

# Score and Wald Tests for Antedependence in Categorical Longitudinal Data

Yunlong Xie<sup>1</sup> and Dale L Zimmerman<sup>2\*</sup>

<sup>1</sup>Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, 6100 Executive Boulevard, Bethesda, MD 20892, USA

<sup>2</sup>Department of Statistics and Actuarial Science, 241 Schaeffer Hall, University of Iowa, Iowa City, IA 52242, USA

## Abstract

Three standard hypothesis testing procedures exist based on likelihood functions: the likelihood ratio, score, and Wald tests. For unstructured antedependence models for categorical longitudinal data, Xie and Zimmerman derived the likelihood ratio test for the order of antedependence as well as likelihood ratio tests for time-invariance of transition probabilities and strict stationarity. In this article, we derive score tests (of Pearson's chi-square form) and Wald tests for all the same purposes. Via simulation, we show that for testing for order of antedependence, a modified likelihood ratio test performs best if the sample is of size 50 or smaller, but otherwise the score test is superior. The Wald test is markedly inferior to both. We also show that the likelihood ratio and score tests for time-invariant transition probabilities and strict stationarity perform about equally well. The methods are applied to data from a longitudinal study of labor force participation of married women, indicating that these data are third-order antedependent with time-invariant transition probabilities of this order.

**Keywords:** Score test; Pearson's chi-square; Markov models; Transition models; Wald test

## Introduction

Three standard ways exist for using the likelihood function to perform hypothesis testing: the likelihood ratio test, score test and Wald test. Xie and Zimmerman [1] gave a thorough presentation of antedependent (AD) multinomial models for categorical longitudinal data and, extending much earlier work of Anderson and Goodman [2], developed maximum likelihood estimators of transition probabilities and testing procedures associated with fitting AD models to categorical longitudinal data with missingness. More specifically, they developed likelihood ratio tests for order of antedependence, time-invariance of transition probabilities, and strict stationarity. Due to the oversensitivity of the likelihood ratio test for order of antedependence (i.e. Type I error rates that are significantly larger than nominal levels), they used Lawley [3] strategy of multiplying the likelihood ratio test statistic by a scale factor to obtain a modified likelihood ratio test statistic with Type I error rate closer to its nominal level. However, this modification requires tedious computation. Moreover, Xie and Zimmerman [1] did not discuss how the performance of the likelihood ratio test (modified or unmodified) compares to that of either the score test or Wald test.

The purpose of this article is to make comparisons among the score, Wald, likelihood ratio, and modified likelihood ratio tests for antedependence and related properties. We will show that the score test is superior to the Wald and likelihood ratio tests with respect to computational simplicity. Moreover, we will show that the score test outperforms the others in terms of size and power.

The remainder of the article is organized as follows. In Section 2, we review unstructured AD multinomial models (of variable and constant order) and maximum likelihood estimation of the parameters of such models. In Section 3, we derive score tests for order of antedependence, time-invariance of transition probabilities, and strict stationarity. We also derive Wald tests for these same purposes, but in somewhat less generality due to the complexity of notation in AD multinomial models. In Section 4, we compare the operating characteristics of the score, Wald, and likelihood ratio tests (both the regular and the modified version) based on data simulated from three AD multinomial models.

In Section 5, we use the testing procedures developed in Section 3 to analyze data from a longitudinal study of labor force participation of married women. Section 6 is a brief summary.

## Antedependence Models for Categorical Longitudinal Data

Time index-ordered random variables  $Y_1, \dots, Y_n$  are said to be antedependent (AD) of order  $(p_1, p_2, \dots, p_k)$  if the  $k$ th variable, conditioned on the  $p_k$  immediately preceding variables, is independent of all further preceding variables [4]. In the special case for which  $p_k = \min(k-1, p)$  for all  $k$  and for some integer  $p$  (with  $0 \leq p \leq n-1$ ), we say alternatively that the variables are  $p$ th-order antedependent.

## Notation and parameterizations

Consider a longitudinal setting, with a fixed number  $N$  of subjects. Let  $Y_i \equiv (Y_{i1}, \dots, Y_{in})'$  denote the vector of values of a categorical characteristic at the  $n$  scheduled observation times for the  $i$ th subject, some of which may not be observed (if a portion of that subject's data is missing). Let  $1, \dots, c \geq 2$  denote the characteristic's categories. Let  $Y_k$  denote the characteristic's value at time point  $k$  for a generic subject. For each possible outcome  $(y_1, \dots, y_n)$  — corresponding to a cell in an  $n$ -dimensional contingency table — let  $\pi_{y_1 \dots y_n} \equiv P(Y_1 = y_1, \dots, Y_n = y_n)$  denote the true cell probability with corresponding cell count  $N_{y_1 \dots y_n}$  (which may or may not be observed), and put  $\pi = (\pi_{y_1 \dots y_n})$ . Accordingly,

$$N = \sum_{(y_1, \dots, y_n) \in C_n} N_{y_1 \dots y_n}, \text{ where } C_n \equiv \{1, \dots, c\}^n \text{ is the set of all } c^n \text{ possible}$$

**\*Corresponding author:** Dale L. Zimmerman, Department of Statistics and Actuarial Science, 241 Schaeffer Hall, University of Iowa, Iowa City, IA 52242, USA, Tel: 319-335-0818; E-mail: [dale-zimmerman@uiowa.edu](mailto:dale-zimmerman@uiowa.edu)

**Received** February 19, 2014; **Accepted** March 27, 2014; **Published** March 31, 2014

**Citation:** Xie Y, Zimmerman DL (2014) Score and Wald Tests for Antedependence in Categorical Longitudinal Data. J Biomet Biostat 5: 188. doi:10.472/2155-6180.1000188

**Copyright:** © 2014 Xie Y, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

outcomes. We assume that the  $Y_i$ 's are independently and identically distributed as Multinomial  $(1, \pi)$  and that covariates, apart possibly from indicator variables for groups, are either unavailable or not used in the analysis.

In this article, we will derive score and Wald tests for various properties of AD models. The score test, like the likelihood ratio test, is well known to be invariant to 1:1 transformations of the parameter space. It is therefore convenient, since antedependence is defined in terms of certain conditional independencies, to parameterize in terms of certain conditional probabilities. Define  $\pi_{y_k|y_1 \dots y_{k-1}} \equiv P(Y_k = y_k | Y_1 = y_1, \dots, Y_{k-1} = y_{k-1})$  for  $k=2, \dots, n$  and  $(y_1, \dots, y_k) \in C_k$ . Under an  $AD(p_1, \dots, p_n)$  model, we may represent the common values of  $\{\pi_{y_k|y_1 \dots y_{k-p_k-1} y_{k-p_k} \dots y_{k-1}} : (y_1, \dots, y_{k-p_k-1}) \in C_{k-p_k-1}\}$  by transition probability parameter  $\pi_{y_k|+ \dots + y_{k-p_k} \dots y_{k-1}}$ . (Here and subsequently, we indicate summation over a subscripted index by replacing that index with a "+"). Thus, the  $AD(p_1, \dots, p_n)$  model may be parameterized by the nonredundant set of parameters

$$\Theta^{(p_1, \dots, p_n)} \equiv \left( \bigcup_{k \geq p_k=0} \left\{ \pi_{y_k|+ \dots + y_{k-p_k} \dots y_{k-1}} : y_k = 1, \dots, c-1 \right\} \right) \cup \left( \bigcup_{k \geq p_k \geq 1} \left\{ \pi_{y_k|+ \dots + y_{k-p_k} \dots y_{k-1}} : y_k = 1, \dots, c-1; (y_{k-p_k}, \dots, y_{k-1}) \in C_{p_k} \right\} \right).$$

Note that  $\dim(\Theta^{(p_1, \dots, p_n)}) = (c-1) \sum_{k=1}^n C_{p_k}$ .

In contrast to the likelihood ratio and score tests, the Wald test is not invariant to parameter transformations. It is inconvenient to express all nonredundant equations implied by equivalence among  $\{\pi_{y_k|y_1 \dots y_{k-p_k-1} y_{k-p_k} \dots y_{k-1}} : (y_1, \dots, y_{k-p_k-1}) \in C_{k-p_k-1}\}$  in terms of  $\pi$  or in a well-organized way to perform a Wald test. Alternatively, Agresti [5] suggests parameterizing in terms of log odds ratios, since convergence of the asymptotic distribution of the estimated log odds ratios to normality is rapid. The natural analogue here is to parameterize in terms of conditional log odds ratios, or COLORS, where the conditioning is on all possible realizations of all intervening variables. More specifically, we define COLORS of lag two and higher as follows:

$$\begin{aligned} \psi_{y_k-h, y_k|y_{k-h+1} \dots y_{k-1}} &\equiv \log \left( \frac{P(Y_{k-h} = c, Y_{k-h+1} = y_{k-h+1}, \dots, Y_{k-1} = y_{k-1}, Y_k = c)}{P(Y_{k-h} = c, Y_{k-h+1} = y_{k-h+1}, \dots, Y_{k-1} = y_{k-1}, Y_k = y_k)} \right) \\ &\times \frac{P(Y_{k-h} = y_{k-h}, Y_{k-h+1} = y_{k-h+1}, \dots, Y_{k-1} = y_{k-1}, Y_k = y_k)}{P(Y_{k-h} = y_{k-h}, Y_{k-h+1} = y_{k-h+1}, \dots, Y_{k-1} = y_{k-1}, Y_k = c)} \\ &= \log \pi_{y_k-h+1 \dots y_{k-1} c} + \log \pi_{y_k-h+1 \dots y_{k-1} y_k} \\ &- \log \pi_{y_k-h+1 \dots y_{k-1} y_k} - \log \pi_{y_k-h+1 \dots y_{k-1} c} \end{aligned} \quad (1)$$

where  $(y_{k-h}, y_k) \in \{1, \dots, c-1\}^2$ ,  $(y_{k-h+1}, \dots, y_{k-1}) \in C_{h-1}$ ,  $k=3, 4, \dots, n$ , and  $h=2, 3, \dots, k-1$ . In addition, we define lag-one COLORS (for which there are no intervening variables) by an expression identical to (1) but with  $h=1$ ; thus, lag-one COLORS coincide with lag-one log odds ratios.

## Structured antedependence models and maximum likelihood estimators

Antedependence (of specified order) is said to be *structured* if additional restrictions (such as stationarity) are imposed; otherwise it is *unstructured*. Two types of structure we consider here are time-invariance of  $p$ th-order transition probabilities and strict stationarity,

defined as follows. If  $p \geq 1$  and  $\pi_{y_{p+1}|y_1^{(k-p)} \dots y_p^{(k-1)}}$  denotes  $P(Y_k = y_{p+1} | Y_{k-p} = y_1, \dots, Y_{k-1} = y_p)$  for  $k=p+1, \dots, n$ , the  $p$ th-order transition probabilities are said to be time-invariant if

$$\pi_{y_{p+1}|y_1^{(1)} \dots y_p^{(p)}} = \pi_{y_{p+1}|y_1^{(2)} \dots y_p^{(p+1)}} = \dots = \pi_{y_{p+1}|y_1^{(n-p)} \dots y_p^{(n-1)}} \quad \text{for all } (y_1, \dots, y_{p+1}) \in C_{p+1} \quad (2)$$

Variables  $Y_1, \dots, Y_n$  are said to be strictly stationary if the joint probabilities of all events are invariant to time shifts. Clearly, strict stationarity implies (2), but not vice versa. For situations with either no missing data or ignorably monotone missing data, Xie and Zimmerman [1] derived closed-form expressions for the maximum likelihood estimators of the parameters of an unstructured  $AD(p_1, \dots, p_n)$  process under the multinomial sampling framework described previously. They also derived maximum likelihood estimators (in closed form) for the parameters of an  $AD(p)$  model with time-invariant  $p$ th-order transition probabilities, and for the parameters of a strictly stationary  $AD(p)$  model. Furthermore, for situations with arbitrarily ignorable missing data, they devised a restricted EM algorithm for computing maximum likelihood estimates. One or more of these estimators appears in expressions for the score tests to be described in the next section.

## Hypothesis tests

**Order of antedependence:** Score test Consider testing  $AD(p_1, \dots, p_n)$  as a null hypothesis against  $AD(q_1, \dots, q_n)$  as the alternative hypothesis, where  $p_k \leq q_k$  for all  $k$  and the inequality is strict for at least one  $k$ . Now, for hypothesis testing in a multinomial sampling context, it is well known that if the alternative hypothesis is saturated, then the score and Pearson's chi-square test statistics for any testable null hypothesis coincide. In the present context the alternative hypothesis need not be saturated [i.e.  $AD(n-1)$ ]; nevertheless the likelihood under the alternative hypothesis is the product of kernels of conditionally independent saturated  $c$ -nomial distributions, whose parameters' only constraints are that they lie in  $[0, 1]$  and sum to one [1]. Furthermore, the parameters in these distributions are all distinct. As a result, the information matrix corresponding to this likelihood is block diagonal [each block on the diagonal being of dimensions  $(c-1) \times (c-1)$ ], and the score test statistic for any testable null hypothesis is merely the summation of the Pearson's chi-square statistics corresponding to the individual  $c$ -nomial distributions. By similar arguments, the score tests for all the other purposes presented in this article also have Pearson's chi-square form. Now, the Pearson's chi-square statistics equal zero for those  $k$  for which  $p_k = q_k$ . Let  $C_{u,k}^+ \equiv \{(y_u, y_{u+1}, \dots, y_k) \in C_{k-u+1} : N_{y_u, \dots, y_k} > 0\}$  and write  $C_k^+$  for  $C_{k,k}^+$ . Therefore the score test statistic for testing for  $AD(p_1, \dots, p_n)$  versus  $AD(q_1, \dots, q_n)$  simplifies to

$$X_{p,q}^2 \equiv \sum_{k=2}^n \left( I(q_k > p_k) \sum_{(y_k-q_k, \dots, y_1) \in C_{q_k-k+1}^+} \left( \frac{N_{y_k-q_k, \dots, y_1} - N_{y_k-q_k, \dots, y_1} \hat{\pi}_{y_k-q_k, \dots, y_1}^{(p_1, \dots, p_n)}}{N_{y_k-q_k, \dots, y_1} + \dots + \hat{\pi}_{y_k-q_k, \dots, y_1}^{(p_1, \dots, p_n)}} \right)^2 \right)$$

Where

$$\delta_{y_k|+ \dots + y_{k-q_k} \dots y_{k-1}}^{(p_1, \dots, p_n)} = I(p_k = 0) \delta_{y_k|+ \dots + y_{k-1}}^{(p_1, \dots, p_n)} + I(p_k \geq 0) \delta_{y_k|+ \dots + y_{k-p_k} \dots y_{k-1}}^{(p_1, \dots, p_n)}.$$

The limiting null distribution of  $X_{p,q}^2$  is chi-square with degrees of freedom equal to  $d(p, q) \equiv \dim(\Theta^{(q_1, \dots, q_n)}) - \dim(\Theta^{(p_1, \dots, p_n)}) = (c-1) \sum_{k=2}^n (C_{p_k} - C_{q_k})$ . Thus, the approximate size- $\alpha$  score test for  $AD(p_1, \dots, p_n)$  versus  $AD(q_1, \dots, q_n)$

rejects the null hypothesis if  $X_{p,q}^2$  exceeds the  $100(1-\alpha)$ th percentile of the aforementioned chi-square distribution.

For the important special case of testing AD(p) versus AD(q), where  $0 \leq p < q \leq n-1$ , the score test statistic simplifies to

$$X_{p,q}^2 \equiv \sum_{j=p+1}^q \sum_{(y_1, \dots, y_j) \in C_{1,j}^+} \frac{\left( N_{y_1, \dots, y_j, +, \dots, +} - N_{y_1, \dots, y_{j-1}, +, \dots, +} \hat{\pi}_{y_1, \dots, y_{j-1}}^{(p)} \right)^2}{N_{y_1, \dots, y_{j-1}, +, \dots, +} \hat{\pi}_{y_1, \dots, y_{j-1}}^{(p)}} + \sum_{k=q+1}^n \sum_{(y_{k-q}, \dots, y_k) \in C_{k-q,k}^+} \frac{\left( N_{+, \dots, +, y_{k-q}, \dots, y_k, +, \dots, +} - N_{+, \dots, +, y_{k-q}, \dots, y_{k-1}, +, \dots, +} \hat{\pi}_{y_{k-q}, \dots, y_{k-1}}^{(p)} \right)^2}{N_{+, \dots, +, y_{k-q}, \dots, y_{k-1}, +, \dots, +} \hat{\pi}_{y_{k-q}, \dots, y_{k-1}}^{(p)}}.$$

Where

$$\hat{\pi}_{y_j | y_1, \dots, y_{j-1}}^{(p)} = I(p=0) \hat{\pi}_{+, \dots, +, y_j, +, \dots, +}^{(0)} + I(p \geq 1) \hat{\pi}_{y_j | +, \dots, +, y_{j-p}, \dots, y_{j-1}}^{(p)} \quad \text{and} \\ \hat{\pi}_{y_k | +, \dots, +, y_{k-q}, \dots, y_{k-1}}^{(p)} = I(p=0) \hat{\pi}_{+, \dots, +, y_k, +, \dots, +}^{(0)} + I(p \geq 1) \hat{\pi}_{y_k | +, \dots, +, y_{k-p}, \dots, y_{k-1}}^{(p)}.$$

The degrees of freedom for the corresponding limiting null chi-square distribution are  $d(p, q) \equiv (n-p-1)c^p - (n-p)c^{p+1} - (n-q-1)c^q + (n-q)c^{q+1}$ .

If some data are (ignorably) missing, the score test statistics may be generalized to expressions identical to  $X_{p,q}^2$  and  $X_{p,q}^2$ , except that the cell counts are replaced with maximum likelihood estimates obtained by carrying out a restricted EM algorithm under AD( $q_1, \dots, q_n$ ) and the maximum likelihood estimates of transition probabilities are obtained by carrying out a restricted EM algorithm under AD( $p_1, \dots, p_n$ ). Similar generalizations to accommodate missing data may be made to all score tests presented subsequently.

## Wald test

For simplicity, we present a Wald test for testing  $p$ th-order antedependence against the saturated model [AD( $n-1$ )]. Wald tests for variable-order antedependence can be derived easily by extending results from this section. Let  $\Psi \equiv (\psi_{y_k-h, y_k | y_{k-h+1}, \dots, y_{k-1}})$  denote the vector of COLORS listed in order from no intervenors (lag one) to  $n-2$  intervenors (lag  $n-1$ ).

Under the saturated model, the maximum likelihood estimator of  $\Psi_{y_k-h, y_k | y_{k-h+1}, \dots, y_{k-1}}$  is given by

$$\hat{\psi}_{y_k-h, y_k | y_{k-h+1}, \dots, y_{k-1}} = \log(N_{+, \dots, +, y_k | y_{k-h+1}, \dots, y_{k-1}, +, \dots, +}) + \log(N_{+, \dots, +, y_k-h, y_k | y_{k-h+1}, \dots, y_{k-1}, +, \dots, +}) - \log(N_{+, \dots, +, y_k-h+1, \dots, y_{k-1}, +, \dots, +}) - \log(N_{+, \dots, +, y_k-h, y_k | y_{k-h+1}, \dots, y_{k-1}, +, \dots, +})$$

Let  $\hat{\Psi} \equiv (\hat{\psi}_{y_k-h, y_k | y_{k-h+1}, \dots, y_{k-1}})$ , with ordering of elements identical to that of  $\Psi$ .

**Lemma 3.1.** As  $N \rightarrow \infty$ ,  $\sqrt{N}(\hat{\Psi} - \Psi)$  converges in distribution to  $N(0, \Sigma)$ , where

$$\Sigma = B[\text{diag}(A\pi)]^{-1} A[\text{diag}(\pi) - \pi\pi'] A'[\text{diag}(A\pi)]^{-1} B',$$

and matrices A and B are defined in the proof in Section 7.1.

In Section 7.2, we create a toy example to illustrate the Wald test for AD(p) against the saturated model.

By Lemma 3.1, we obtain the limiting distribution of all conditional log odds ratios. Under an AD(p) model (with  $p < n-1$ ), all COLORS of

lag  $p+1$  and higher are equal to zero, as is easily verified. Consequently,  $p$ th-order antedependence is equivalent to the condition  $M_p \Psi = 0$ , where

$$M_p \equiv \begin{bmatrix} 0 & d_p \times (d_0 - d_p) \\ I_{d_p \times d_p} \end{bmatrix}.$$

0 and I are null and identity matrices respectively, and

$$d_p = (c-1)^2 \sum_{k=p}^{n-2} (n-1-k)c^k \quad \text{or, equivalently,}$$

$$d_p = c^n - 1 - \dim(\Theta^{(p_1, \dots, p_n)}) = c^n - (n-p)c^{p+1} + (n-1-p)c^p.$$

Then a Wald test statistic is given by

$$T_p^2 \equiv N(M_p \hat{\Psi})' (M_p \hat{\Sigma} M_p')^{-1} (M_p \hat{\Psi})$$

It follows from Lemma 3.1 and Slutsky's Theorem that the limiting null distribution of  $T_p^2$  is chi-square with  $d_p$  degrees of freedom. However, the Wald test is affected by empty cells. If any cell count is zero, 0.5 may be added to each cell count to avoid zero denominators in the maximum likelihood estimates of COLORS.

## Time-invariant transition probabilities

**Score test:** Xie and Zimmerman [1] derived the likelihood ratio test for time-invariant  $p$ th-order transition probabilities. For this testing purpose, the score statistic can be expressed as

$$X_t^2 = \sum_{(y_1, \dots, y_{p+1})} \sum_k \frac{\left( N_{+, \dots, +, y_1^{(k-p)}, \dots, y_p^{(k)} | +, \dots, +} - N_{+, \dots, +, y_1^{(k-p)}, \dots, y_p^{(k-1)} | +, \dots, +} \hat{\pi}_{y_1, \dots, y_p}^{(p)} \right)^2}{N_{+, \dots, +, y_1^{(k-p)}, \dots, y_p^{(k-1)} | +, \dots, +} \hat{\pi}_{y_1, \dots, y_p}^{(p)}}$$

where  $\hat{\pi}_{y_1, \dots, y_p}^{(p)}$  is the maximum likelihood estimator of the common

$p$ th-order transition probability [see Theorem 2 of Xie and Zimmerman [1] for an expression for this estimator] and the summation is over those  $(y_1, \dots, y_{p+1}) \in C_{p+1}$  and  $k=p+1, \dots, n$  for which the denominator is positive. The limiting null distribution of each test statistic is chi-square, with degrees of freedom equal to the number of nonredundant conditions in (2), or  $(c-1)[n-(p+1)]c^p$ .

**Wald test:** Due to the complexity of the general situation, we illustrate the derivation of a Wald test for time-invariant transition probabilities for the special case of an AD(1) model, with  $c=2$  and  $n=3$ . A Wald test for situations with other  $p$ ,  $c$ , and  $n$  can be obtained in similar fashion.

Using notation from (2), the null hypothesis of time-invariant transition probabilities is  $H_0: \pi_{1(2)|1(1)} = \pi_{1(3)|1(2)} = \pi_{1(2)|2(1)} = \pi_{1(3)|2(2)}$ .

Maximum likelihood estimators of the parameters appearing in  $H_0$  are as follows:

$$\hat{\pi}_{1(2)|1(1)} \equiv \frac{N_{11+}}{N_{1++}}, \quad \hat{\pi}_{1(3)|1(2)} \equiv \frac{N_{11+}}{N_{1++}}, \quad \hat{\pi}_{1(2)|2(1)} \equiv \frac{N_{21+}}{N_{2++}}, \quad \hat{\pi}_{1(3)|2(2)} \equiv \frac{N_{21+}}{N_{2++}},$$

and these are all asymptotically independent with limiting normal distributions

$$N \left( \pi_{1(2)|1(1)}, \frac{\pi_{1(2)|1(1)}(1-\pi_{1(2)|1(1)})}{N_{1++}} \right), N \left( \pi_{1(3)|1(2)}, \frac{\pi_{1(3)|1(2)}(1-\pi_{1(3)|1(2)})}{N_{1++}} \right), \\ N \left( \pi_{1(2)|2(1)}, \frac{\pi_{1(2)|2(1)}(1-\pi_{1(2)|2(1)})}{N_{2++}} \right), N \left( \pi_{1(3)|2(2)}, \frac{\pi_{1(3)|2(2)}(1-\pi_{1(3)|2(2)})}{N_{2++}} \right),$$

respectively. Thus, a Wald test statistic is given by

$$T_t^2 \equiv \begin{bmatrix} \hat{\pi}_{1(2)|1(1)} - \hat{\pi}_{1(3)|1(2)} \\ \hat{\pi}_{1(2)|2(1)} - \hat{\pi}_{1(3)|2(2)} \end{bmatrix} \begin{bmatrix} \hat{\sigma}_{11}^2 & 0 \\ 0 & \hat{\sigma}_{12}^2 \end{bmatrix} \begin{bmatrix} \hat{\pi}_{1(2)|1(1)} - \hat{\pi}_{1(3)|1(2)} \\ \hat{\pi}_{1(2)|2(1)} - \hat{\pi}_{1(3)|2(2)} \end{bmatrix}$$

Where

$$\hat{\sigma}_{1j}^2 = \frac{\hat{\pi}_{1(2)|2(1)}(1 - \hat{\pi}_{1(2)|2(1)})}{N_{j++}} + \frac{\hat{\pi}_{1(3)|2(2)}(1 - \hat{\pi}_{1(3)|2(2)})}{N_{j++}} \quad j = 1, 2$$

A Wald test may be carried out by comparing  $T_t^2$  to a chosen percentile of the chi-square distribution with two degrees of freedom.

### Strict stationarity

**Score test:** Maximum likelihood estimators of cell probabilities of a strictly stationary AD(p) model cannot be expressed in closed form, but they can be obtained numerically using Lang [6] algorithm. Let  $\hat{\pi}^{s,(p)}$  be the vector of estimates so obtained. Then, the score statistic for testing for strict stationarity under an unstructured AD(p) model is

$$X_s^2 = \sum_{(y_1 \dots y_p)} \frac{(N_{y_1 \dots y_p} + \dots - N \hat{\pi}_{y_1 \dots y_p}^{s,(p)})^2}{N \hat{\pi}_{y_1 \dots y_p}^{s,(p)}} \\ \sum_{(y_1 \dots y_p)} \sum_{k=p+1}^n \frac{(N_{y_1 \dots y_{p+1}}^{(k-p)} \dots y_{p+1}^{(k)} + \dots - N_{y_1 \dots y_p}^{(k-1)} + \dots + \hat{\pi}_{y_1 \dots y_p}^{s,(p)} \dots y_p^{(k-p)})^2}{N_{y_1 \dots y_p}^{(k-p)} \dots y_p^{(k-1)} + \dots + \hat{\pi}_{y_1 \dots y_p}^{s,(p)} \dots y_p^{(k-1)}},$$

respectively, where the summations are taken over those terms for which the denominators are positive. The limiting null distribution of each test statistic is chi-square with  $(c-1)[n-(p+1)]c^p + (c^p-1)$  degrees of freedom.

**Wald test:** As we did for the Wald test for time-invariant transition probabilities, we present a Wald test for strict stationarity only for the special case of an AD(1) with  $c=2$  and  $n=3$ . By Lemma 1 of Xie and Zimmerman [1], strict stationarity imposes one additional restriction [under AD(1)] beyond those imposed by time-invariant transition probabilities, namely  $h(\pi_{1++}, \pi_{1(2)|1(1)}, \pi_{1(2)|2(1)}) \equiv \pi_{1++} - \pi_{1(2)|1(1)}\pi_1 - \pi_{1(2)|2(1)}(1 - \pi_{1++}) = 0$ . Thus, the null hypothesis of strict stationarity is

$$H_0: \pi_{1(2)|1(1)} = \pi_{1(3)|1(2)}, \pi_{1(2)|2(1)} = \pi_{1(3)|2(2)}, h(\pi_{1++}, \pi_{1(2)|1(1)}, \pi_{1(2)|2(1)}) = 0$$

Let  $\theta = (\pi_{1++}, \pi_{1(2)|1(1)}, \pi_{1(2)|2(1)}, \pi_{1(3)|1(2)}, \pi_{1(3)|2(2)})'$ . Maximum likelihood estimators of all elements of  $\theta$  but the first were given in section 3.2.2; the maximum likelihood estimator of the first element is  $\hat{\pi}_{1++} = \frac{N_{1++}}{N}$  which has limiting normal distribution  $N(\pi_{1++}, \frac{\pi_{1++}(1-\pi_{1++})}{N})$ . Furthermore, the maximum likelihood estimators of the five components of  $\theta$  are asymptotically independent.

Note that  $\frac{\partial h}{\partial \theta} = (1 - \pi_{1(2)|1(1)} + \pi_{1(2)|2(1)}, -\pi_{1++}, 0, \pi_{1++}, 0)'$ . Thus, by the delta method, the limiting distribution of  $h(\hat{\pi}_{1++}, \hat{\pi}_{1(2)|1(1)}, \hat{\pi}_{1(2)|2(1)})$   $N(h(\pi_{1++}, \pi_{1(2)|1(1)}, \pi_{1(2)|2(1)}), \sigma_h^2)$  is where

$$\sigma_h^2 = (1 - \pi_{1(2)|1(1)} + \pi_{1(2)|2(1)})^2 \frac{\pi_{1++}(1 - \pi_{1++})}{N} + \pi_{1++}^2 \frac{\pi_{1(2)|1(1)}(1 - \pi_{1(2)|1(1)})}{N_{1++}} \\ + \pi_{1++}^2 \frac{\pi_{1(2)|2(1)}(1 - \pi_{1(2)|2(1)})}{N_{2++}}$$

Thus, a Wald test statistic for strict stationarity is given by

$$T_s^2 = \begin{bmatrix} \hat{\pi}_{1(2)|1(1)} - \hat{\pi}_{1(3)|1(2)} \\ \hat{\pi}_{1(2)|2(1)} - \hat{\pi}_{1(3)|2(2)} \\ h(\hat{\pi}_{1++}, \hat{\pi}_{1(2)|1(1)}, \hat{\pi}_{1(2)|2(1)}) \end{bmatrix} \begin{bmatrix} \hat{\sigma}_{11}^2 & 0 & \hat{\sigma}_{h,1|1} \\ 0 & \hat{\sigma}_{12}^2 & \hat{\sigma}_{h,1|2} \\ \hat{\sigma}_{h,1|1} & \hat{\sigma}_{h,1|2} & \hat{\sigma}_h^2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\pi}_{1(2)|1(1)} - \hat{\pi}_{1(3)|1(2)} \\ \hat{\pi}_{1(2)|2(1)} - \hat{\pi}_{1(3)|2(2)} \\ h(\hat{\pi}_{1++}, \hat{\pi}_{1(2)|1(1)}, \hat{\pi}_{1(2)|2(1)}) \end{bmatrix}$$

where  $\hat{\sigma}_{h,1|1}$  and  $\hat{\sigma}_{h,1|2}$  are estimators of

$$\sigma_{h,1|1} \equiv \lim_{N \rightarrow \infty} \text{Cov}(h(\hat{\pi}_{1++}, \hat{\pi}_{1(2)|1(1)}, \hat{\pi}_{1(2)|2(1)}), \hat{\pi}_{1(2)|1(1)} - \hat{\pi}_{1(3)|1(2)}) = -\pi_{1++} \frac{\pi_{1(2)|1(1)}(1 - \pi_{1(2)|1(1)})}{N_{1++}}$$

$$\text{and } \sigma_{h,1|2} \equiv \lim_{N \rightarrow \infty} \text{Cov}(h(\hat{\pi}_{1++}, \hat{\pi}_{1(2)|1(1)}, \hat{\pi}_{1(2)|2(1)}), \hat{\pi}_{1(2)|2(1)} - \hat{\pi}_{1(3)|2(2)}) = -(1 - \pi_{1++}) \frac{\pi_{1(2)|2(1)}(1 - \pi_{1(2)|2(1)})}{N_{2++}},$$

respectively, obtained by replacing the parameters with their maximum likelihood estimates. A Wald test may be carried out by comparing  $T_s^2$  to a chosen percentile of the chi-square distribution with three degrees of freedom.

### Simulation studies

**Tests for order of antedependence:** In this section, we report results of several simulation studies of the tests developed in earlier sections, beginning with tests for order of antedependence. Data were simulated from two binary linear processes observed at four time points. The first process is defined as follows:

$$Y_1 = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ 2 & \text{w.p. } \frac{1}{2} \end{cases}, \quad Y_2 | Y_1 = \begin{cases} Y & \text{w.p. } \frac{1}{3} \\ 1 & \text{w.p. } \frac{1}{3} \\ 2 & \text{w.p. } \frac{1}{3} \end{cases} \quad (3)$$

$$Y_3 | (Y_1, Y_2) = \begin{cases} Y_1 & \text{w.p. } \frac{\theta}{4} \\ Y_2 & \text{w.p. } \frac{2-\theta}{8} \\ 1 & \text{w.p. } \frac{1}{4} \\ 2 & \text{w.p. } \frac{1}{4} \end{cases}, \quad Y_4 | (Y_1, Y_2, Y_3) = \begin{cases} Y_1 & \text{w.p. } \frac{\theta}{8} \\ Y_2 & \text{w.p. } \frac{\theta}{8} \\ Y_3 & \text{w.p. } \frac{2-\theta}{4} \\ 1 & \text{w.p. } \frac{1}{4} \\ 2 & \text{w.p. } \frac{1}{4} \end{cases} \quad (4)$$

Here  $\theta \in [0, 2]$  controls the degree of departure from first-order antedependence: when  $\theta=0$  the process is AD(1), when  $\theta>0$  the process is AD(3), and as  $\theta$  increases, the departure from AD(1) is larger. The second process differs from the first only by redefining  $Y_3$  and  $Y_4$  as follows:

$$Y_3 | (Y_1, Y_2) = \begin{cases} Y_1 & \text{w.p. } \frac{1}{4} \\ Y_2 & \text{w.p. } \frac{1}{4} \\ 1 & \text{w.p. } \frac{1}{4} \\ 2 & \text{w.p. } \frac{1}{4} \end{cases}, \quad Y_4 | (Y_1, Y_2, Y_3) = \begin{cases} Y_1 & \text{w.p. } \frac{\theta}{4} \\ Y_2 & \text{w.p. } \frac{2-\theta}{8} \\ Y_3 & \text{w.p. } \frac{2-\theta}{8} \\ 1 & \text{w.p. } \frac{1}{4} \\ 2 & \text{w.p. } \frac{1}{4} \end{cases} \quad (5)$$



This difference renders the process AD(2) [rather than AD(1)] when  $\theta=0$ .

For each process, each of four sample sizes ( $N = 50, 100, 200, 1000$ ), and each  $\theta$  over a range of values, 10,000 samples were generated. Rejection rates for (nominal) size-0.05 score, Wald, likelihood ratio (LRT) and modified likelihood ratio (MLRT) tests of AD(1) versus AD(3) for (3) and (4), and of AD(2) versus AD(3) for (3) and (5), are listed in Table 1. Rejection rates for testing (3) and (4) are plotted in Figure 1, while those for testing

(3) and (5) are similar hence not shown. These results indicate that:

1. Regardless of the specific test statistic, rejection rates for testing AD(1) versus AD(3) are higher than those for testing AD(2) versus AD(3), as expected.

2. The Wald test is the least powerful and its test statistic converges in distribution to  $\chi^2(8)$  relatively slowly. As the sample size increases, rejection rates for the score test, LRT and MLRT become very similar to each other.

3. When the sample size is small ( $N=50$ ), rejection rates are discernibly higher for the MLRT than for the score test. In particular, the MLRT makes more Type I errors, which is the cost it pays for having higher rejection rates than the score test at nonzero values of  $\theta$ .

4. When the sample size is moderate ( $N=100$  and  $N=200$ ), the power curves for score test and MLRT cross. The point of intersection is closer to the null ( $\theta=0$ ) when  $N=200$  than when  $N=100$ . Thus the larger the sample size, the better the relative performance of the score test to the MLRT.

We also investigated the effect of missing data on these results,

N	$\theta$	(3) and (4)				(3) and (5)			
		Score	LRT	MLRT	Wald	Score	LRT	MLRT	Wald
50	0	0.019	0.073	0.029	0.002	0.017	0.097	0.038	0.003
	0.2	0.029	0.1	0.047	0.004	0.023	0.111	0.044	0.003
	0.4	0.044	0.138	0.068	0.005	0.04	0.14	0.062	0.009
	0.6	0.076	0.202	0.113	0.01	0.063	0.183	0.087	0.014
	0.8	0.119	0.285	0.181	0.016	0.094	0.24	0.123	0.023
	1	0.187	0.396	0.268	0.026	0.147	0.322	0.185	0.047
100	0	0.024	0.078	0.04	0.01	0.035	0.081	0.046	0.013
	0.2	0.051	0.103	0.057	0.029	0.057	0.099	0.06	0.024
	0.4	0.113	0.164	0.099	0.074	0.108	0.146	0.098	0.055
	0.6	0.22	0.267	0.183	0.165	0.182	0.221	0.159	0.106
	0.8	0.39	0.434	0.328	0.32	0.291	0.334	0.25	0.188
	1	0.588	0.628	0.518	0.528	0.438	0.477	0.385	0.315
200	0	0.037	0.068	0.044	0.02	0.047	0.063	0.049	0.028
	0.2	0.087	0.102	0.073	0.045	0.087	0.091	0.073	0.058
	0.4	0.232	0.231	0.184	0.1	0.193	0.186	0.16	0.149
	0.6	0.476	0.47	0.406	0.208	0.36	0.351	0.31	0.304
	0.8	0.75	0.745	0.687	0.372	0.577	0.565	0.524	0.523
	1	0.924	0.924	0.895	0.586	0.78	0.772	0.738	0.741
1000	0	0.051	0.053	0.051	0.044	0.053	0.055	0.051	0.047
	0.2	0.293	0.277	0.272	0.161	0.209	0.2	0.196	0.205
	0.4	0.871	0.865	0.862	0.571	0.692	0.682	0.675	0.689
	0.6	0.999	0.999	0.999	0.927	0.973	0.97	0.968	0.973
	0.8	1	1	1	0.997	0.999	0.999	0.999	0.999
	1	1	1	1	1	1	1	1	1

**Table 1:** Empirical rejection rates of tests for order of antedependence for data simulated from processes (3) and (4) and (3) and (5). Empirical sizes (rejection rates when  $\theta=0$ ) more than two estimated standard errors from the nominal size (0.05) are set in bold type.

but to save space we do not list detailed results. The overall effect is a reduction in the power; for example, when 25% of the data is missing, the power of the score, likelihood ratio, and modified likelihood ratio tests is reduced by roughly 10-20% and the power of the Wald test is reduced by roughly 20-30%.

### Tests for time-invariant transition probabilities

For our second simulation study, we define binary AD(1) random variables  $Y_1, Y_2, Y_3$  and  $Y_4$  as follows:

$$Y_1 = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ 2 & \text{w.p. } \frac{1}{2} \end{cases}, P(Y_t = 1 | Y_{t-1} = 1) = \begin{cases} \frac{1}{2} + \frac{(t-1)\lambda}{10} & \text{if } Y_{t-1} = 1 \\ \frac{1}{3} + \frac{(t-1)\lambda}{10} & \text{if } Y_{t-1} = 2 \end{cases} \quad (6)$$

where  $0 \leq \lambda \leq 5/3$ . For each fixed  $(t, \lambda)$ ,  $P(Y_t = 1 | Y_{t-1} = 1) \neq P(Y_t = 1 | Y_{t-1} = 2)$ ; hence the variables are not AD(0) for any  $\lambda$ . When  $\lambda = 0$ , however, the transition probabilities are time-invariant; otherwise, they are not. Ten thousand realizations were simulated from (6) for each of several values of  $\lambda$ , and the likelihood ratio and score tests for time-invariant transition probabilities were performed.

Empirical rejection rates for the tests are listed in Table 2. These indicate that the likelihood ratio and score tests perform about equally well, with a very slight power advantage to the likelihood ratio test. When 25% of the data is missing, the power of both tests is reduced by roughly 15-20%.

### Tests for strict stationarity

Finally, consider binary AD(1) random variables defined as follows:

$$Y_1 = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ 2 & \text{w.p. } \frac{1}{2} \end{cases}, P(Y_t = 1 | Y_{t-1} = 1) = \frac{\gamma}{3} \text{ for } t = 2, 3, 4 \text{ and } 0.4 \leq \gamma \leq 1.8 \text{ for } t = 2, 3, 4 \text{ and } 0.4 \leq \gamma \leq 1.8 \quad (7)$$

3, 4 and  $0.4 \leq \gamma \leq 1.8$

Note that for arbitrary  $\gamma \neq 0$ ,  $P(Y_t = 1 | Y_{t-1} = 2)$  is always 50% larger than  $P(Y_t = 1 | Y_{t-1} = 1)$ . Clearly, process (7) has time-invariant transition probabilities. Furthermore,

$$P(Y_2 = 1) = P(Y_2 = 1 | Y_1 = 1)P(Y_1 = 1) + P(Y_2 = 1 | Y_1 = 2)P(Y_1 = 2) = \frac{5\gamma}{12}.$$

Thus, when  $\gamma = \frac{6}{5}$ ,  $P(Y_2 = 1) = \frac{1}{2} = P(Y_1 = 1)$ , which guarantees that process (7) is of strict stationarity by Lemma 1 of Xie and Zimmerman [1]; otherwise, it is not strictly stationary. Ten thousand realizations were simulated from (7) for each of several values of  $\gamma$  ranging between 0.4 and 1.8, and the likelihood ratio and score tests for strict stationarity were performed.

Empirical rejection rates are listed in Table 3. When  $\gamma$  is far away from 1.2 and the sample size is insufficiently large, the simulated data yield too many empty cells, with the consequence that strict stationarity cannot be tested by either the likelihood ratio test or score test due to zero expected cell counts. We use an "X" in Table 3 to indicate such cases. We observe that the likelihood ratio test slightly outperforms the score test for strict stationarity since the empirical size of the likelihood ratio test is not significantly different from the nominal size 0.05, except when the sample size is small, and the power of the likelihood ratio test is uniformly higher (by a small amount) than that of the score test. Since the likelihood ratio test performs well here, we do not develop on any modification to it as Xie and Zimmerman [1] did for determining the order of antedependence. As expected, the further  $\gamma$  is from 1.2, the higher the empirical rejection rate.

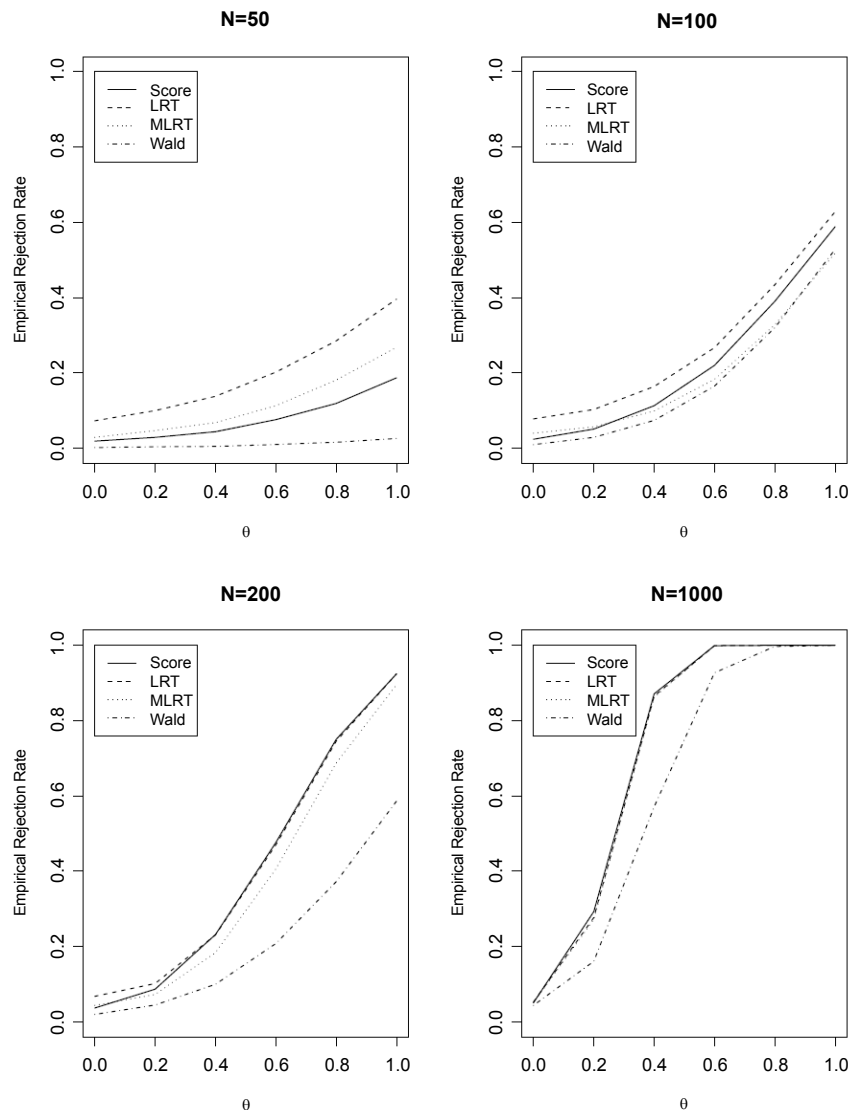


Figure 1: Empirical rejection rates for tests for AD(1) versus AD(3) under model (3) and (4), as a function of  $\theta$ .

$\lambda$	N=50		N=200		N=1000	
	LRT	Score	LRT	Score	LRT	Score
0	0.058	0.053	0.53	0.50	0.048	0.047
0.1	0.062	0.056	0.061	0.062	0.093	0.091
0.2	0.067	0.061	0.090	0.086	0.260	0.258
0.3	0.079	0.074	0.140	0.137	0.560	0.557
0.4	0.095	0.089	0.220	0.214	0.839	0.839
0.5	0.118	0.112	0.333	0.329	0.969	0.968
0.6	0.149	0.141	0.470	0.464	0.997	0.998
0.7	0.191	0.180	0.620	0.614	1	1
0.8	0.238	0.221	0.762	0.751	1	1
0.9	0.299	0.280	0.870	0.859	1	1
1.0	0.375	0.350	0.941	0.935	1	1

Table 2: Empirical rejection rates for tests of time-invariant transition probabilities for data simulated from process (6). Empirical sizes (rejection rates when  $\lambda=0$ ) more than two estimated standard errors from the nominal size (0.05) are set in bold type.

Example: Labor Force Data

Lindsey [7] provides data on annual employment status (employed or unemployed) from 1967 to 1971 for 1583 U.S. women from a cross-sectional survey. The data are displayed in a longitudinal multinomial format in Table 4, where  $Y_1$  through  $Y_5$  correspond to the years 1967 through 1971. No data are missing and only one cell is empty. The cell counts indicate that employment status is rather persistent over the time period: in particular, the outcomes “employed for all 5 years” and “unemployed for all 5 years” have relatively high proportions, 0.269 and 0.353 respectively. All remaining cells have relatively small proportions (less than 0.05).

The penalized log-likelihood model selection procedure proposed by Xie and Zimmerman [1] selects AD(0,1,2,3,3) [or equivalently AD(3)] when AIC is the criterion. Likelihood ratio, modified likelihood ratio, score, and Wald tests are all highly significant ( $P<10^{-8}$ ) for AD(0) against AD(1), AD(1) against AD(2), and AD(2) against AD(3), but

Y	N=50		N=200		N=1000	
	LRT	Score	LRT	Score	LRT	Score
0.4	X	X	X	X	1.000	1.000
0.6	X	X	0.999	0.999	1.000	1.000
0.8	X	X	0.845	0.839	1.000	1.000
1.0	0.096	0.089	0.243	0.236	0.903	0.903
1.2	0.058	0.054	0.051	0.050	0.049	0.050
1.4	0.099	0.089	0.234	0.224	0.877	0.877
1.6	X	X	0.772	0.765	1.000	1.000
1.8	X	X	X	X	1.000	1.000

**Table 3:** Empirical rejection rates for tests of strict stationarity for data simulated from process (7). Empirical sizes (rejection rates when  $\gamma=1.2$ ) more than two estimated standard errors from the nominal size (0.05) are set in bold type.

Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	Y <sub>5</sub>	Count	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	Y <sub>5</sub>	Count
1	1	1	1	1	426	2	1	1	1	1	73
1	1	1	1	2	38	2	1	1	1	2	11
1	1	1	2	1	16	2	1	1	2	1	7
1	1	1	2	2	47	2	1	1	2	2	17
1	1	2	1	1	11	2	1	2	1	1	9
1	1	2	1	2	2	2	1	2	1	2	3
1	1	2	2	1	12	2	1	2	2	1	5
1	1	2	2	2	28	2	1	2	2	2	24
1	2	1	1	1	21	2	2	1	1	1	54
1	2	1	1	2	7	2	2	1	1	2	16
1	2	1	2	1	0	2	2	1	2	1	6
1	2	1	2	2	9	2	2	1	2	2	28
1	2	2	1	1	8	2	2	2	1	1	36
1	2	2	1	2	3	2	2	2	1	2	24
1	2	2	2	1	5	2	2	2	2	1	35
1	2	2	2	2	43	2	2	2	2	2	559

**Table 4:** Labor force data (1=employed, 2=unemployed).

those for AD(3) against AD(4) are not significant ( $P=0.340$ ,  $P=0.384$ ,  $P=0.418$ , and  $P=0.458$  for the likelihood ratio, modified likelihood ratio, score, and Wald tests respectively). Thus we conclude that AD(3) is the best constant-order AD model.

Next, we test for time-invariant third-order transition probabilities under the assumption of an AD(3) model, finding insufficient evidence against this hypothesis ( $P = 0.205$  for the score test and  $P = 0.201$  for the modified likelihood ratio test). Maximum likelihood estimates of the time-invariant third-order transition probabilities under AD(3) are as follows:

$$\begin{aligned}\hat{\pi}_{1+|111}^{(3)} &= 0.896, \quad \hat{\pi}_{1+|112}^{(3)} = 0.257, \quad \hat{\pi}_{1+|121}^{(3)} = 0.774, \quad \hat{\pi}_{1+|122}^{(3)} = 0.219, \\ \hat{\pi}_{1+|211}^{(3)} &= 0.772, \quad \hat{\pi}_{1+|212}^{(3)} = 0.214, \quad \hat{\pi}_{1+|221}^{(3)} = 0.651, \quad \hat{\pi}_{1+|222}^{(3)} = 0.077.\end{aligned}$$

From these we note that among the transition probabilities that the women were employed given that they were unemployed only once during the preceding three years,  $\hat{\pi}_{1+|211}^{(3)}$  and  $\hat{\pi}_{1+|212}^{(3)}$  are very similar and are much larger than  $\hat{\pi}_{1+|112}^{(3)}$ , indicating that the employment statuses of those women at each of the last two years were more closely related to those of the immediately preceding year than to those at the two further preceding years. This is consistent with expectation. Similarly,  $\hat{\pi}_{1+|122}^{(3)}$  and  $\hat{\pi}_{1+|212}^{(3)}$  are very similar and much smaller than  $\hat{\pi}_{1+|221}^{(3)}$ .

Although the transition probabilities do not vary significantly over

time, strict stationarity is strongly rejected ( $P < 9.0 \times 10^{-5}$  for the score and modified likelihood ratio tests). An examination of each year's marginal probability of employment suggests that the rejection of strict stationarity is due largely to a significantly smaller level of employment among surveyees in 1967 ( $\hat{\pi}_{1++++} = 0.427$ ) and a significantly larger level in 1969 ( $\hat{\pi}_{1++1++} = 0.490$ ) than in the three other years ( $\hat{\pi}_{1+++} = 0.461$ ,  $\hat{\pi}_{+++1+} = 0.469$ ,  $\hat{\pi}_{++++1} = 0.457$ ).

Based on these findings, we recommend that any additional model development [e.g. fitting an autoregressive binary time series model, such as those featured by Cox and Snell [8]] take the order of the model to be three, take third-order transition probabilities to be time-invariant, and allow marginal probabilities to vary over time.

## Conclusions

In this article, we derived score and Wald tests for order of antedependence, timeinvariance of transition probabilities, and strict stationarity of categorical longitudinal data. This work generalizes that of Anderson and Goodman [2] for testing for firstorder antedependence to testing for antedependence of any variable order, and for testing for time-invariance of transition probabilities in a first-order antedependence model to testing for such invariance, and for the even stronger property of strict stationarity, in an antedependence model of any constant order. These tests are first-order equivalent and asymptotically optimal; however, their performance differs for finite sample samples. To perform the likelihood ratio test, one must estimate the models under both the null and alternative hypotheses; an advantage of the score and Wald tests is that they require only one model to be estimated. However, an important disadvantage of the Wald test is its lack of invariance to the parameterization used.

Based on our simulation results, for testing for the order of antedependence we recommend the modified likelihood ratio test if the sample is of size 50 or smaller, and the score test otherwise. The Wald test for order of antedependence proposed herein is markedly inferior to both and should not be used. Furthermore, the score and likelihood ratio tests for time-invariant transition probabilities and for strict stationarity appear to perform about equally well. Therefore, for the sake of consistency, we recommend that the practitioner test for time-invariant transition probabilities and strict stationarity using the same type of test (likelihood ratio or score, depending on whether the sample size exceeds 50) as he/she uses to test for the order of antedependence.

## Acknowledgements

We thank Professor Joseph B. Lang in the Department of Statistics and Actuarial Sciences at the University of Iowa for his review and comments. We also thank two anonymous referees, whose comments led to several improvements in the content and clarity of this article.

## References

- Xie Y, Zimmerman DL (2013) Antedependence models for nonstationary categorical longitudinal data with ignorable missingness: likelihood-based inference. Stat Med 32: 3274-3289.
- Anderson TW, Goodman LA (1957) Statistical Inference About Markov Chains. Ann Math Statist 28: 89-110.
- Lawley DN (1956) A General Method for Approximating to the Distribution of Likelihood Ratio Criteria. Biometrika 43: 295-303.
- Zimmerman DL, Nunez-Anton V (2010) Antedependence Models for Longitudinal Data, Boca Raton, Florida: CRC Press.
- Agresti A (2002) Categorical Data Analysis, Boca Raton, FL: John Wiley & Sons, Inc.
- Lang JB (2004) Multinomial-Poisson Homogeneous Models for Contingency Tables. Annals of Statistics 32: 340-383.

7. Lindsey JK (1993) Models for Repeated Measurements, New York: Oxford University Press.
8. Cox DR, Snell EJ (1989) Analysis of Binary Data. (2ndedn), London: Chapman and Hall/CRC Press.