The Annals of Applied Statistics
2012, Vol. 6, No. 3, 1256–1279
DOI: 10.1214/12-AOAS546

ORDER SELECTION IN NONLINEAR TIME SERIES MODELS WITH APPLICATION TO THE STUDY OF CELL MEMORY[1]

BY YING HUNG

Rutgers University

Cell adhesion experiments are biomechanical experiments studying the binding of a cell to another cell at the level of single molecules. Such a study plays an important role in tumor metastasis in cancer study. Motivated by analyzing a repeated cell adhesion experiment, a new class of nonlinear time series models with an order selection procedure is developed in this paper. Due to the nonlinearity, there are two types of overfitting. Therefore, a double penalized approach is introduced for order selection. To implement this approach, a global optimization algorithm using mixed integer programming is discussed. The procedure is shown to be asymptotically consistent in estimating both the order and parameters of the proposed model. Simulations show that the new order selection approach outperforms standard methods. The finite-sample performance of the estimator is also examined via a simulation study. The application of the proposed methodology to a T-cell experiment provides a better understanding of the kinetics and mechanics of cell adhesion, including quantifying the memory effect on a repeated unbinding force experiment and identifying the order of the memory.

1. Introduction. Cell adhesion plays an important role in many physiological and pathological processes, especially in tumor metastasis in cancer study. Cell adhesion experiments refer to biomechanical experiments that study the binding of cells at the molecular level. The binding is mediated by specific interaction between cell adhesion proteins, called receptors, and the molecules that they bind to, called ligands. The resulting bond is called the receptor-ligand bond. There are various types of measurements in the cell adhesion experiments to study different aspects of the binding, such as the binding frequency and bond lifetime measurements [Zarnitsyna et al. (2007), Huang et al. (2010)]. This research is inspired by analyzing a specific type of cell adhesion experiment known as the unbinding force assay [Marshall et al. (2003, 2005)].

Receptor-ligand bonds that mediate cell adhesion are often subjected to forces that regulate their dissociation; therefore, an important issue is to study the unbinding force of a receptor-ligand bond. To address this issue, the unbinding force assay is developed by using a high-tech version of the micropipette known as the biomembrane force probe [Chen et al. (2008)]. A biomembrane force probe is

Received July 2011; revised January 2012.
[1]Supported by NSF Grants DMS-09-05753 and CMMI 0927572.
Key words and phrases. Consistency, micropipette experiment, order selection, single molecule, threshold autoregressive model.

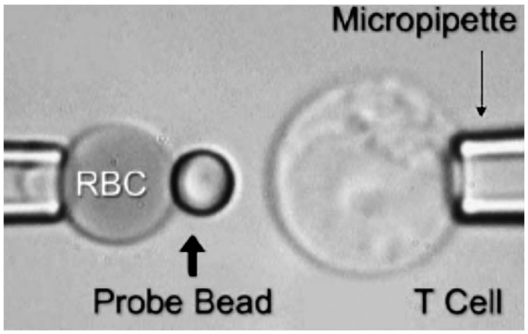

FIG. 1. *Illustration of the biomembrane force probe.*

illustrated in Figure 1 where a probe bead (left) is attached to the apex of the micropipette-aspirated red blood cell to allow tracking of the deflection of another cell (right). Figure 2 illustrates one cycle of the unbinding force assay. It includes an approaching stage where the probe bead and the T-cell are brought into contact. In the next stage, the touch of the two subjects is controlled with a given contact time so that a receptor-ligand bond might occur. In the last stage, the probe bead and the T-cell are retracted at a constant rate until they go back to the unbinding position that indicates the bond failure. The y-axis in Figure 2 represents the applied force in the foregoing process. The unbinding force is measured by the force difference observed at the point of bond failure.

Two interesting questions are raised in analyzing the repeated unbinding force tests where the unbinding force assay (i.e., approaching, contact and retraction) is performed repeatedly for each pair of experimental units, including a T-cell and a probe bead attached to a red blood cell. Such repeated assays are conducted for different pairs of units as replicates. The objective of the experiments is to study the dependence of the repeated unbinding force measurements because it was discovered recently that cells appear to have the ability to "remember" the previous

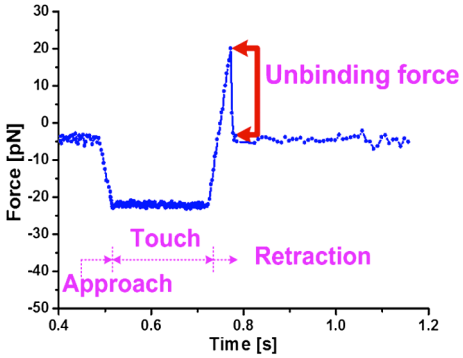

FIG. 2. *One cycle of the unbinding force experiment.*

adhesion events. Zarnitsyna et al. (2007), Hung et al. (2008) and Huang et al. (2010) demonstrated that in some biological systems the occurrence of binding in the immediate past assay could either increase or decrease the likelihood for the next assay to result in a binding. Such memory effects can affect not only through the binding frequency but also the unbinding force. Hence, the first question is how to model the memory effect on the repeated unbinding force assays. Apart from this, different receptor-ligand bonds can have a different order of the memory due to their string strength difference. Specifying the order of the memory for receptor-ligand bonds is important because it can be used to classify the bonds into groups for further biological study. Therefore, the other question is how to identify the order of the memory.

To answer the foregoing questions, a naive approach is to study the memory on the unbinding force by a time series model. However, the standard time series models cannot be applied directly. The reason is as follows. Due to the inherent stochastic nature of single molecular interaction, any particular assay has two random outcomes, either a receptor-ligand bond occurs or not. An unbinding force is representative and the resulting memory effects are considered only if the corresponding assay is associated with the occurrence of a bond. Theoretically, a distribution function might be used to capture the chance of a bond formation with respect to unbinding force. However, the related studies are mainly developed based on the independent assumption on the repeated adhesion experiments [Marshall et al. (2005)]. Being the first attempt to study the memory, we assume that the occurrence of a bond is determined by having the unbinding force above some threshold, which can be interpreted as the average unbinding force for bond dissociation. That is, if a bond occurs during the contact, the unbinding force would be larger than some threshold. The threshold, however, is unknown and has to be estimated from the data because of the detection limits and measurement errors. For example, Figure 3 is an example of the experiments with 20 repeated unbinding force assays generated from Hung et al. (2008). For each cycle of the assay, unbinding forces can be easily measured as described in Figure 2. A threshold has

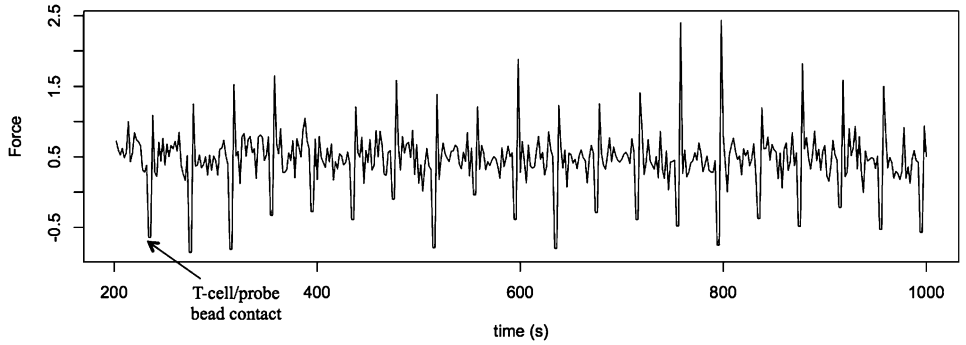

FIG. 3. *The measurements from the repeated unbinding force assays.*

to be determined so that time series models can be applied to those forces that are above the threshold. Failing to include such a threshold term can lead to a systematic bias in the successive adhesion assays. Because of the unknown threshold, conventional time series modeling techniques cannot be used. Furthermore, to identify the order of the memory, a new order selection approach that takes into account the foregoing features is called for.

A new time series model is proposed in this article to study the memory effect on the repeated unbinding force assays. It is a multiple nonlinear time series model with an unknown threshold parameter. Even though there are numerous studies on nonlinear time series modeling [Tong and Lim (1980), Tsay (1989), Fan and Yao (2003)], most of them are developed based on a single series of observations and focus on the situation where nonlinearity is determined by a particular variable. For example, the threshold autoregressive model [Tong (1983, 2007)] is constructed for a single series of observations with a delay parameter indicating the variable where the threshold is applied. The proposed nonlinear model is different from the existing nonlinear time series models in that there is no specific delay parameter involved. Instead, the threshold is applied to all the historical observations. Moreover, there is a hierarchical structure imposed upon the nonlinear model that makes the model more interpretable. Besides, this model handles multiple time series by incorporating random effects to take into account the heterogeneity among experimental units.

Identifying the order of the memory is equivalent to specifying the correct order of the proposed time series model. This is different from standard order selection problems because there are two types of overfitting associated with the proposed nonlinear time series model. Thus, a double penalized approach is developed and a global optimization algorithm using mixed integer programming (MIP) is introduced to implement this approach. The order selection consistency and asymptotic properties for the proposed method are discussed. The discontinuity of the conditional mean function of the new model results in nonstandard asymptotics for the estimators.

Although the methodology is motivated by the analysis of biomechanical experiments, it can be applied to a wide variety of studies, such as longitudinal data analysis [Diggle et al. (2002)], econometrics and influenza modeling. For example, in influenza modeling [Hyman and LaForce (2003)], the proposed method can be applied to model the spread of a disease, such as SARS. Because an epidemic threshold is used to indicate the take off and die out of an epidemic, the spread of the disease is of interest only when the threshold is reached, such as the infected population exceeding some amount. These thresholds are often unknown and estimated from the data. Therefore, the proposed model can be desirable for these studies.

The remainder of the paper is organized as follows. In Section 2 the nonlinear time series model is introduced. The estimation and order selection procedures

with a global optimization algorithm are introduced. In Section 3 the order selection consistency and some asymptotic properties of this model are discussed. The performance of the new model and the order selection procedure is demonstrated via simulations in Section 4. The proposed model is applied to an unbinding force assay in Section 5. Summary and concluding remarks are given in Section 6.

2. New class of nonlinear time series models.

2.1. *Modeling*. A new multiple nonlinear time series model is introduced in this section. Assume y_{it} represents the unbinding force observed from the ith subject at time t, where $i = 1, \ldots, n$, $t = 1, \ldots, m$ and the sample size $N = mn$. Define τ as a threshold parameter. Having the unbinding force above τ indicates that the corresponding contact results in a receptor-ligand bond and no bond otherwise. A random effect $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ is incorporated to take into account a variety of situations with the multiple time series, including subject heterogeneity, unobserved covariates and other forms of overdispersion. The random effects α_i's are assumed to be mutually independent and normal distributed with mean 0 and variance σ^2 in this paper. The following model is proposed to quantify the memory effect on the unbinding forces that are associated with receptor-ligand bonds:

$$
\begin{cases}
y_{it} = \alpha_i + \beta_0 + \varepsilon_{it}, & \text{if } y_{i,t-1} \le \tau, \\
y_{it} = \alpha_i + \beta_0 + \beta_1 y_{i,t-1} + \varepsilon_{it}, & \text{if } y_{i,t-1} > \tau,\ y_{i,t-2} \le \tau, \\
y_{it} = \alpha_i + \beta_0 + \beta_1 y_{i,t-1} & \\
\qquad + \beta_2 y_{i,t-2} + \varepsilon_{it}, & \text{if } y_{i,t-1} > \tau,\ y_{i,t-2} > \tau,\ y_{i,t-3} \le \tau, \\
\quad \vdots & \quad \vdots \\
y_{it} = \alpha_i + \beta_0 + \beta_1 y_{i,t-1} & \\
\qquad + \cdots + \beta_k y_{i,t-k} + \varepsilon_{it}, & \text{if } y_{i,t-1} > \tau, \ldots, y_{i,t-k} > \tau,
\end{cases}
\tag{1}
$$

where β_i's are the fixed effects and the error terms ε_{it} are independent with distribution $N(0, \sigma_\varepsilon^2)$.

The first equation in (1) corresponds to the situation where no receptor-ligand bond occurs in the previous test (i.e., $y_{i,t-1} \le \tau$). It amounts to modeling the unbinding forces in a sequence of independent adhesion tests. Let the mean unbinding force be β_0. The estimated value for β_0 is the average unbinding force in independent adhesion assays and can change with different settings of the experimental variables, such as different contact durations. Extensions can be easily achieved by incorporating these experimental variables into the model. The second equation in (1) describes the unbinding force when a receptor-ligand bond occurs in the previous test (i.e., $y_{i,t-1} > \tau$) but no bond in $y_{i,t-2}$ (i.e., $y_{i,t-2} \le \tau$). In this situation, a memory could be carried over from the previous observations. Thus, a first-order autoregressive model is considered. This autoregressive modeling continues to the previous k assays. Similar interpretation can be given to the rest of the model. The

value k represents the upper bound of the memory order; detailed discussions on identifying the order of the memory are given in Section 2.2.

The above model can be written in a concise form as follows:

$$\begin{aligned} y_{it} &= \mathbf{z}_i'\boldsymbol{\alpha} + \beta_0 + \beta_1 y_{i,t-1} I[y_{i,t-1} > \tau] + \beta_2 y_{i,t-2} I[y_{i,t-1} > \tau, y_{i,t-2} > \tau] \\ &\quad + \cdots + \beta_k y_{i,t-k} I[y_{i,t-1} > \tau, \ldots, y_{i,t-k} > \tau] + \varepsilon_{it} \\ &= g(\boldsymbol{\beta}, \tau, \sigma^2 \mid H_{it}) + \varepsilon_{it}, \end{aligned} \tag{2}$$

where $I(y_{i,t-1} > \tau)$ is an indicator function which takes value one if $y_{i,t-1} > \tau$ and zero otherwise. The fixed effects are denoted by $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$, the information from previous observations are included in $H_{it} = (1, y_{i,t-1}, \ldots, y_{i,t-k})$, and $\mathbf{z}_i = \{z_{i,1}, \ldots, z_{i,n}\}'$ is the design matrix for the random effects $\boldsymbol{\alpha}$ such that $\mathbf{z}_i'\boldsymbol{\alpha} = \alpha_i$. Since the proposed model is not limited to the analysis of unbinding force assay, the random intercept alone may not be sufficient to capture the variation exhibited in other applications. Hence, we use a general random effect structure hereafter. We call this new nonlinear time series model the multiple threshold autoregressive (MUTARE) model.

The MUTARE model is very general and includes an interesting special case with a single series of observations. Assuming that the time series observations are y_t, $t = 1, \ldots, m$, the special case of the MUTARE model can be written as

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} I[y_{t-1} > \tau] + \cdots \\ &\quad + \beta_k y_{t-k} I[y_{i,t-1} > \tau, \ldots, y_{t-k} > \tau] + \varepsilon_t. \end{aligned} \tag{3}$$

This is different from the conventional nonlinear time series models. The closest model in the literature is the threshold autoregressive models introduced by Tong (1983, 1990). There are various extensions of the threshold autoregressive models [Samia, Chan and Stenseth (2007)] and the nonlinearity therein is determined by a particular variable with which the threshold parameter is defined. The MUTARE model, however, has the threshold applied to all the historical observations. Furthermore, different from the threshold autoregressive model where piecewise linear submodels are fitted separately, a hierarchical structure is imposed upon the submodels in MUTARE as illustrated in (1), which makes the model easier to interpret.

2.2. *Estimation and order selection procedure.* A crucial step in this study is to specify the order of the memory, denoted by k_0. This is an order selection problem but different from standard ones in that there are two types of overfitting. By maximizing the log likelihood function, the resulting model may overfit the data with some small values of nonzero β_j's (type I overfitting) and/or with a large estimated order (type II overfitting). This is not surprising given the same problem experienced in estimating parameters in finite mixture models [Chen and Khalili (2008)]. Therefore, we propose to penalize type I overfitting by a function

$P_{\lambda_1}(|\beta_k|)$ and penalize type II overfitting by the estimated order ($\max_j\{j:\beta_j \neq 0\}$). The reason to consider type II overfitting is because the MUTARE model has a hierarchical structure as shown in (1). Once the order of the model (i.e., $\max_j\{j:\beta_j \neq 0\}$) is determined, all the previous equations have to be considered. So a double penalized likelihood is defined as

$$\text{(4)}\quad \text{pl}(\boldsymbol{\beta},\sigma^2,\tau) = 2\log L(\boldsymbol{\beta},\sigma^2,\tau) - \sum_{j=1}^{k} P_{\lambda_1}(|\beta_j|) - \lambda_2 \max_j\{j:\beta_j \neq 0\},$$

where L is the likelihood function. By maximizing (4), the solutions, $\hat{\boldsymbol{\beta}}$ and $\max_j\{j:\hat{\beta}_j \neq 0\}$, are the estimated parameters and order of the memory.

To prevent the first type of overfitting, there are different penalty functions discussed in the literature [Donoho and Johnstone (1994), Tibshirani (1996, 1997), Fan and Li (2001)]. Here we focus on the adaptive Lasso [Zou (2006)] where $P_{\lambda_1}(|\beta_j|) = \lambda_1 \nu_j |\beta_j|$ and $\nu_1,\ldots,\nu_k$ are known weights. The specification of ν_j can be fairly flexible and more discussions can be found in Zou (2006). We consider a weight vector suggested in Zou (2006) with $\hat{\nu}_j = |\hat{\beta}_j|^{-\rho}$, where $\rho > 0$ and $\hat{\beta}_j$ is a root-n-consistent estimator of β_j. In Hung (2011), it is shown that the MLE of $\boldsymbol{\beta}$ is root-n-consistent under model (2), therefore it can be applied.

By the following proposition, we can have a closer look at how the double penalized approach works. The proof is straightforward and is omitted.

PROPOSITION 1. *The penalized likelihood function in* (4) *is equivalent to*

$$\begin{aligned}\text{(5)}\quad \text{pl}(\boldsymbol{\beta},\sigma^2,\tau) &= 2\log L(\boldsymbol{\beta},\sigma^2,\tau) - \sum_{j=1}^{k} P_{\lambda_1}(|\beta_j|) - \lambda_2 \sum_{j=1}^{k} I(\beta_j \neq 0) \\ &\quad - \lambda_2 \sum_{j=1}^{k} I(\beta_j = 0, \textit{at least one } \beta_{j+p} \neq 0,\ p = 1,\ldots,k-j).\end{aligned}$$

Equation (5) connects the penalty for type II overfitting with the L_0 penalty, which directly controls the number of nonzero coefficients in the model. Therefore, the double penalized approach is closely related to a combination of L_0 and L_1 penalties, which is carefully studied by Liu and Wu (2007) and found to deliver better variable selection than the L_1 penalty while yielding a more stable model than the L_0 penalty.

2.3. *Mixed integer programming.* In this section a global optimization algorithm is introduced using the idea of MIP. MIP is an active research area in operations research with many applications. The objective here is to solve β_j's by maximizing the double penalized likelihood function (4). It is achieved by the following proposition.

PROPOSITION 2. *The penalized likelihood function in* (4) *is equivalent to*

$$\begin{aligned}\mathrm{pl}(\boldsymbol{\beta}, \sigma^2, \tau) = 2\log L(\boldsymbol{\beta}, \sigma^2, \tau) - \sum_{j=1}^{k} P_{\lambda_1}(|\beta_j|) \\ - \lambda_2 \sum_{j=1}^{k} \bigl(1 - I(\beta_j = \cdots = \beta_k = 0)\bigr).\end{aligned} \tag{6}$$

As discussed in Proposition 2, this problem is equivalent to the maximization of (6). Substitute variable β_j by two nonnegative variables β_j^+ and β_j^- with $\beta_j = \beta_j^+ - \beta_j^-$. Then, we have $|\beta_j| = \beta_j^+ + \beta_j^-$, and the maximization problem in (6) can be converted into a MIP problem with maximization of

$$2\log L(\boldsymbol{\beta}^+ - \boldsymbol{\beta}^-, \sigma^2, \tau) - \sum_{j=1}^{k} P_{\lambda_1}(\beta_j^+ + \beta_j^-) - \lambda_2 \sum_{j=1}^{k} z_j,$$

subject to

$$\begin{aligned}\beta_1^+ + \beta_1^- + \beta_2^+ + \beta_2^- + \cdots + \beta_k^+ + \beta_k^- &\le M z_1, \\ \beta_2^+ + \beta_2^- + \cdots + \beta_k^+ + \beta_k^- &\le M z_2, \\ &\vdots \\ \beta_k^+ + \beta_k^- &\le M z_k, \\ \beta_j^+, \beta_j^- &\ge 0, \qquad j = 1, \ldots, k, \\ z_j &\in \{0, 1\},\end{aligned}$$

where M is a very large constant and we can choose it to be the smallest upper bound of $\sum_j |\beta_j|$ if the prior knowledge is available. In the simulations, we apply the setting $M = 50$ and it works reasonably well in practice. In general, M can be even larger (e.g., $M = 1000$) for those problems with large k. Note that since $\beta_j^+ + \beta_j^-$ are to be minimized, β_j^+ and β_j^- would not be both positive in the optimal solution.

To solve the foregoing MIP problem, there are numerous methods such as the most popular branch-and-bound algorithm. More details about algorithms and the related issues can be found in Nemhauser and Wolsey (1999). The examples we considered in this article are solved by the C language with a GLPK package (available at http://www.gnu.org/software/glpk). Some other commercial optimization software such as CPLEX is also available to solve such a problem. The complexity of MIP can be considerably affected by introducing too many integer variables (i.e., z_j's), but it is in general not a critical concern. This is because the number of integer variables incorporated increases with the order k, and it is usually in a

manageable size in this application. For other applications with a large value of k, one can obtain a reasonably good solution (not necessarily optimal) by setting a restriction on the computing time to achieve efficiency.

Next we discuss the choice of the tuning parameters, λ_1, λ_2 and ρ. There are different approaches available in the literature for selecting tuning parameters [Stone (1974), Craven and Wahba (1979), Fan and Gijbels (1996)]. Burman, Chow and Nolan (1994) introduced the h-block cross-validation for dependent data. The idea is to modify the leave-one-out cross-validation and reduce the training set by removing the h observations preceding and following the observation in each test set. Such blocking allows near independence between the training and test set. This approach is further improved by Racine (2000) to achieve asymptotic consistency. That is, instead of leave-one-out, the size of the validation set is increased to n_v. So the training set has size n_c and $n_v + n_c + 2h = m - k$. In this paper, we implement Racine's approach with the setting $h = (m-k)/4$ and n_c being the integer part of $m^{0.5}$, which appears to work well in a wide range of situations in practice [Racine (2000)].

The rest of the parameters can be estimated by the standard maximum likelihood approach. Denote the observation by vector $Y = (\mathbf{y}_1, \ldots, \mathbf{y}_n)'$, where the observations for subject i are denoted by $\mathbf{y}_i = (y_{i1}, \ldots, y_{im})'$. Given the historical information H_{it} and the random effects, the associated likelihood as a function of the fixed effects $\boldsymbol{\beta}$ and the threshold parameter can be written as

$$L(\boldsymbol{\beta}, \tau \mid \boldsymbol{\alpha}) = \prod_{i=1}^{n} \prod_{t=1}^{m} l(y_{it} \mid \boldsymbol{\alpha}, H_{it}),$$

where $l(\cdot)$ is the likelihood for each observation y_{it} given $\boldsymbol{\alpha}$ and the corresponding historical information. Considering the normality of the error ε and random effects $\boldsymbol{\alpha}$, the joint log likelihood can be easily derived as

$$\begin{aligned} &2\log L(\boldsymbol{\beta}, \sigma^2, \tau) \\ &\quad = -\log|\mathbf{W}| - \bigl(Y - g(\boldsymbol{\beta}, \tau, \sigma^2 \mid H)\bigr)' \mathbf{W}^{-1} \bigl(Y - g(\boldsymbol{\beta}, \tau, \sigma^2 \mid H)\bigr), \end{aligned} \tag{7}$$

where $g(\boldsymbol{\beta}, \tau, \sigma^2 \mid H)$ is the mean vector, $H = (H_1', \ldots, H_n')'$, $H_i = (H_{i1}', \ldots, H_{im}')'$, Z is the design matrix for the random effects with rows $\mathbf{z}_i'$, and $\mathbf{W} = \sigma_\varepsilon^2 \mathbf{I} + \sigma^2 ZZ'$. Note that σ_ε^2 is assumed to be known for notational convenience. The variance component σ^2 is estimated by maximizing the original likelihood throughout the paper and the estimator can be further improved by the restricted maximum likelihood [McCulloch and Searle (2008)]. Such a version of the variance components developed for the linear mixed model can be easily extended to the multiple threshold autoregressive model so that the estimated variance component is invariant to the values of the fixed effects and the degrees of freedom for the fixed effects can be taken into account implicitly.

3. Large sample properties. The consistency of the order selection procedure and the asymptotic properties of the resulting estimators in the MUTARE model are studied in this section. The parameter space of $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \tau, \sigma^2)$ is denoted by Ω and the true parameter is denoted by $\boldsymbol{\gamma}_0 = (\boldsymbol{\beta}_0, \tau_0, \sigma_0^2)$. Assumptions and proofs are deferred to the Appendix.

Lemma 1 shows that the maximum penalized likelihood estimator for the MUTARE model is stochastically bounded.

LEMMA 1. *Under Assumptions A1–A4, there exists a $\nu > 0$ such that, for m and m sufficiently large, the maximum penalized likelihood estimator of the parameter $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \tau, \sigma^2)$ lies in a compact space $\Omega_1 = \{\boldsymbol{\gamma} \in \Omega : |\boldsymbol{\gamma} - \boldsymbol{\gamma}_0| \leq \nu\}$ almost surely.*

The convergence rate of the estimated threshold parameter is derived in Theorem 1 for the MUTARE model. This result is analogous to Chan (1993) for the least squares estimator of the threshold autoregressive model. Not surprisingly, the estimated threshold parameter in the MATARE model has a fast convergence rate $[O(1/N)]$ which is similar to that in the threshold autoregressive model, and the fast convergence rate is also due to the discontinuity of the conditional mean function [Chan (1993), Hansen (2000)]. Note that, as a special case, the estimated threshold parameter in (3) obtains a convergence rate $O(1/m)$.

THEOREM 1. *Under Assumptions A1–A4, the maximum likelihood estimator of the threshold has the property that $\hat{\tau} = \tau_0 + O_p(1/N)$, based on the MUTARE model.*

Define $\tilde{H} = (\tilde{H}_1', \ldots, \tilde{H}_n')$, $\tilde{H}_i = (\tilde{H}_{i1}', \ldots, \tilde{H}_{im}')$, and

$$\tilde{H}_{it} = \big(1, y_{i,t-1} I(y_{i,t-1} > \tau), \ldots, y_{i,t-k} I(y_{i,t-1} > \tau, \ldots, y_{i,t-k} > \tau)\big).$$

Let $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{(1)}', \boldsymbol{\beta}_{(2)}')'$, where $\boldsymbol{\beta}_{(1)}'$ is a vector with all the nonzero parameters and the rest of the parameters are denoted by $\boldsymbol{\beta}_{(2)}'$. Furthermore, assume $\frac{\tilde{H}' \mathbf{W}^{-1} \tilde{H}}{N} \to \Lambda$, where Λ is positive definite and can be written as

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}$$

according to $\boldsymbol{\beta}_{(1)}'$ and $\boldsymbol{\beta}_{(2)}'$.

In the next theorem, we show that the penalized likelihood estimator of $\boldsymbol{\beta}$ enjoys the oracle properties [Fan and Li (2001)], which indicates the consistency in variable selection and the asymptotic normality. This result also implies the order selection consistency of the proposed order selection procedure.

THEOREM 2. *Suppose that* $\lambda_1/\sqrt{N} \to 0$ *and* $\lambda_1 N^{(\rho-1)/2} \to \infty$. *Under Assumptions* A1–A4, *for any* η, $0 < \eta < \infty$, *the maximum penalized likelihood estimator of* $\boldsymbol{\beta}$ *in the MUTARE model satisfies the following two properties as* $n \to \infty$ *and* $m \to \infty$:

(i) $\hat{\boldsymbol{\beta}}_{(2)} = 0$ *with probability* 1,

(ii) $\sup_{|\hat{\tau}-\tau_0|\le\eta/N, |\hat{\sigma}^2-\sigma_0^2|<\eta/\sqrt{N}} \sqrt{N}(\hat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1)}) \to^d N(0, \Lambda_{11}^{-1})$.

Apart from the fixed effects, asymptotic distributions of the estimated variance components deserve more investigation. Numerous works have appeared in the literature addressing methods of variance component estimation in linear models and the associated asymptotic properties [Jiang (1996), McCulloch and Searle (2008)]. Strong consistency of the estimated variance component in nonlinear mixed effect models [Nie (2006)] is expected to be extended to the MUTARE model. A rigorous theoretical proof along the lines of Nie (2006) is not attempted here, and remains the subject of ongoing theoretical work. However, it is briefly noted that the asymptotic conditions, such as Assumptions A3 and A4, required for the results here are indeed met by the requirement in Nie (2006). The requirement of $n \to \infty$ for the main theorems is based upon the asymptotic study in Nie (2006) and it is expected to be further relaxed by the techniques developed in Jiang (1996).

4. Finite-sample performance and empirical application. In this section simulations are conducted to examine the finite-sample performance of the proposed models. Two examples are considered. The first example demonstrates the performance of the estimators in the MUTARE model and the second example compares the double penalized order selection procedure with a standard approach.

4.1. *Example* 1. Consider the following MUTARE model with $k = 2$:

$$y_{it} = \alpha_i + \beta_0 + \beta_1 y_{i,t-1} I[y_{i,t-1} > \tau] + \beta_2 y_{i,t-2} I[y_{i,t-1} > \tau, y_{i,t-2} > \tau] + \varepsilon_{it}.$$

The coefficients of this model are fixed at $\boldsymbol{\gamma}_0 = (\boldsymbol{\beta}_0, 0.1, 0.5)$, where the fixed effects are $\boldsymbol{\beta}_0 = (0, 0.5, 0.4)$. The random error ε_{it} is generated from a normal distribution with mean 0 and variance 0.5. The sample size combinations used are $(m = 30, n = 10)$, $(m = 40, n = 15)$, and $(m = 60, n = 25)$. For each combination, the simulations are conducted based on 1000 replicates. In this example, tuning parameters are determined by minimizing the mean squared prediction error of new generated testing data with the same size and then fixed for all the replicates.

The simulation results are reported in Table 1. For each sample size combination, the sample means and standard deviations of the estimates are listed. The empirical coverage probabilities of the fixed effects, denoted by "CP," are listed in the last row of each setting. They are calculated based on the 90% confidence intervals of the corresponding regression parameters. As shown in the table, the sample mean of the estimates becomes closer to the true value and the associated standard

TABLE 1
Summary of simulation results in example 1

	τ	β_0	β_1	β_2	σ^2
		$m = 30, n = 10$			
Mean	0.114	0.038	0.491	0.390	0.385
sd	0.029	0.174	0.083	0.080	0.188
CP		0.906	0.859	0.866	
		$m = 40, n = 15$			
Mean	0.111	0.035	0.484	0.393	0.431
sd	0.028	0.155	0.058	0.047	0.140
CP		0.915	0.868	0.889	
		$m = 60, n = 25$			
Mean	0.105	0.036	0.501	0.398	0.477
sd	0.019	0.123	0.041	0.032	0.090
CP		0.918	0.878	0.898	
True	0.1	0	0.5	0.4	0.5

deviation becomes smaller as the sample size increases. These results confirm the asymptotic consistency discussed in Section 3. Moreover, when the sample size increases, the empirical coverage probabilities for the fixed effects are closer to the nominal coverage probabilities.

To assess the asymptotic normality, normal Q–Q plots are reported in Figure 4. It is plotted based on the three estimated fixed effects, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$, with the

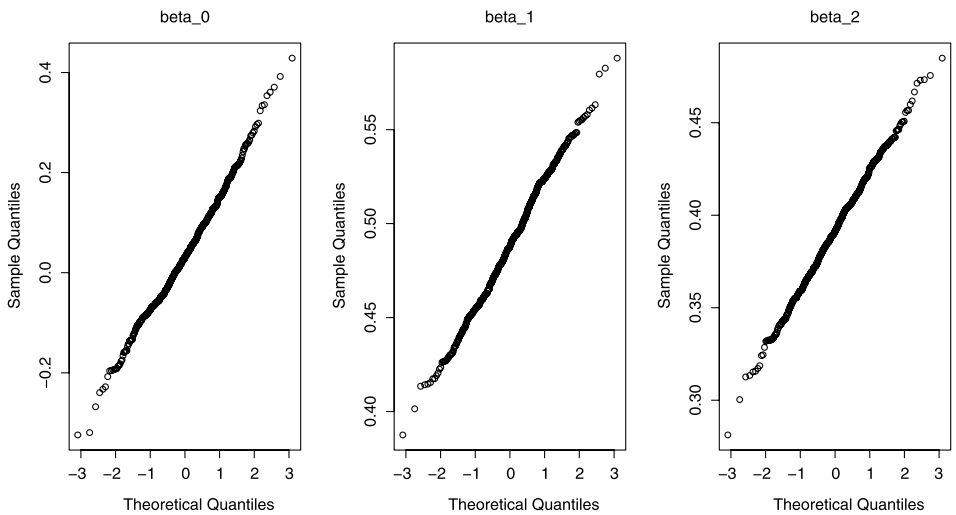

FIG. 4. *Normal Q–Q plots in example* 1.

TABLE 2
Parameter values in example 2

Model	β_1	β_2	β_3	β_4	β_5
1	0.4	0.4	0	0	0
2	0.5	0.3	0.1	0	0
3	0.3	0.2	0.1	0.05	0

sample size combination $m = 60$ and $n = 25$. In general, the data points being close to straight lines in the Q–Q plots confirms that the estimates are normally distributed.

4.2. *Example* 2. In this example we study the performance of the proposed order selection procedure. Since there is no existing approach available, we compare the double penalized approach with a naive Akaike information criterion [AIC; Akaike (1973)], which is suggested for order selection in the threshold autoregressive models [Tong (1980)], and the Bayesian information criterion [BIC; Schwarz (1978)]. Three different models following equation (3) are considered with parameters given in Table 2 and sample size 200. The threshold parameters are assumed to be 0.01 and the random errors are generated from a normal distribution with mean 0 and variance 0.1. The tuning parameters are determined as in example 1.

Table 3 shows the order selection performance of AIC, BIC and the double penalized approach. The column k_0 indicates the true order. For both methods, we report the percentage of times that the estimated order equals a number of values (i.e., 1 to 5) out of 1000 replicates. The numbers with boldface indicate the most selected orders. For model 1, all the three methods select the right order with their highest frequency. The double penalized approach and BIC perform equally well in this model and both of them perform better than AIC. For example, the double penalized approach has a 30% $[= (0.751 - 0.580)/0.580]$ higher chance to select the right order compared with AIC. For models 2 and 3, both AIC and BIC tend to underestimate the order and the double penalized approach selects the correct order with probability higher than 65%. These results indicate that the double penalized approach outperforms the other two methods in terms of order selection. The computational efficiency of the double penalized approach is reasonably close to AIC and BIC in the simulation. The average computing times are 4.38 seconds for AIC, 4.45 seconds for BIC and 4.92 seconds for the double penalized approach.

5. Application in unbinding force experiments. In this section we revisit the repeated unbinding force experiments and apply the proposed method to study the memory effect on such repeated assays. There are 15 pairs of experimental subjects and each pair includes a T-cell and a probe bead attached to a red blood cell

TABLE 3
Simulation results in example 2

		AIC				
Model	k_0	**1**	**2**	**3**	**4**	**5**
1	2	0.178	**0.580**	0.193	0.014	0.020
2	3	0.142	**0.574**	0.150	0.101	0.031
3	4	**0.522**	0.325	0.111	0.042	0.000
		BIC				
Model	k_0	**1**	**2**	**3**	**4**	**5**
1	2	0.001	**0.749**	0.152	0.088	0.001
2	3	0.243	**0.536**	0.151	0.058	0.012
3	4	**0.553**	0.322	0.110	0.015	0.000
		Double penalized				
Model	k_0	**1**	**2**	**3**	**4**	**5**
1	2	0.103	**0.751**	0.091	0.050	0.004
2	3	0.000	0.053	**0.659**	0.167	0.121
3	4	0.023	0.081	0.248	**0.645**	0.003

as described in Figure 1. For each cell adhesion cycle, a T-cell and a probe bead are brought into contact (i.e., touch) for 4 seconds and then retracted to the unbinding position (see Figure 2). Such a cycle is performed repeatedly on the same pair of experimental subjects for 50 times. Figure 5 is three randomly selected samples of the repeated unbinding forces from such experiments. For each sample, the forces are plotted based on observations in 1000 seconds with 50 repeated adhesion cycles completed.

The unbinding forces are collected according to the definition in Figure 2. Prior knowledge [Zarnitsyna et al. (2007), Hung et al. (2008)] indicates that a reasonable order of the memory in this process should be less than 5. Therefore, we first fit the MUTARE model with $k = 5$ and then the double penalized order selection procedure is applied. The order of the memory is identified as two and the memory effect on the repeated unbinding force experiments can be quantified by the MUTARE model as

$$\hat{y}_{it} = \alpha_i + 0.245 y_{i,t-1} I[y_{i,t-1} > \hat{\tau}] + 0.11 y_{i,t-2} I[y_{i,t-1} > \hat{\tau}, y_{i,t-2} > \hat{\tau}],$$

where $i = 1, \ldots, 15$, $t = 1, \ldots, 50$, the random effect α_i follows normal distribution with mean -0.072 and variance 0.389. The estimated order of the memory in this experiment is consistent with that in Hung et al. (2008) with a similar setting but different measurements. Such consistency provides important evidence of

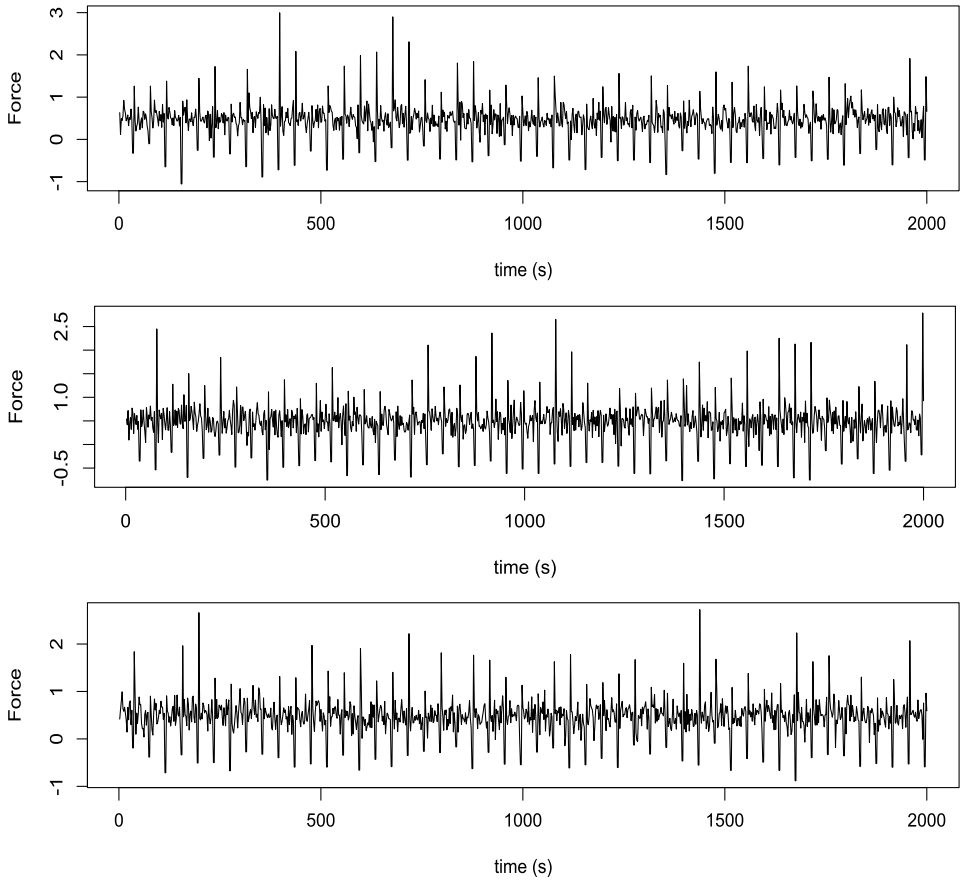

FIG. 5. *The measurements from the repeated unbinding force assays.*

a unified underlying kinetic mechanism in the adhesion process. The estimated threshold, $\hat{\tau} = 0.089$, indicates that an adhesion leads to a bond only if the unbinding force is larger than $0.089pN$. Based on this result, the occurrence of a bond, although unobservable, can be easily studied by measuring the corresponding unbinding forces. Since random effects are considered, the fitted model can be used to make inference beyond the 15 pairs of experimental subjects.

6. Summary and concluding remarks. Despite numerous results available in modeling nonlinear time series, their applications are limited. For example, they are mainly constructed for a single series of observations and focus on the case where the nonlinearity is determined based on one variable. Furthermore, there is no order selection procedure available with theoretical justification for such models. Motivated by the analysis of the repeated unbinding force experiments, a new

nonlinear time series model, MUTARE, and a double penalized order selection procedure are introduced.

The proposed model handles multiple time series by incorporating random effects to borrow strength across different subjects. Thus, inference and predictions can be made beyond the experimental units in the study. Moreover, the proposed methodology provides a new nonlinear time series model that is easy to interpret and captures the autoregressive behavior of the observations above some unknown threshold. The double penalized procedure can be used to efficiently identify the order and can be easily implemented by a global optimization algorithm using mixed integer programming. The selection consistency and asymptotic normality of the estimators are derived. Apart from the asymptotic results, the finite-sample performance is examined via simulations.

As an application, the MUTARE model is illustrated by modeling the memory effect on the repeated unbinding force assays. The fitted model provides a better understanding of how force regulates receptor-ligand interactions. This work is one of the first few studies considering memory effects in the cell adhesion experiments. More studies are needed to construct a rigorous and interpretable biological model. An ongoing project includes theoretical development for the estimated threshold, relaxation of the constant threshold assumption, and taking into account important process variables, such as contact duration, into the model.

APPENDIX A: ASSUMPTIONS

ASSUMPTION A1. The process y_{it} is stationary, ergodic and has finite second moments.

ASSUMPTION A2. The autoregressive function is discontinuous, that is, there exists a $H^* = (1, y^*_{t-1}, \ldots, y^*_{t-k})$ such that $H^*(A_s - A_t) \neq 0$ and $y_{t-1} = \cdots = y_{t-j} = \tau$, where $A_1 = (\beta_0, 0, \ldots, 0)', \ldots, A_{k-1} = (\beta_1, \ldots, \beta_k)'$, $(s,t) \in (1, \ldots, k-1)$, and $j = 1, \ldots, k$.

ASSUMPTION A3. There exists a $M_1 > 0$ such that $E[\mathrm{tr}(\mathbf{W}_i^{-1} Z_i Z_i' \times \mathbf{W}_i^{-1} Z_i Z_i')]^2 \leq M_1$, and $E[\mathrm{tr}(\mathbf{W}_i^{-1} Z_i Z_i') - (\mathbf{y}_i - g(H_i, \boldsymbol{\beta}_0, \tau_0, \sigma^2))' \mathbf{W}_i^{-1} Z_i Z_i' \times \mathbf{W}_i^{-1} (\mathbf{y}_i - g(H_i, \boldsymbol{\beta}_0, \tau_0, \sigma^2)]^2 \leq M_1$ for all i, where $\mathbf{W}_i$ and Z_i are the matrices of the covariance and random effects for the ith subject.

ASSUMPTION A4. $\liminf_{n\to\infty} \lambda_n = \lambda > 0$, where λ_n is the smallest eigenvalue of $-\frac{1}{n} \sum_i E[\mathrm{tr}(\mathbf{W}_i^{-1} Z_i Z_i' \mathbf{W}_i^{-1} Z_i Z_i')]$.

Assumptions A1 and A2 are necessary for the strong consistency of the fixed effect and threshold parameter estimators. Assumptions A3 and A4 are used for

the strong consistency of the variance components. More discussions can be found in Nie (2006).

APPENDIX B: PROOF OF LEMMA 1

The proof relies on verifying the following two claims.

CLAIM 1. *There exists a* $M_2 > 0$ *such that, for* m *and* n *sufficiently large, the maximum likelihood estimator of* $\boldsymbol{\gamma}$ *lies in* $\Omega_2 = \{\boldsymbol{\gamma} \in \Omega : |\beta_1 - \beta_{1,0}| \le M_2, \ldots, |\beta_k - \beta_{k,0}| \le M_2, |\sigma^2 - \sigma_0^2| \le M_2\}$ *almost surely.*

VERIFICATION OF CLAIM 1. Recall $\boldsymbol{\gamma}_0 = (\boldsymbol{\beta}_0, \tau, \sigma_0^2)$ and define $\boldsymbol{\beta}_0 = (\beta_{0,0}, \ldots, \beta_{k,0})'$. To prove Claim 1, it suffices to show that for m and n sufficiently large and uniformly for $\boldsymbol{\gamma}$ not belonging to Ω_2, we have $(mn)^{-1}(\mathrm{pl}(\boldsymbol{\gamma}) - \mathrm{pl}(\boldsymbol{\gamma}_0)) < 0$ almost surely:

$$\begin{aligned}\frac{\mathrm{pl}(\boldsymbol{\gamma}) - \mathrm{pl}(\boldsymbol{\gamma}_0)}{mn} &= \frac{\mathrm{pl}(\boldsymbol{\beta}, \tau, \sigma^2) - \mathrm{pl}(\boldsymbol{\beta}_0, \tau_0, \sigma^2)}{mn} \\ &\quad + \frac{\mathrm{pl}(\boldsymbol{\beta}_0, \tau_0, \sigma^2) - \mathrm{pl}(\boldsymbol{\beta}_0, \tau_0, \sigma_0^2)}{mn}.\end{aligned} \tag{8}$$

We first examine the first part on the right-hand side of (8). Assuming that the variance component is consistent along the lines of Nie (2006), the study of the first part can be transformed into the study of $Y^* = \mathbf{W}^{-1/2}(Y - Z\boldsymbol{\alpha})$, which is used in the derivation for both $\mathrm{pl}(\boldsymbol{\beta}, \tau, \sigma^2)$ and $\mathrm{pl}(\boldsymbol{\beta}_0, \tau_0, \sigma^2)$. We have

$$\begin{aligned}&\mathrm{pl}(\boldsymbol{\beta}, \tau, \sigma^2) - \mathrm{pl}(\boldsymbol{\beta}_0, \tau_0, \sigma^2) \\ &\quad = 2\log L(\boldsymbol{\beta}, \tau, \sigma^2) - 2\log L(\boldsymbol{\beta}_0, \tau_0, \sigma^2) \\ &\qquad + \sum_{j=1}^{k} [P_{\lambda_1}(|\beta_{j,0}|) - P_{\lambda_1}(|\beta_j|)] \\ &\qquad + \lambda_2 \Big[\max_j \{j : \beta_{j,0} \neq 0\} - \max_j \{j : \beta_j \neq 0\} \Big].\end{aligned}$$

First, up to an additive constant, we have

$$2\log(\boldsymbol{\beta}, \tau, \sigma^2) = -\sum_i \sum_t (y_{it}^* - g(\boldsymbol{\beta}, \tau, \sigma^2 \mid H_{it}))^2.$$

Due to the nonlinearity, the derivation for a general MUTARE model can be lengthy in nature. Therefore, we illustrate the detailed derivation by a smaller model and consider the case where $\tau > \tau_0$. The same argument can be easily applied and extended to the MUTARE model and the case $\tau \le \tau_0$ in general.

Consider a MUTARE model with $k = 2$:

$$
(9)\qquad \begin{cases} y_{it} = \alpha_i + \beta_0 + \varepsilon_{it}, & \text{if } y_{i,t-1} \le \tau, \\ y_{it} = \alpha_i + \beta_0 + \beta_1 y_{i,t-1} + \varepsilon_{it}, & \text{if } y_{i,t-1} > \tau,\ y_{i,t-2} \le \tau, \\ y_{it} = \alpha_i + \beta_0 + \beta_1 y_{i,t-1} + \beta_2 y_{i,t-2} + \varepsilon_{it}, & \text{if } y_{i,t-1} > \tau,\ y_{i,t-2} > \tau, \end{cases}
$$

the corresponding log likelihood function can be decomposed by

$$
\begin{aligned}
&2\log L(\boldsymbol{\beta}, \tau, \sigma^2) \\
&\quad = -\sum_i \sum_t (y_{it}^* - \beta_0)^2 I[y_{i,t-1} \le \tau_0] \\
&\qquad - \sum_i \sum_t (y_{it}^* - \beta_0)^2 I[\tau_0 < y_{i,t-1} \le \tau,\ y_{i,t-2} \le \tau_0] \\
&\qquad - \sum_i \sum_t (y_{it}^* - \beta_0)^2 I[\tau_0 < y_{i,t-1} \le \tau,\ y_{i,t-2} > \tau_0] \\
&\qquad - \sum_i \sum_t (y_{it}^* - \beta_0 - \beta_1 y_{i,t-1}^*)^2 I[y_{i,t-1} > \tau,\ y_{i,t-2} \le \tau_0] \\
&\qquad - \sum_i \sum_t (y_{it}^* - \beta_0 - \beta_1 y_{i,t-1}^*)^2 I[y_{i,t-1} > \tau,\ \tau_0 < y_{i,t-2} \le \tau] \\
&\qquad - \sum_i \sum_t (y_{it}^* - \beta_0 - \beta_1 y_{i,t-1}^* - \beta_2 y_{i,t-2}^*)^2 I[y_{i,t-1} > \tau,\ y_{i,t-2} > \tau] \\
&\quad = R_1(\boldsymbol{\beta}, \tau, \sigma^2) + \cdots + R_6(\boldsymbol{\beta}, \tau, \sigma^2).
\end{aligned} \tag{10}
$$

Defining $A_1 = (\beta_{0,0}, 0, 0)'$, $A_2 = (\beta_{0,0}, \beta_{1,0}, 0)'$, $A_3 = (\beta_{0,0}, \beta_{1,0}, \beta_{2,0})'$, $B_1 = (\beta_0, 0, 0)'$, $B_2 = (\beta_0, \beta_1, 0)'$, and $B_3 = (\beta_0, \beta_1, \beta_2)'$, we have

$$
\begin{aligned}
&R_4(\boldsymbol{\beta}, \tau, \sigma^2) - R_4(\boldsymbol{\beta}_0, \tau_0, \sigma^2) \\
&\quad = \sum_i \sum_t [-(y_{it}^* - \beta_0 - \beta_1 y_{i,t-1}^*)^2 + (y_{it}^* - \beta_{0,0} - \beta_{1,0} y_{i,t-1}^*)^2] \\
&\qquad\qquad \times I(y_{i,t-1} > \tau,\ y_{i,t-2} \le \tau_0) \\
&\quad = \sum_i \sum_t [-(y_{it}^* - H_{i,t-1}^* B_2)^2 + (y_{it}^* - H_{i,t-1}^* A_2)^2] \\
&\qquad\qquad \times I(y_{i,t-1} > \tau,\ y_{i,t-2} \le \tau_0) \\
&\quad = 2|B_2 - A_2| \sum_i \sum_t H_{i,t-1}^* \frac{(B_2 - A_2)}{|B_2 - A_2|} (y_{it}^* - H_{i,t-1}^* A_2) \\
&\qquad\qquad\qquad \times I(y_{i,t-1} > \tau,\ y_{i,t-2} \le \tau_0) \\
&\qquad - |B_2 - A_2|^2 \sum_i \sum_t \left(H_{i,t-1}^* \frac{(B_2 - A_2)}{|B_2 - A_2|} \right)^2 I(y_{i,t-1} > \tau,\ y_{i,t-2} \le \tau_0).
\end{aligned}
$$

Therefore, based on the uniform law of large numbers [Pollard (1984), page 8], we have

$$\begin{aligned}
&\frac{1}{mn}\bigl(\mathrm{pl}(\boldsymbol{\beta},\tau,\sigma^2)-\mathrm{pl}(\boldsymbol{\beta}_0,\tau_0,\sigma^2)\bigr)\\
&\quad\le 2(|B_1-A_1|+|B_1-A_2|+|B_1-A_3|\\
&\qquad+|B_2-A_2|+|B_2-A_3|+|B_3-A_3|)\varepsilon\\
&\qquad-(|B_1-A_1|^2+|B_1-A_2|^2+|B_1-A_3|^2\\
&\qquad\quad+|B_2-A_2|^2+|B_2-A_3|^2+|B_3-A_3|^2)(K-\varepsilon)\\
&\qquad+(mn)^{-1}\Biggl\{\sum_{j=1}^{k}[P_{\lambda_1}(|\beta_{j,0}|)-P_{\lambda_1}(|\beta_j|)]\\
&\qquad\qquad+\lambda_2\Bigl[\max_j\{j:\beta_{j,0}\neq 0\}-\max_j\{j:\beta_j\neq 0\}\Bigr]\Biggr\}\\
&\quad=2\varepsilon\Delta_1-\Delta_2(K-\varepsilon)+\Delta_3,
\end{aligned}$$

where

$$K=\inf_{\beta}\min_{i\le j}E\left(\left(H_{i,t-1}\frac{(B_i-A_j)}{|B_i-A_j|}\right)^2 I_{ij}\right)$$

and I_{ij} is the corresponding indicator function as listed in (10). Note that the uniform law of large numbers in Pollard [(1984), page 8] assumes that the data are independent and identically distributed. This assumption is relaxed to a stationary ergodic process by Samia and Chan (2011). Therefore, the uniform law of large numbers can be applied here. Based on the Cauchy–Schwarz inequality, we have $\Delta_1\le\sqrt{6\Delta_2}\le 6\Delta_2$ for sufficiently large M_2. For sufficiently large m and n, $\Delta_3<\varepsilon\Delta_2$. Thus, by selecting $\varepsilon<K/14$, it follows that $(mn)^{-1}(l(\boldsymbol{\beta},\tau,\sigma^2)-l(\boldsymbol{\beta}_0,\tau_0,\sigma^2))<0$.

For the second term on the right-hand side of (8), under Assumptions A3 and A4, the maximum likelihood estimator of the variance component almost surely converges based on the results in Nie (2006). Therefore, we have $(mn)^{-1}(l(\boldsymbol{\beta}_0,\tau_0,\sigma^2)-l(\boldsymbol{\beta}_0,\tau_0,\sigma_0^2))<0$ and Claim 1 follows.

CLAIM 2. *There exists a $M_3>0$ such that, for m and n sufficiently large, the maximum likelihood estimator of $\boldsymbol{\gamma}$ lies in $\Omega_3=\{\boldsymbol{\gamma}\in\Omega_2:|\tau-\tau_0|\le M_3\}$ almost surely.*

VERIFICATION OF CLAIM 2. Similar to Claim 1, it suffices to show that, for m and n sufficiently large, $(mn)^{-1}(\mathrm{pl}(\boldsymbol{\gamma})-\mathrm{pl}(\boldsymbol{\gamma}_0))<0$ for $\boldsymbol{\gamma}$ not belonging to Ω_3. We apply the same decomposition as in Lemma 1 and focus on the first part on the

right-hand side of (8). Applying the uniform law of large numbers and the same transformation as described in Claim 1, for m and n sufficiently large, it holds that

$$\begin{aligned}
&\frac{\mathrm{pl}(\boldsymbol{\beta},\tau,\sigma^2)-\mathrm{pl}(\boldsymbol{\beta}_0,\tau_0,\sigma^2)}{mn}\\
&\quad= mn^{-1}\Bigg\{2\log L(\boldsymbol{\beta},\tau,\sigma^2)\\
&\qquad\qquad -2\log L(\boldsymbol{\beta}_0,\tau_0,\sigma^2)\sum_{j=1}^{k}[P_{\lambda_1}(|\beta_{j,0}|)-P_{\lambda_1}(|\beta_j|)]\\
&\qquad\qquad +\lambda_2\Big[\max_j\{j:\beta_{j,0}\neq 0\}-\max_j\{j:\beta_j\neq 0\}\Big]\Bigg\}\\
&\quad\le E\big\{(-(y_{it}^*-H_{i,t-1}^*B_1)^2+(y_{it}-H_{i,t-1}A_1)^2)I[y_{i,t-1}\le\tau_0]\big\}\\
&\qquad+E\big\{(-(y_{it}^*-H_{i,t-1}^*B_1)^2+(y_{it}^*-H_{i,t-1}^*A_2)^2)\\
&\qquad\qquad\times I[\tau_0<y_{i,t-1}\le\tau,\,y_{i,t-2}\le\tau_0]\big\}\\
&\qquad+E\big\{(-(y_{it}^*-H_{i,t-1}^*B_1)^2+(y_{it}^*-H_{i,t-1}^*A_3)^2)\\
&\qquad\qquad\times I[\tau_0<y_{i,t-1}\le\tau,\,y_{i,t-2}>\tau_0]\big\}\\
&\qquad+E\big\{(-(y_{it}^*-H_{i,t-1}^*B_2)^2+(y_{it}^*-H_{i,t-1}^*A_2)^2)\\
&\qquad\qquad\times I[y_{i,t-1}>\tau,\,y_{i,t-2}\le\tau_0]\big\}\\
&\qquad+E\big\{(-(y_{it}^*-H_{i,t-1}^*B_2)^2+(y_{it}^*-H_{i,t-1}^*A_3)^2)\\
&\qquad\qquad\times I[y_{i,t-1}>\tau,\,\tau_0<y_{i,t-2}\le\tau]\big\}\\
&\qquad+E\big\{(-(y_{it}^*-H_{i,t-1}^*B_3)^2+(y_{it}^*-H_{i,t-1}^*A_3)^2)\\
&\qquad\qquad\times I[y_{i,t-1}>\tau,\,y_{i,t-2}>\tau]\big\}+\varepsilon.
\end{aligned}$$

Considering the situation where $\tau>\tau_0$, we have

$$\frac{l(\boldsymbol{\beta},\tau,\sigma^2)-l(\boldsymbol{\beta}_0,\tau_0,\sigma^2)}{mn}\le J+\varepsilon,$$

where

$$\begin{aligned}
J&=E\big\{(-(y_{it}^*-H_{i,t-1}^*B_1)^2+(y_{it}^*-H_{i,t-1}^*A_1)^2)I[y_{i,t-1}\le\tau_0]\big\}\\
&\quad+E\big\{(-(y_{it}^*-H_{i,t-1}^*B_1)^2+(y_{it}^*-H_{i,t-1}^*A_2)^2)\\
&\qquad\times I[\tau_0<y_{i,t-1}\le\tau,\,y_{i,t-2}\le\tau_0]\big\}\\
&\quad+E\big\{(-(y_{it}^*-H_{i,t-1}^*B_1)^2+(y_{it}^*-H_{i,t-1}^*A_3)^2)\\
&\qquad\times I[\tau_0<y_{i,t-1}\le\tau,\,y_{i,t-2}>\tau_0]\big\}.
\end{aligned}$$

When $\tau = \infty$, the model becomes a linear mixed model; therefore, by the dominated convergence theorem and a similar argument in Samia and Chan (2011), it holds almost surely that, for m and n sufficiently large and for any $M_3 > 0$, $(mn)^{-1}(l(\boldsymbol{\beta}, \tau, \sigma^2) - l(\boldsymbol{\beta}_0, \tau_0, \sigma^2)) < 0$ for $\tau \geq \tau_0 + M_3$. Similar derivation can be applied to the case $\tau < \tau_0$, thus the detail is omitted.

Following the same argument for Claim 1, the second part on the right-hand side of (8) is smaller than 0 with Assumptions A3 and A4. Therefore, Lemma 1 holds.

APPENDIX C: PROOF OF THEOREM 1

Without loss of generality, the parameter space can be restricted to $\Omega_\delta = \{\boldsymbol{\gamma} \in \Omega : |\boldsymbol{\beta} - \boldsymbol{\beta}_0| < \delta, |\sigma^2 - \sigma_0^2| < \delta, |\tau - \tau_0| < \delta\}$ according to Lemma 1. To simplify the notation, we assume that $\tau_0 = 0$. Because the derivation for a general model is lengthy, we consider the same model in Lemma 1, the MUTARE model with $k = 2$ in (9), and assuming $\tau > 0$, we have

$$
\begin{aligned}
&\mathrm{pl}(\boldsymbol{\beta}, \tau, \sigma^2) - \mathrm{pl}(\boldsymbol{\beta}, 0, \sigma^2) \\
&\quad = 2\log L(\boldsymbol{\beta}, \tau, \sigma^2) - 2\log L(\boldsymbol{\beta}, 0, \sigma^2) \\
&\quad = -\sum_i \sum_t \{[(y_{it}^* - H_{i,t-1}^* B_1)^2 - (y_{it}^* - H_{i,t-1}^* B_2)^2]Q_1 \\
&\qquad\qquad + [(y_{it}^* - H_{i,t-1}^* B_1)^2 - (y_{it}^* - H_{i,t-1}^* B_3)^2]Q_2 \\
&\qquad\qquad + [(y_{it}^* - H_{i,t-1}^* B_2)^2 - (y_{it}^* - H_{i,t-1}^* B_3)^2]Q_3\} \\
&\quad \leq -\sum_i \sum_t \{\big[2H_{i,t-1}^*(B_2 - B_1)\varepsilon_{it} + \big(H_{i,t-1}^*(A_2 - B_1)\big)^2 \\
&\qquad\qquad - \big(H_{i,t-1}^*(A_2 - B_2)\big)^2\big]Q_1 \\
&\qquad\qquad + \big[2H_{it-1}^*(B_3 - B_1)\varepsilon_{it} + \big(H_{i,t-1}^*(A_3 - B_1)\big)^2 \\
&\qquad\qquad - \big(H_{i,t-1}^*(A_3 - B_3)\big)^2\big]Q_2 \\
&\qquad\qquad + \big[2H_{it-1}^*(B_3 - B_2)\varepsilon_{it} + \big(H_{i,t-1}^*(A_3 - B_2)\big)^2 \\
&\qquad\qquad - \big(H_{i,t-1}^*(A_3 - B_3)\big)^2\big]Q_3\},
\end{aligned}
$$

where $Q_1 = I(0 < y_{i,t-1} \leq \tau, y_{i,t-2} \leq 0)$, $Q_2 = I(0 < y_{i,t-1} \leq \tau, y_{i,t-2} > 0)$, $Q_3 = I(\tau < y_{i,t-1}, 0 < y_{i,t-2} \leq \tau)$. If δ is sufficiently small, based on Assumption A2, we have $\sum_i \sum_j [(H_{i,t-1}^*(A_s - B_j))^2 - (H_{i,t-1}^*(A_s - B_s))^2]Q_k \geq 0$, for $k = 1, 2, 3$ and $s > j$. Therefore, by the same argument in Proposition 1 of Chan (1993), it holds that for all $\varepsilon > 0$, there exists a T such that with probability greater than $1 - \varepsilon$, $\boldsymbol{\gamma} \in \Omega_\delta$, $\tau > T/N$, implies $l(\boldsymbol{\beta}, \tau, \sigma^2) - l(\boldsymbol{\beta}, 0, \sigma^2) < 0$. Similar derivation can be extended to the case where $\tau < -T/N$. Hence, Theorem 1 holds.

APPENDIX D: PROOF OF THEOREM 2

We first prove the asymptotic normality. Based on the adaptive lasso penalty,

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}} npl(\mathbf{u}),$$

where $npl(\mathbf{u}) = -2\log L(\boldsymbol{\beta} + \mathbf{u}, \sigma^2, \tau) + \lambda_1 \sum_{j=1}^{k} \nu_j(|\beta_j + u_j|) + \lambda_2 \max_j\{j : \beta_j + u_j \neq 0\}$. By the Taylor expansion, we have

$$\begin{aligned} npl(\mathbf{u}) = {} & npl(\mathbf{0}) - \mathbf{u}'\tilde{H}'\mathbf{W}^{-1}(\sigma)\big(Y - g(H, \boldsymbol{\beta}, \tau, \sigma^2)\big) \\ & + \frac{1}{2}\sqrt{N}\mathbf{u}'\left(\frac{\tilde{H}\mathbf{W}^{-1}(\sigma)\tilde{H}'}{N}\right)\sqrt{N}\mathbf{u} \\ & + \lambda_1 \sum_{j=1}^{k} \nu_j(|\beta_j + u_j| - |\beta_j|) \\ & + \lambda_2\Big(\max_j\{j : \beta_j + u_j \neq 0\} - \max_j\{j : \beta_j \neq 0\}\Big). \end{aligned}$$

The last term on the right-hand side equals 0 if $u_j = 0$ and $\beta_j = 0$, combining with the fact that [Zou (2006)]

$$\lambda_1 \nu_j(|\beta_j + u_j| - |\beta_j|) \to_{\mathcal{P}} \begin{cases} 0, & \text{if } \beta_j \neq 0, \\ 0, & \text{if } \beta_j = 0 \text{ and } u_j = 0, \\ \infty, & \text{if } \beta_j = 0 \text{ and } u_j \neq 0, \end{cases} \tag{11}$$

we have for every u

$$\begin{aligned} & npl(\mathbf{u}) - npl(\mathbf{0}) \\ & \quad \to_D \begin{cases} -\mathbf{u}_{(1)}'\tilde{H}(1)'\mathbf{W}^{-1}(\sigma)\big(Y - g(H, \boldsymbol{\beta}, \tau, \sigma^2)\big) \\ \qquad + \dfrac{(\sqrt{N}\mathbf{u}_{(1)})'\Lambda_{11}(\sqrt{N}\mathbf{u}_{(1)})}{2}, & \text{if } u_{(2)} = 0, \\ \infty, & \text{otherwise.} \end{cases} \end{aligned}$$

By the same argument of Theorem 2 in Zou (2006), the asymptotic normality holds by the martingale central limit theorem [Hall and Heyde (1980)].

For consistency, it suffices to show that $P(\hat{\beta}_{(2)} \neq 0) \to 0$. Using the Karush–Kuhn–Tucker (KKT) optimality conditions, it follows that

$$2\tilde{H}(1)'\mathbf{W}^{-1}(\sigma)\big(Y - g(H, \hat{\boldsymbol{\beta}}, \tau, \sigma^2)\big) = \lambda_1 \nu_{(1)},$$

where $\nu_{(1)}$ are the weights corresponding to the first q variables. Note that $\lambda \frac{\nu_{(1)}}{\sqrt{N}} \to_{\mathcal{P}} \infty$ [Theorem 2, Zou (2006)] and $2\frac{\tilde{H}(1)'\mathbf{W}^{-1}(\sigma)(Y - g(H, \hat{\boldsymbol{\beta}}, \tau, \sigma^2))}{\sqrt{N}}$ is asymptotically normal. Therefore,

$$P\big(\hat{\beta}_{(2)} \neq 0\big) \leq P\big(2\tilde{H}(1)'\mathbf{W}^{-1}(\sigma)\big(Y - g(H, \hat{\boldsymbol{\beta}}, \tau, \sigma^2)\big) = \lambda_1 \nu_{(1)}\big) \to 0,$$

and Theorem 2 holds.

Acknowledgments. The author is grateful to the Editor, Associate Editor and two referees for their helpful comments and suggestions, and the author would like to thank C. F. Jeff Wu and C. Zhu for helpful discussions.

REFERENCES

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (*Tsahkadsor*, 1971) (B. N. Petrov and F. Csaki, eds.) 267–281. Akad. Kiadó, Budapest. MR0483125

BURMAN, P., CHOW, E. and NOLAN, D. (1994). A cross-validatory method for dependent data. *Biometrika* **81** 351–358. MR1294896

CHAN, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21** 520–533. MR1212191

CHEN, J. and KHALILI, A. (2008). Order selection in finite mixture models with a nonsmooth penalty. *J. Amer. Statist. Assoc.* **103** 1674–1683. MR2722574

CHEN, W., EVANS, E. A., MCEVER, R. P. and ZHU, C. (2008). Monitoring receptor-ligand interactions between surfaces by thermal fluctuations. *Biophys. J.* **94** 694–701.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **31** 377–403. MR0516581

DIGGLE, P. J., HEAGERTY, P. J., LIANG, K.-Y. and ZEGER, S. L. (2002). *Analysis of Longitudinal Data*, 2nd ed. *Oxford Statistical Science Series* **25**. Oxford Univ. Press, Oxford. MR2049007

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. Chapman and Hall, London. MR1383587

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581

FAN, J. and YAO, Q. (2003). *Nonlinear Time Series*: *Nonparametric and Parametric Methods*. Springer, New York. MR1964455

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. MR0624435

HANSEN, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* **68** 575–603. MR1769379

HUANG, J., ZARNITSYNA, V. I., LIU, B., EDWARDS, L. J., CHIEN, Y. H., JIANG, N., EVAVOLD, B. D. and ZHU, C. (2010). The kinetics of two-dimensional TCR and pMHC interactions determine T-cell responsiveness. *Nature* **464** 932–936.

HUNG, Y. (2011). Maximum likelihood estimation of nonlinear time series models. Technical report, Dept. Statistics and Biostatistics, Rutgers Univ., Piscataway, NJ.

HUNG, Y., ZARNITSYNA, V., ZHANG, Y., ZHU, C. and WU, C. F. J. (2008). Binary time series modeling with application to adhesion frequency experiments. *J. Amer. Statist. Assoc.* **103** 1248–1259. MR2462896

HYMAN, J. M. and LAFORCE, T. (2003). Modeling the spread of influenza among cities. In *Biomathematical Modeling Applications for Homeland Security* (T. Banks and C. Castillo-Chavez, eds.). SIAM, Philadelphia, PA.

JIANG, J. (1996). REML estimation: Asymptotic behavior and related topics. *Ann. Statist.* **24** 255–286. MR1389890

LIU, Y. and WU, Y. (2007). Variable selection via a combination of the L_0 and L_1 penalties. *J. Comput. Graph. Statist.* **16** 782–798. MR2412482

MARSHALL, B. T., LONG, M., PIPER, J. W., YAGO, T., MCEVER, R. P. and ZHU, C. (2003). Direct observation of catch bonds involving cell-adhesion molecules. *Nature* **423** 190–193.

MARSHALL, B. T., SARANGAPANI, K. K., LOU, J., MCEVER, R. P. and ZHU, C. (2005). Force history dependence of receptor-ligand dissociation. *Biophys. J.* **88** 1458–1466.

MCCULLOCH, C. E. and SEARLE, S. R. (2008). *Generalized, Linear, and Mixed Models*, 2nd ed. Wiley, New York.

NEMHAUSER, G. and WOLSEY, L. (1999). *Integer and Combinatorial Optimization*. Wiley, New York. MR1699321

NIE, L. (2006). Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed-effects models. *Metrika* **63** 123–143. MR2242536

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York. MR0762984

RACINE, J. (2000). Consistent cross-validatory model-selection for dependent data: *hv*-block cross-validation. *J. Econometrics* **99** 39–61.

SAMIA, N. I. and CHAN, K.-S. (2011). Maximum likelihood estimation of a generalized threshold stochastic regression model. *Biometrika* **98** 433–448. MR2806439

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** 101–118. MR2367827

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

STONE, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *J. Roy. Statist. Soc. Ser. B* **36** 111–147. MR0356377

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

TIBSHIRANI, R. J. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* **16** 385–395.

TONG, H. (1980). Threshold autoregression, limit cycles and cyclical data. *J. Roy. Statist. Soc. Ser. B* **42** 245–292.

TONG, H. (1983). *Threshold Models in Nonlinear Time Series Analysis. Lecture Notes in Statistics* **21**. Springer, New York. MR0717388

TONG, H. (1990). *Nonlinear Time Series: A Dynamical System Approach. Oxford Statistical Science Series* **6**. Oxford Univ. Press, New York. MR1079320

TONG, H. (2007). Birth of the threshold time series model. *Statist. Sinica* **17** 8–14.

TONG, H. and LIM, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. Roy. Statist. Soc. Ser. B* **42** 245–292.

TSAY, R. S. (1989). Testing and modeling threshold autoregressive processes. *J. Amer. Statist. Assoc.* **84** 231–240. MR0999683

ZARNITSYNA, V. I., HUANG, J., ZHANG, F., CHIEN, Y. H., LECKBAND, D. and ZHU, C. (2007). Memory in receptor-ligand mediated cell adhesion. *Proc. Natl. Acad. Sci. USA* **104** 18037–18042.

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

DEPARTMENT OF STATISTICS
AND BIOSTATISTICS
RUTGERS UNIVERSITY
PISCATAWAY, NEW JERSEY 08854
USA
E-MAIL: yhung@stat.rutgers.edu