second derivative, see Strook and Varadhan [28]. 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.44). 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,

$$\begin{aligned}
&\hat{A}^\epsilon f^\epsilon(v) - Af(x^\epsilon(v)) \\
&= \frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1} E_v^\epsilon\{f(x^\epsilon(v+k\epsilon)+J^\epsilon_{k+[Tv]+1}) - f(x^\epsilon(v)+J^\epsilon_{k+[Tv]+1})\} \\
&\qquad - \frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1} E_v^\epsilon\{f(x^\epsilon(v+k\epsilon)) - f(x^\epsilon(v))\} \\
&= \frac{1}{\tau_\epsilon}\sum_{k=0}^{m_\epsilon-1} E_v^\epsilon[\{\dot{f}(x^\epsilon(v+k\epsilon)+\chi_1 J^\epsilon_{k+[Tv]+1}) \\
&\qquad\qquad - \dot{f}(x^\epsilon(v)+\chi_1 J^\epsilon_{k+[Tv]+1})\}J^\epsilon_{k+[Tv]+1}], \qquad \text{(A.47)}
\end{aligned}$$

for some χ_1 between 0 and 1. Therefore,

$$\begin{aligned}
&\hat{A}^\epsilon f^\epsilon(v) - Af(x^\epsilon(v)) \\
&\quad = \frac{1}{\tau_\epsilon} \sum_{k=0}^{m_\epsilon - 1} E_v^\epsilon[\ddot{f}(x^\epsilon(v) + \chi_2(x^\epsilon(v+k\epsilon) - x^\epsilon(v)) + \chi_1 J_{k+[Tv]+1}^\epsilon) \\
&\qquad\qquad \times J_{k+[Tv]+1}^\epsilon \times \{x^\epsilon(v+k\epsilon) - x^\epsilon(v)\}], \qquad \text{(A.48)}
\end{aligned}$$

for some χ_1 and χ_2 between 0 and 1.

Denote $\frac{1}{\tau_\epsilon}[\ddot{f}(x^\epsilon(v) + \chi_2(x^\epsilon(v+k\epsilon) - x^\epsilon(v)) + \chi_1 J_{k+[Tv]+1}^\epsilon) \times J_{k+[Tv]+1}^\epsilon \times \{x^\epsilon(v+k\epsilon) - x^\epsilon(v)\}]$ 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.48) can be decomposed as

$$\sum_{k=0}^{m_\epsilon - 1} E_v^\epsilon(B_{v,k} I_1) + \sum_{k=0}^{m_\epsilon - 1} E_v^\epsilon(B_{v,k} I_2). \qquad \text{(A.49)}$$

Note that the first sum is bounded by $\eta K_1 \frac{1}{\tau_\epsilon} \sum_{k=0}^{m_\epsilon - 1} E_v^\epsilon(J_{k+[Tv]+1}^\epsilon)$ for some finite $K_1 > 0$. Using similar arguments as above, it can be checked that $\frac{1}{\tau_\epsilon} \sum_{k=0}^{m_\epsilon - 1} E_v^\epsilon(J_{k+[Tv]+1}^\epsilon)$ is $O_p(1)$ for all sufficiently small ϵ. The fact that x^ϵ is 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.49) is bounded by $K_2 E_v^\epsilon(I_2) \times \frac{1}{\tau_\epsilon} \sum_{k=0}^{m_\epsilon - 1} E_v^\epsilon\{(J_{k+[Tv]+1}^\epsilon)^2\}$. 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.8. Proof of Theorem 3.4.

Proof. Because $\hat{r}$ is T-consistent and by Lemma 3.3, 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 [31], 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.

□

A.9. Proof of (7) for Example 1.

PROOF. Let $w_t = [\{\Lambda(a_t, x_t, y_t)I(-\Delta \le z_t - r_0 \le \Delta), z_t I(-\Delta \le z_t - r_0 \le \Delta)\}']$ and U be the interval $[r_0-\Delta, r_0+\Delta]$. Let k be a positive integer greater than 1. Consider two random variables $f(w_{t+k})$ and $g(w_t)$ that are of zero mean and finite variance. Clearly, $E(f(w_{t+k})|\mathcal{F}_{t+k})$ can be written as $\tilde{f}(z_{t+k}I(z_{t+k}\in U))$, for some function $\tilde{f}$. Also $g(w_t)=\tilde{g}(z_t, z_{t+1})$ for some function $\tilde{g}$. Now,

$$
\begin{aligned}
&E\{f(w_{t+k})g(w_t)\}\\
&= E\{\tilde{f}(z_{t+k}I(z_{t+k}\in U))\tilde{g}(z_t, z_{t+1})\}\\
&= \int\int\int \tilde{f}(z_{t+k}I(z_{t+k}\in U))\{p^{k-1}(z_{t+1}, z_{t+k}) - \pi(z_{t+k})\}\pi_{t,t+1}(z_t, z_{t+1})\\
&\qquad \times \tilde{g}(z_t, z_{t+1})dz_t dz_{t+1} dz_{t+k}\\
&= \int_U\int\int \frac{\tilde{f}(z_{t+k}I(z_{t+k}\in U))}{\pi(z_{t+k})}\pi(z_{t+k})\{p^{k-1}(z_{t+1}, z_{t+k}) - \pi(z_{t+k})\}\\
&\qquad \times \pi_{t,t+1}(z_t, z_{t+1})\tilde{g}(z_t, z_{t+1})dz_t dz_{t+1} dz_{t+k}\\
&\quad + \tilde{f}(0)\int_{U^c}\int\int \{p^{k-1}(z_{t+1}, z_{t+k}) - \pi(z_{t+k})\}\pi_{t,t+1}(z_t, z_{t+1})\\
&\qquad \times \tilde{g}(z_t, z_{t+1})dz_t dz_{t+1} dz_{t+k}.
\end{aligned}
$$

Because the stationary density π is a positive and continuous function, there exists $K>1$ such that $1/\pi(z)\le K$ over U. Consequently, in view of (2.10) and by increasing K if necessary,

$$
\begin{aligned}
|E\{(f(w_{t+k})g(w_t)\}| &\le K\rho^{k-1}E\{|\tilde{f}(z_{t+k}I(z_{t+k}\in U))|\}E\{|H(z_{t+1})\tilde{g}(z_t, z_{t+1}|\}\\
&\le \rho^{k-1}KE^{1/2}\{H^2(z_{t+1})\}\sqrt{\operatorname{var}(f(w_{t+k})\operatorname{var}(g(w_t))}.
\end{aligned}
$$

Recall $H(z_t)$ has finite second moment. Hence, the correlation between $f(w_{t+k})$ and $g(w_t)$ decays to 0 geometrically as $k\to\infty$. Thanks to the Markov property of $\{z_t\}$, the geometric decay rate holds for general f and g measurable w.r.t. the σ-algebra generated by $\{w_{t+j}, j\ge k\}$ and that by $\{w_{t+j}, j\le 0\}$, respectively, so that the required ρ-mixing condition holds as the mixing coefficients decay to 0 geometrically. □

REFERENCES

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

[2] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York.

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

[4] Chan, K.S. (1990). A note on the geometric ergodicity of a Markov chain. *Advances in Applied Probability*, **21**, 702-704.

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

[6] Chan, K.S., Petruccelli, J.D., Tong, H. and Woolford, S.W. (1985). A multiple threshold AR(1) model, *Journal of Applied Probability*, **22**, 267-279.

[7] Chan, K.S. and Tong, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. in Appl. Probab.* **17**, 666–678.

[8] 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.

[9] Cox, D.R. (1981). Statistical analysis of time series: some recent developments. *Scan. J. Statist.*, **8**, 93–115.

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

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

[12] 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.

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

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

[15] 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.

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

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

[18] 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.

[19] Lawrance, A.J. and Lewis, P.A.W. (1985). Modelling and residual analysis of non-linear autoregressive time series in exponential variables (with Discussion). *J. R. Stat. Soc.*, **B47**, 165-202.

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

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

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

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

[24] 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.

[25] Samia, N.I. and Chan, K.S. (2009). Maximum likelihood estimation of a generalized threshold model. Technical Report **398**, Department of Statistics and Actuarial Science, University of Iowa.

[26] Samia, N.I., Kausrud, K.L., Ageyev, V.S., Begon, M., Chan, K.S. and Stenseth, N.C. (2010). The dynamics of the plague-wildlife-human system in Central Asia is controlled by two epidemiological thresholds. *Submitted.*

[27] 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.

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

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

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

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

[32] 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