Sankhyā : The Indian Journal of Statistics
2004, Volume 66, Part 3, pp 536-547

Approximating Dependence Structures of Repeated Stochastic Processes

Bertrand Clarke
University of British Columbia, Vancouver, Canada
Peter X.-K. Song
York University, Toronto, Canada

Abstract

We present a general procedure for joint modelling of the mean structure and the stochastic dependence for longitudinal data. To reveal the underlying dependence mechanism, we proceed in three steps. First, we use cross-sectional regression to relocate the data to achieve marginal stationarity. Second, we discretize the relocated data. Third, we model the dependence structure of the discretized data as a stationary Markov chain with sufficiently high order. The procedure is primarily developed for continuous responses, but it is applicable for discrete responses that emerge from an underlying continuous process. Two data analysis examples are presented to illustrate our procedure.

AMS (2000) *subject classification.* Primary 62M10; secondary 62-10.
Keywords and phrases. Discretization, longitudinal data, Markovity, nonstationarity.

1 Introduction

Dependence and nonstationarity are, essentially, always present in data observed sequentially over time. It seems there has been little study of statistical models that jointly address mean and dependence structures, see Diggle et al. (1994, Chap. 10, p.207). Statistical inference can become very challenging when modelling repeated stochastic processes is of primary interest. Rather than treating the dependence structure as nuisance as most of current methods do in the analysis of longitudinal data, here we focus on the development of statistical procedure to explore the mechanism of stochastic dependence.

We present a general approach to a large class of repeated stochastic processes. It is based on treating dependence and nonstationarity disjointly.

We do not deal with the dependence structure nonstationarity except to remain aware we are using stationary approximations. To deal with marginal nonstationarity, we assume a location model in which it is enough to model the mean of each random variable in the sequence as a function of the covariates of interest. To tackle possible dependence structures in the relocated sequence, we discretize and then "Markovize" by approximating it with a stationary Markov process of a sufficiently high order.

A key assumption is that the relocation by a regression function results in a stationary residual process. Thus, we imagine that the mean of the response Y_t drifts over time according to a deterministic function of the covariates. For each subject i, $i = 1, \ldots, K$ we relocate $Y_{i,t}$ to $Z_{i,t} = Y_{i,t} - \beta_{i,t}$ where the mean for subject i has a linear model

$$\beta_{i,t} = \sum_{j=1}^{p} \gamma_{t,j} X_{i,t,j}, \tag{1.1}$$

in which the $\gamma_{t,j}$'s are regression coefficients for the p covariates $X_{i,t,j}$, with $j = 1, \ldots, p$ at time t ($p < K$). In some practical settings, the variance of responses is time-varying, such as volatility in financial time series, one then has to rescale the response by replacing $Y_{i,t}$ with $Z_{i,t} = (Y_{i,t} - \beta_{i,t})/\sigma_{i,t}$ for each subject, where $\beta_{i,t}$ is modelled by (1.1) and $\sigma_{i,t}$ is modelled analogously by

$$\log \sigma_{i,t}^2 = \sum_{j=1}^{p} \tilde{\gamma}_{t,j} X_{i,t,j}. \tag{1.2}$$

For ease of exposition we ignore this added complexity by assuming that the marginals have the same shape after relocating and that the linear regression is sufficient to remove time-varying marginal patterns. Moreover, either the mean model (1.1) or variance model (1.2) may be specified as a nonparametric regression fashion, such as the generalized additive models (Hastie and Tibshirani, 1990), if the data suggest so.

In Section 2, we briefly outline the key steps in our approximation. Section 3 examines the key choices in detail. Section 4 situates our methods in a general stochastic process context to indicate the role of our approximation. In Section 5, we demonstrate our approach in two applications.

2 General Approach

For convenience, we suppress the subject index i in this section. Let $\{Y_t, t = 1, 2, \ldots, \}$ be a stochastic process taking quantitative values in an

ordered set. Assume that for a given integer n, the joint density of $Y_1, \ldots, Y_n$ is a member of a parametric family, defined by

$$p_{\boldsymbol{\beta},\boldsymbol{\alpha}}(y_1, \ldots, y_n) = p_{\boldsymbol{\alpha}}(y_1 - \beta_1, \cdots, y_n - \beta_n) = p_{\boldsymbol{\alpha}}(z_1, \ldots, z_n), \qquad (2.1)$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$ represents the location of each response variable and $Z_t = Y_t - \beta_t$ are relocated random variables having the same sample space $\mathcal{X}$ and mean 0. The densities $p_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and $p_{\boldsymbol{\alpha}}$ in (2.1) are taken with respect to a common dominating measure μ.

Suppose we have a finite partition $\mathcal{P} = (S_1, \ldots, S_\ell)$ of the sample space $\mathcal{X}$. Then for fixed $\boldsymbol{\alpha}$ and n, the joint distribution of $Z_1, \ldots, Z_n$ can be approximated by

$$p_{\boldsymbol{\alpha}}(z_1, \ldots, z_n)\mu(S_{i_1}) \times \cdots \times \mu(S_{i_n}) \approx P_{\boldsymbol{\alpha}}(Z_1 \in S_{i_1}, \ldots, Z_n \in S_{i_n}), \quad (2.2)$$

for $z_1 \in S_{i_1}, \ldots, z_n \in S_{i_n}$. Writing $y_i^{i+j} = (y_i, y_{i+1}, \ldots, y_{i+j})$, we can factor the probability on the right-hand side of equation (2.2) as

$$P_{\boldsymbol{\alpha}}(Z_1 \in S_{i_1}) P_{\boldsymbol{\alpha}}(Z_2 \in S_{i_2} | Z_1 \in S_{i_1}) \cdots P_{\boldsymbol{\alpha}}(Z_n \in S_{i_n} | Z_1^{n-1} \in S_{i_1}^{i_{n-1}}) \quad (2.3)$$

which can be further approximated by

$$\prod_{j=1}^{n} P_{\boldsymbol{\alpha}}(Z_j \in S_{i_j} | Z_{j-r}^{j-1} \in S_{i_{j-r}}^{i_{j-1}}) \qquad (2.4)$$

where r is the order of the approximating Markov chain.

For the special case that $r = 1$, we get the transition equation

$$\begin{bmatrix} P(Z_t \in S_1) \\ \vdots \\ P(Z_t \in S_\ell) \end{bmatrix} = [P_{\boldsymbol{\alpha}}(Z_t \in S_i | Z_{t-1} \in S_j)]_{i,j=1,\ldots,\ell} \begin{bmatrix} P(Z_{t-1} \in S_1) \\ \vdots \\ P(Z_{t-1} \in S_\ell) \end{bmatrix}. \qquad (2.5)$$

Denote the transition matrix in (2.5) by $T(\boldsymbol{\alpha}, t)$ with entries $T_{ij}(\boldsymbol{\alpha}, t)$ that only depend on the parameter $\boldsymbol{\alpha}$. The transition matrix $T(\boldsymbol{\alpha}, t)$ is the ultimate object that we aim to obtain through our procedure to examine the stochastic dependence structure of the data. Note that parameter $\boldsymbol{\alpha}$ does not appear in the marginal probability vectors due to the marginal stationarity of the relocated process Z_t, which is crucial to proceed approximation by a stationary Markov chain.

The above procedure rests on the fact that Markovity is a property of probabilities, rather than expectations, and so is compatible with many mean structures. This generality allows us to cover both continuous and discrete responses. Moreover, the discretization is helpful to overcome the difficulty of estimating transition intensities under a continuous sample space.

3 Choice of Approximations

The relocation task is easy to accomplish by running the cross-sectional linear least squares regression with model (1.1) at each time, $t = 1, \ldots, n$. The resulting vectors of residuals are then $(Z_{1,t}, \ldots, Z_{K,t})$ at time $t = 1, \ldots, n$. These vectors are then further standardized by division by $\sqrt{1 - h_{t,ii}}$ where $h_{t,ii}$ is the i-th diagonal element of the hat matrix in the t-th regression.

For each of the n standardized vectors, $(Z_{1,t}, \ldots, Z_{K,t}), t = 1, \ldots, n$, we discretize and approximate by a Markov chain so as to fit a model of the form (2.5). This gives a transition matrix $T(\boldsymbol{\alpha}) = T(\boldsymbol{\alpha}, t)$ which is independent of time, but is a function of $\boldsymbol{\alpha}$ which we must estimate.

3.1 Choosing the order of the Markov chain. The relocated processes are marginally stationary and can be approximated by a stochastic process whose autocorrelation depends only on time lags. The autocorrelation function (ACF) is $\rho(h) = \gamma(h)/\gamma(0), h \geq 1$, where $\gamma(h) = \mathrm{cov}(Z_t, Z_{t+h}), h \geq 1$ is the autocovariance function (ACVF) and $\gamma(0)$ is the variance. The sample ACF for the sequence of standardized residuals leads to K estimated Markov orders, each for one subject. If all of these empirical orders are the same there is no problem. Otherwise, we form a histogram from these estimated orders and use its mode. As usual, we estimate the order of the process by the maximum lag for which the sample ACF exceeds the approximate 95% confidence limits $\pm 2/\sqrt{n}$, see Brockwell and Davis (1996, Chap. 6).

3.2 Choosing a partition. Next, choose the partition $\mathcal{P}$ for given $\ell = \mathrm{card}(\mathcal{P})$. Consider any univariate random variable $Z : \Omega \to \mathcal{X}$ and the case $\ell = 2$. Approximating Z by a step function in L_1, we want to find a_1 and a_2 as well as disjoint sets A_1 and A_2 (with $A_1 \cup A_2 = \Omega$) to achieve

$$\min_{a_1, a_2, A_1, A_2} E|Z - a_1\chi_{A_1} - a_2\chi_{A_2}|. \tag{3.1}$$

The argument of the minimum in (3.1) equals $E|(Z - a_1)\chi_{A_1}| + E|(Z - a_2)\chi_{A_2}|$, and these two terms can be separately minimized to see that a_1 must be the median of $Z\chi_{A_1}$ and a_2 must be the median of $Z\chi_{A_2}$. When Z is symmetric about τ, the natural choices for A_1 and A_2 are $A_1 = Z^{-1}(S_1) = \{\omega \in \Omega : Z(\omega) \in S_1\}$ and $A_2 = Z^{-1}(S_2)$, where $S_1 = (-\infty, \tau]$ and $S_2 = (\tau, \infty)$. Then one would choose the sample median as an estimator of τ. This argument generalizes to the case $\ell \geq 3$, leading to the following.

PROPOSITION 3.1 *: Suppose the support of the density of Z is a connected set and let the integer $\ell \geq 2$. An optimal choice for $\mathcal{P}$ under L_1 distance is*

the collection of sets defined by the j/ℓ percentiles, where $j = 1, ..., \ell-1$. That is, the partition can be optimally defined by the collection of $\ell - 1$ boundary points dividing the distribution into ℓ equally probable connected regions.

The proof of this result is as follows. First, the argument is analogous to the case $\ell = 2$ when ℓ is even. When ℓ is odd, there will be a middle bin on which one takes the median in addition to the even number of bins on either side of the point of symmetry.

It remains to choose ℓ. Let $m(z)$ be the (marginal) density for Z. For each ℓ, we can approximate $m(z)$ by

$$m_\ell(z) = \sum_{i=1}^{\ell} m(\text{med}_i)\, \chi_{S_i}(z)$$

where $\text{med}_i = \text{median}(Z\chi_{A_i})$ and the S_i's are the sets in $\mathcal{P}$ for the given ℓ. It is seen that $m_\ell(z)$ is a step function motivated by Proposition 1, and as $\ell \to \infty$, $E|m(z) - m_\ell(z)| \to 0$. Thus, we can choose the smallest ℓ for which $E|m(z) - m_\ell(z)| \leq \epsilon$, with a pre-specified precision ϵ.

3.3 The categorical response case. Following McCullagh (1980), our procedure first represents c-level categorical responses, $\{Y_{i,t} \in \{1, \ldots, c\}, t = 1, \ldots, n\}$, in terms of underlying continuous stochastic processes $\{U_{i,t}, t = 1, \ldots, n\}$, each for one subject $i = 1, \ldots, K$. Thus, the defining relation between $Y_{i,t}$ and the latent process $U_{i,t}$ is given by

$$Y_{i,t} = h \;\Leftrightarrow\; a_{t,h-1} < U_{i,t} \leq a_{t,h},\; h = 1, \ldots, c,$$

where $-\infty = a_{t,0} < a_{t,1} < \cdots < a_{t,c-1} < a_{t,c} = \infty$ are time-dependent, but subject-independent, threshold points, which can be estimated cross-sectionally. Naturally, we apply a variant of our procedure from Section 3.2 to the underlying continuous $U_{i,t}$. We use the cumulative probit regression model to get marginal stationarity. To proceed, we assume $U_{i,t} \sim N(\beta_{i,t}, 1)$ with mean $\beta_{i,t}$ being defined in (1.1). The resulting model takes the from

$$\eta_{i,t,h} = \Phi^{-1}\{P(Y_{i,t} \leq h)\} = a_{t,h} - \beta_{i,t}, \tag{3.2}$$

and the cumulative probability is

$$P(Y_{i,t} \leq h) = P(U_{i,t} \leq a_{t,h}) = P(\xi_t \leq a_{t,h} - \beta_{i,t}) = \Phi(a_{t,h} - \beta_{i,t}),\; h = 1, \ldots, c, \tag{3.3}$$

where $(\xi_{t=1}, \ldots, \xi_{t=n})'$ is an independent copy of the relocated $(U_{i,t=1}, \ldots, U_{i,t=n})'$, for any i, and so has $\xi_t \sim N(0, 1)$ marginally. The regression coefficients and the threshold points in (3.2) can be obtained by maximum likelihood estimation (McCullagh and Nelder, 1989).

A first-order Markov model requires that the conditional probabilities

$$P(Y_{i,t+1}=h|Y_{i,t}=h')=P(a_{t+1,h-1}<U_{i,t+1}\leq a_{t+1,h}\,|\,a_{t,h'-1}<U_{i,t}\leq a_{t,h'}). \quad (3.4)$$

Using the MLE's $\widehat{a}_{t,h}$ and $\widehat{\beta}_{i,t}$ to get $\widehat{\eta}_{i,t,h}$, we approximate in (3.4) by

$$P(\widehat{\eta}_{i,t+1,h-1} < \xi_{t+1} \leq \widehat{\eta}_{i,t+1,h}|\widehat{\eta}_{i,t,h'-1} < \xi_t \leq \widehat{\eta}_{i,t,h'}). \quad (3.5)$$

(Subtract $\beta_{i,t}$ from $U_{i,t}$ in (3.4), and use MLE's to get the fitted values $\hat{\eta}_{i,t,h}$'s.) In order to pool the cross-sectional data to estimate transition probabilities, we must remove the dependence on i in $\widehat{\eta}_{i,t,h}$'s. Also we must remove the dependence on t, so our approximating Markov chain will have stationary transition probabilities.

To obtain threshold points that depend only on h, let $\eta_{med,t,h}$ be the cross-sectional median of the $\eta_{i,t,h}$ over subjects i, and let $\eta_{med,med,h}$ be the global median of the $\eta_{med,t,h}$ over times t. We informally justify this choice by the following argument. For a sequence $\eta_{med,t,h}, t = 1, \ldots, n$ with h fixed, we want to find the optimal value of η, $\eta_{med,med,h}$ which achieves

$$\min_{\eta} \sum_{t} |P(\xi_t \leq \eta_{med,t,h}) - P(\xi_t \leq \eta)|.$$

Note $\Phi(\eta) = P(\xi_t \leq \eta)$, which is strictly increasing,

$$\eta_{med,med,h} = \arg\min_{\eta} \sum_{t} |\Phi(\eta_{med,t,h}) - \Phi(\eta)| = \arg\min_{\eta} \sum_{t} |\eta_{med,t,h} - \eta|.$$

Thus our approximate transition probabilities of $h' \hookrightarrow h$ are

$$\begin{aligned} P(Y_{t+1} \;=\; & h|Y_t = h') \approx P\,(\eta_{med,med,h-1} \\ & < \xi_{t+1} \leq \eta_{med,med,h}|\eta_{med,med,h'-1} < \xi_t \leq \eta_{med,med,h'})\,. \end{aligned}$$

Approximations for higher order conditional probabilities can be derived analogously.

4 Properties of the Approximation Procedure

We examine goodness-of-fit of the procedure in the sense that the model we analyse is close to the data generation processes. First note that the subjective features introduced are clearly identified, are amenable to robustness analyses, and can be related to the phenomenon under investigation. Because our procedure is for exploratory data analysis, we neglect standard

errors to focus on model selection, which seems more important than estimation within a model.

Given the completion of the relocation by (1.1) and the rescaling by (1.2), we write $p(z_1^n)$ for the density of the stochastic process generating the residuals. This $p(z_1^n)$ is similar to the right-hand side of (2.1); however, (2.1) has a subscript $\boldsymbol{\alpha}$ which arises only after the Markovization and discretization have been done.

For a given partition $\mathcal{P}$ and a given order r of a Markov chain, we obtain a parameterization following the procedure of Section 2.1, characterized by the (transition) parameter $\boldsymbol{\alpha}$. Thus, the joint density of the approximating Markov chain is $p_{\mathcal{P},r,\boldsymbol{\alpha}}(z_1^n)$. Our task is to identify an estimator $\hat{\boldsymbol{\alpha}}$ so that for $\mathcal{P}$ rich enough, r high enough and n large enough, we will have

$$E_{Z^n}|p(Z_1^n) - p_{\mathcal{P},r,\hat{\boldsymbol{\alpha}}}(Z_1^n)| \leq \epsilon \tag{4.1}$$

for pre-assigned ϵ. Note that the left-hand side of (4.1) only goes to zero when card($\mathcal{P}$), r, and n all go to infinity together.

Expression (4.1) can not be used as it stands because it involves the unknown density $p(z_1^n)$. However when the required convergence holds and the two densities $p(Z_1^n)$ and $p_{\mathcal{P},r,\hat{\boldsymbol{\alpha}}}(Z_1^n)$ are bounded, we can replace the expectation in (4.1) to get

$$E_{p_{\mathcal{P},r,\boldsymbol{\alpha}}}|p(Z_1^n) - p_{\mathcal{P},r,\hat{\boldsymbol{\alpha}}}(Z_1^n)| \leq \epsilon \tag{4.2}$$

in which the expectation is now with respect to $p_{\mathcal{P},r,\boldsymbol{\alpha}}(z_1^n)$, where $\boldsymbol{\alpha}$ is the true value. Note that going from (4.1) to (4.2) is a change in the density defining the L_1 distance.

If we choose card($\mathcal{P}$) and r large enough, the density $p_{\mathcal{P},r,\boldsymbol{\alpha}_0}$ generated from our approximation technique will be so close to the density $p(z_1^n)$ of the actual stochastic process that we can replace $p(z_1^n)$ by $p_{\mathcal{P},r,\boldsymbol{\alpha}_0}$ in (4.2). Consequently for card($\mathcal{P}$) and r large enough and any $\boldsymbol{\alpha}_0$ taken as true, we can make n large enough that

$$E_{p_{\mathcal{P},r,\boldsymbol{\alpha}_0}}|p_{\mathcal{P},r,\boldsymbol{\alpha}_0}(Z_1^n) - p_{\mathcal{P},r,\hat{\boldsymbol{\alpha}}}(Z_1^n)| \leq \epsilon$$

for any pre-assigned value of ϵ.

5 Computational Illustrations

We present two examples in this section to illustrate the proposed procedure, one in an economic context and the other in a biostatistical context.

5.1 Continuous response example. We tested our techniques by analysing an economic data set ('strikes') that we found at the website lib.stat.cmu.edu. This data set contains penta-variate time series of annual observations from 1951 to 1980 for 18 OECD countries. (The time series from country 3 ends at 1980, so we truncated time series from all countries at 1980 for simplicity.) We regard each country as a subject, so we have five measurements on each subject at each of 30 years. We took the first of the five variables, unemployment rate, as the response $Y_{i,t}$, and the other four variables, strike volume $X_{i,t,1}$, inflation $X_{i,t,2}$, parliamentary representation of social democratic and labour parties $X_{i,t,3}$, and a time-invariant measure of union centralization $X_{i,t,4}$, as the covariates. For more details, see Western (1996).

TABLE 1. HISTOGRAMS OF EMPIRICAL ORDERS FOR SEVERAL MODELS

Model	Expression	Histogram
(1)	$Y \sim X_1$	5 6 2 3 1 1
(2)	$Y \sim X_1 + X_2$	7 6 0 4 1
(3)	$Y \sim X_1 + X_2 + X_3$	9 3 0 4 1
(4)	$Y \sim X_1 + X_2 + X_4$	9 3 3 2 1
(5)	$Y \sim X_1 + X_2 + X_3 + X_4$	7 5 1 3 1

The first stage is to find a collection of explanatory variables so that the vector of residuals has a recognizable dependence structure. We used $X_{i,t,1}$, $X_{i,t,2}$, $X_{i,t,3}$ and $X_{i,t,4}$. For each model, we fit the response to the covariates for each time and we combined the residual vectors from each time into a single 18 by 30 matrix. Next we treated each row as a time series and produced an ACF plot. Each ACF plot gave one empirical order of autoregressive dependence. Following Section 3.1, we used $\pm 2/\sqrt{30}$ as the cutoff. So, we chose the largest significant time-lag from each ACF plot as the subject-specific empirical Markov order. Thus we got 18 empirical orders, one for each country, which we used to generate a histogram. Table 1 lists the models we examined. The column labelled Histogram lists the number of subjects (countries) that had empirical orders 1 to 6.

It is seen that the fourth model has the strongest mode at lag 1 and is unimodal. So, we examine the dependence structure of the residuals from model (4).

Next, we need a partition size for the discretization. For the economic variables here, there are periods in which the economy is growing, shrinking, and flat. So, we chose a partition of cardinality 3. By Proposition 1, we want the 33rd and 67th percentiles from the distribution of residuals from model (4). We estimated these percentiles by the sample percentiles obtained from pooling the entries of the 18 by 30 matrix of residuals, -0.8773 and

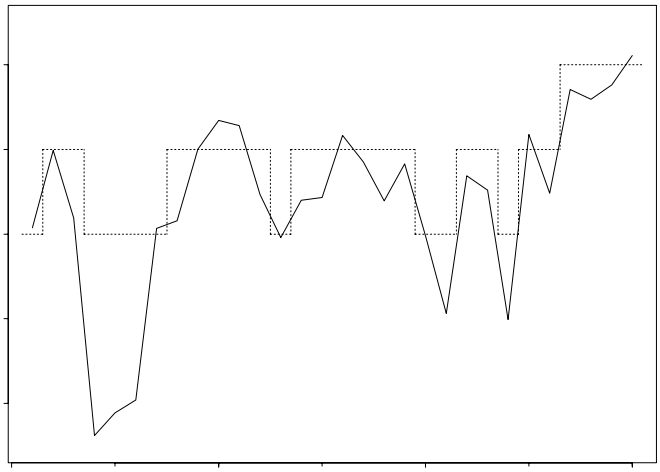

FIGURE 1. RESIDUAL AND DISCRETIZED PROCESSES FOR SUBJECT 1

0.4883. If a residual is equal to -0.8773 or less, it is replaced by -1; if a residual is between -0.8773 and 0.4883, it is replaced by 0; otherwise, it is replaced by 1. For instance, Figure 1 displays the first row of the matrix of residuals (solid line) and the discretized form of these residuals (dotted line).

We notice that in the discretized process for subject 1, -1 occurs 9 times. There are four years in which a -1 is followed by another -1; there are five years in which a -1 is followed by a 0; and no years in which -1 is followed by 1. So, country 1 contributes four to the frequency of transition from -1 to -1, five to the frequency of transition from -1 to 0, and zero to the frequency of transition from -1 to 1. Doing this over all countries and taking the sum of the counts relative to the total number of transitions provides an estimate of the 3×3 matrix of transition probabilities. For model (4), with partition size $\ell = 3$ of $S_1 = (-\infty, -0.8773] \cup S_2 = (-0.8773, 0.4883] \cup S_3 = (0.4883, \infty)$, the matrix of estimated transition probabilities, as in (2.5), is

$$\left[\widehat{P}_{\widehat{\boldsymbol{\alpha}}}(Z_t \in S_i | Z_{t-1} \in S_j)\right]_{3\times 3} = \begin{bmatrix} 0.7572 & 0.2139 & 0.0289 \\ 0.1808 & 0.6271 & 0.1921 \\ 0.0349 & 0.1802 & 0.7849 \end{bmatrix}.$$

This estimated transition matrix suggests that the unemployment status tends to stay in its current state with relatively high probabilities 60%–78%

(the diagonal elements) and that with about 0.2 probability the unemployment status will change from one state to a neighbouring state, either better or worse.

5.2 Discrete response example. To illustrate our method in Section 3.3, we use a data set from the Betaseron longitudinal clinical trial on multiple sclerosis (MS) which compared three dosage levels of interferon beta-1b, namely zero (placebo), low and high. For more details on the data, refer to Dyachkova et al. (1997). We model the binary indicator variable of exacerbation ($Y_{i,t} = 1(Yes)$; $0(No)$) in terms of the Kurtzke Expanded Disability Status Scale (EDSS or $X_{i,t,1}$), and the burden of disease ($X_{i,t,2}$). We use 13 (K) subjects for whom we had complete data collected from 17 (n) equally spaced time points.

Given an underlying Gaussian process $\{U_{i,t}\}$ and the threshold points a_t, the correspondence relationship is defined by

$$Y_{i,t} = 0 \Leftrightarrow \{U_{i,t} \leq a_t\} \text{ and } Y_{i,t} = 1 \Leftrightarrow \{U_{i,t} > a_t\}.$$

Thus, the probit regression model can be specified by assuming $U_{i,t} \sim N(\beta_{i,t}, 1)$ with $\beta_{i,t} = \gamma_{t,1}X_{i,t,1} + \gamma_{t,2}X_{i,t,2}$. That is, $\Phi^{-1}(\mu_{i,t}) = -\eta_{i,t} = -a_t + \gamma_{t,1}X_{i,t,1} + \gamma_{t,2}X_{i,t,2}$. In addition, the marginal probability of exacerbation at time t is $\mu_{i,t} = P(U_{i,t} > a_t)$, and $U_{i,t} - \beta_{i,t} \sim N(0, 1)$. The maximum likelihood estimates of the parameters $(a_t, \gamma_{t,1}, \gamma_{t,2})$ for $t = 1, \ldots, 17$ can be obtained by a statistical software such as S-plus.

It is worth noting that the probit regression at time step 9 failed because the $Y_{i,9}$'s are zero. The estimated threshold point $a_9 = \infty$ and the other two γ parameters are both zero. However, this unusual estimates will not affect our further analysis because we will use a robust median statistic to estimate the global threshold point.

Our goal is to approximate the first-order transition probabilities. Based on a 13 by 17 matrix of fitted values of $\hat{\eta}_{i,t}$, we yielded the global median $\hat{\eta}_{med,med} = 1.129$ for the threshold point. Thus, the stationary transition probability, for example, $P(Y_t = 1 | Y_{t-1} = 0)$ is approximated by

$$P_\rho(\xi_t > \eta_{med,med} | \xi_{t-1} \leq \eta_{med,med}) = \frac{\Phi(\hat{\eta}_{med,med}) - \Phi_\rho(\hat{\eta}_{med,med}, \hat{\eta}_{med,med})}{\Phi(\hat{\eta}_{med,med})}, \tag{5.1}$$

where $(\xi_t, \xi_{t-1}) \sim N_2(0, 0, 1, 1, \rho)$. To estimate the correlation ρ in (5.1) we

invoke a simple method of moments technique. Note that

$$\begin{aligned} E_\rho(Y_{i,t}Y_{i,t-1}) &= P_\rho(\xi_t > \eta_{i,t}, \xi_{t-1} > \eta_{i,t-1}) \\ &\approx P_\rho(\xi_t > \hat{\eta}_{i,t}, \xi_{t-1} > \hat{\eta}_{i,t-1}) \\ &= 1 - \Phi(\hat{\eta}_{i,t}) - \Phi(\hat{\eta}_{i,t-1}) + \Phi_\rho(\hat{\eta}_{i,t}, \hat{\eta}_{i,t-1}). \end{aligned}$$

The approximation error above is negligible since $\hat{\eta}_{i,t}$ converges to $\eta_{i,t}$. Solving an unbiased estimating equation,

$$\sum_{i=1}^{K}\sum_{t=2}^{n}\{Y_{i,t}Y_{i,t-1} - E_\rho(Y_{i,t}Y_{i,t-1})\} = 0, \tag{5.2}$$

leads to a consistent estimator of ρ (Heyde, 1997). Through the grid point search algorithm, we found the solution to (5.2), $\hat{\rho} = -0.26$. Analogously, we get the estimated first-order transition matrix

$$T(\hat{\rho}) = \begin{bmatrix} P_{\hat{\rho}}(0|0) & P_{\hat{\rho}}(0|1) \\ P_{\hat{\rho}}(1|0) & P_{\hat{\rho}}(1|1) \end{bmatrix} = \begin{bmatrix} 0.864 & 0.945 \\ 0.136 & 0.055 \end{bmatrix}.$$

The (1,2) entry says that a patient experiencing an exacerbation at a given time step is very likely, nearly 95%, to recover within the next time step of 6 weeks. The $(1,1)$ entry says that a patient who is not experiencing an exacerbation at a given time step is relatively unlikely, about 85%, to experience an exacerbation within the next 6 weeks.

Acknowledgements. The authors are grateful to the Associate Editor and the Editor for their constructive suggestions that helped to improve the exposition of the paper. This work was supported by NSERC Operating Grants held by both authors.

References

Brockwell, P.J. and Davis, R.A. (1996). *Time Series: Theory and Methods*, 2nd ed., Springer, New York.

Diggle, P., Liang, K.-Y. and Zeger, S.L. (1994). *Analysis of Longitudinal Data*, Clarendon Press, Oxford.

Dyachkova, Y., Petkau, J. and White, R. (1997). Longitudinal analyses for magnetic resonance imaging outcomes in multiple sclerosis clinical trials, *Journal of Biopharmaceutical Statistics*, **7**, 501-531.

Hastie, T.J. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall, London.

HEYDE, C.C. (1997). *Quasi-likelihood and Its Application*, Springer, New York.

MCCULLAGH, P. (1980). Regression models for ordinal data (with discussion), *J. Roy. Statist. Soc. B.*, **42**, 109-142.

MCCULLAGH, P. and NELDER, J.A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

WESTERN, B. (1996). Vague theory and model uncertainty in Macrosociology, *Sociological Methodology*, **26**, 165-192.

BERTRAND CLARKE
DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, B.C. V6T 1Z2
CANADA
E-mail: riffraff@stat.ubc.ca

PETER X.-K. SONG
DEPARTMENT OF MATHEMATICS AND STATISTICS
YORK UNIVERSITY
TORONTO, ONTARIO M3J 1P3
CANADA
E-mail: song@mathstat.yorku.ca

Paper received: May 2003; revised June 2004.