COPING WITH NONSTATIONARITY IN CATEGORICAL TIME SERIES

Monnie McGee and Ian R. Harris

Department of Statistical Science

Southern Methodist University Dallas, TX 75275 mmcgee@smu.edu & iharris@smu.edu

Key Words: Categorical data, DAR models, stationary models, time series

ABSTRACT

Categorical time series are time sequenced data in which the values at each time point are categories rather than measurements. A categorical time series is considered stationary if the marginal distribution of the data is constant over the time period for which it was gathered and the correlation between successive values is a function only of their distance from each other, and not of their position in the series. However, there are many examples of categorical series which do not fit this rather strong definition of stationarity. Such data show various non-stationary behavior, such as a change in the probability of the occurrence of one or more categories. In this paper we introduce an algorithm which corrects for nonstationarity in categorical time series. The algorithm produces series which are not stationary in the traditional sense of its definition often used for stationary categorical time series. The form of stationarity is weaker, but still useful for parameter estimation. Simulation results show that this simple algorithm applied to a DAR(1) model can dramatically improve the parameter estimates in some cases.

1. INTRODUCTION

Categorical time series are serially correlated data for which an observation at a time point is recorded in terms of a state (or a category). Some such series are continuous series which one can analyze as categorical ones, for example a sequence of rainfall data in which successive days were recorded as "wet" or "dry" (Chang, Kavvas, and Delleur, 1984 a,b). Other series are truly categorical in nature. Examples include geomagnetic reversals in the polarity of the earth from "normal" polarity to "reverse" polarity (Negi, *et. al.*, 1993), and

records of brain waves during a person's sleep using an EEG, where readings are classified into one of six possible states (Stoffer *et. al.* , 1988). Regardless of the origin of the series, it is clear such series are in fact quite common, although they have received much less attention in the literature than continuous variable time series.

A categorical time series $\{X_1, X_2, \ldots, X_t\}$ is considered to be stationary if any two n-tuples, say $\{X_1, \ldots, X_n\}$ and $\{X_{1_h}, \ldots, X_{n+h}\}$, have the same distribution for every $n \geq 1$ and $h \geq 0$. (Jacobs and Lewis, 1978a,b) . This definition is too strong for most applications, as it involves strict assumptions on the joint distribution of consecutive sequences. Another possible definition, implied in Stoffer, *et. al.* , 1993, is such that $P(X_t = c_j)$ is constant in $t = 1, 2, 3, \ldots$, for every $j = 1, 2, \ldots, C$, where C is the number of possible categories. One presumes that the correlation between two values is not dependent on the position of the values in the series; only on their distance from each other. More precisely, that $P(X_t = c_j \cap X_{t+h} = c_k) = f_{j,k}(h)$, where we define $Cov(X_t, X_{t+h}) = f(h)$. The latter definition is analogous to "weak" stationarity in numerical time series.

Many categorical time series are not stationary in either the strong or the weak sense. As an example, the top panel of Figure 1 shows El Niño data gathered from 1525 to 1987 (Quinn, *et. al.*, 1987). In this series, 1 indicates presence of the El Niño and 0 indicates its absence. There is a distinct change in probability of El Niño occurrences around time value 290. This change in probability could be due either to better recording of events after this point (time 290 is roughly the year 1815), or to a real change in probability due to a change in weather patterns. Since the probability of an El Niño year changes quite abruptly, it is clear that these data do not fit the definition of stationarity used in categorical time series.

The bottom panel shows data indictating the winner of Major League Baseball's All-Star Game from 1950 to 2005. An American League win is coded as 1, and a National League win is coded as a 0. The data exhibit clear signs of non-stationarity: the National League dominated until roughly the 1980's or 1990's, and the American League has dominated in the last fifteen years or so. Another example (not shown) is data dealing with geomagnetic reversals of the polarity of the earth from North polarity to South polarity (Negi, *et. al.*,

1993). In that article, the authors state that they are unable to use all of the data that they have because it is clearly nonstationary. Instead, they choose to use a portion of the data that looks stationary, according to a time plot.

The focus of this work is examining the effects of non-stationarity on parameter estimation in categorical time series and introducing an algorithm to induce a form of stationarity in nonstationary series. In Section 2, a simple flipping algorithm is introduced, which can be applied to certain non-stationary categorical time series to make them stationary. Simulation results, which show that the correlation parameter estimator from a stationary model can be dramatically improved after applying the algorithm, are given in Section 3. However, the stationarity resulting from the flipping algorithm is not distributional stationarity, but something weaker. We define this form of stationarity and discuss its properties in Section 4. In Section 5, the detrending algorithm is illustrated with data from the sequence of league wins in Major League Baseball's All–Star game from 1950 to 2005.

2. THE FLIPPING ALGORITHM

In this section, we introduce a simple algorithm which takes a non-stationary series and transforms it to one that is stationary. For simplicity, the initial focus is on series with a binary outcome, where we arbitrarily denote the categories by 0 and 1. The flipping scheme assumes that one of the categories is more common at the beginning of the series, and that there is then a transition so that by the end of the series the other category is more common. Without loss of generality, the category that is more common at the beginning of the series will be labeled the 0 category. For simplicity, we will examine in detail only the case with one transition, from $(0 \rightarrow 1)$, although the algorithm can be extended to multiple transitions.

The algorithm is as follows:

1. Denote the original non-detrended series by $X_1, X_2, \ldots, X_T$.

2. Create T new series where the kth series is created by "flipping" observations $X_1, X_2, \ldots, X_k$. For example, the first series would be the same as the original series, except that the first observation would be changed from 0 to 1 (or 1 to 0). The next series would result

3

from flipping the first two observations, etc. The last series would be the complete opposite of the original series.

3. Count the number of ones in each of the $T + 1$ series (the original series and the T "new" ones).

4. The series with the highest number of ones is the detrended series. In case of a tie for the highest number of ones, choose the first series in the sequence with the highest number of ones (that is, the one with the fewest flips).

As a simple example, consider the sequence $0, 1, 0, 0, 1, 0, 1, 1$. There are nine sequences to consider: the original one, and the eight new ones generated by flipping as above. The sequences are given in Table 1 . The first row of the table gives the original sequence, and the next eight rows the generated sequences obtained from the original. There are two sequences with the maximum number of 1's, the $k = 4$ and $k = 6$ sequences, obtained by flipping the first four and first six respectively. Both of these sequences have six ones. By convention the tie is broken by using the earlier sequence $(k = 4)$, as it is the "least disturbed" compared to the original one.

In general one would apply this algorithm to sequences which exhibit a trend in the number of ones. That is, sequences for which there exists a point k such that 0's are more common than 1's for $t < k$ and 1's more common than 0's for $t > k$. By design this algorithm will then produce a sequence where 1's are more common than 0's both for $t < k$ and $t > k$, and hence one can say the trend has been removed.

We can also extend the algorithm to series with more than two categories in the following manner. Suppose there are three categories, 1, 2 and 3. In this situation, without loss of generality, one can label the category that is most likely early in the sequence as category 1 and the category that is most likely at the end of the sequence as category 3. We assume there are either two transitions in terms of the most likely category, $(1 \rightarrow 2 \rightarrow 3)$, or one transition $(1 \rightarrow 3)$. It would also be possible to extend the idea to more than one transition using a similar scheme to the one described below.

To transform the series to a stationary sequence, we create a new sequence such that category 3 is most likely everywhere. To do this, we define two cut points k_1 and k_2 such that $k_1 \leq k_2$ and k_1, k_2 are chosen from $0, \ldots, n$. If the sequence is of length n there are then $\binom{n+2}{2}$ choices for (k_1, k_2). For each pair of cutpoints (k_1, k_2) create a new sequence where if $t \leq k_1$ then flip categories 1 and 3, but leave 2 unchanged, if $k_1 < t \leq k_2$ then flip categories 2 and 3, but leave 1 unchanged, and if $k > t$ we leave the categories unchanged. Now for each sequence count the number of category 3 responses, and choose as the detrended sequence the one with the highest number of 3's. To break any tie choose the sequence with smallest k_2 and then smallest k_1. Extensions to a greater number of categories than three are possible, but the algorithm becomes much more computationally demanding.

3. NONSTATIONARY SERIES AND THE EFFECT OF THE FLIPPING ALGORITHM

Simulation results given in this section show the effects of non-stationarity and the flipping algorithm on the fit of the Discrete Autoregressive Model (DAR) model of Jacobs and Lewis (1978a, b). The DAR model is used as an illustration of a simple stationary model, without implying that this model is necessarily the best one to use for categorical data in general. The primary motivation here is to show that non-stationarity can seriously compromise the fit, but detrending, while not a panacea, can help. First, we describe the DAR model, then give the simulation results.

The sequence $\{X_t\}$ has a binary DAR structure when it is formed according to the probabilistic linear model

$$X_t = I_t X_{t-1} + (1 - I_t) Y_t, \qquad t = 1, 2, 3, \ldots.$$

where $\{Y_t\}$ is a sequence of independent binary random variables with $P(Y_t = 1) = p_t$, and $\{I_t\}$ is a sequence of independent binary random variables, also independent from $\{X_t\}$ and $\{Y_t\}$. Let $P(I_t = 1) = q$, where $0 \leq q \leq 1$ is fixed. Typically one assumes that $X_1 = Y_1$. Note that X_t is also binary, and it is a simple matter to show that if $p_t = p$ then $P(X_t = 1) = p$, and hence $\{X_t\}$ is a stationary series.

Figure 2 shows data simulated from a DAR(1) model with three different values of q.

5

Note that the value of q controls the probability of X_t staying in the same state (0 or 1). As the figures show, if q is very large, it is very likely that $X_t = X_{t-1}$, and long runs in the series dominate.

Our primary interest here is in estimation of q, as that is the most useful parameter in one-step forecasting, and in the non-stationary cases p has no clear meaning. Indeed it is easy to show that if one assumes the stationary DAR model then $cor(X_t, X_{t-1}) = q$. A simple method of moments estimator of q would then be the lag one correlation between successive observations.

The data for the simulations consist of series of length 200 generated from a DAR(1) model with one transition at a chosen point k where the probability of a 1 changes from p to $1 - p$. The value of q, the correlation parameter, was one of $q = 0.1, 0.3, 0.5, 0.7$. The value of p was chosen to be either 0.1, 0.3, or 0.5. Values less than 0.5 were chosen because of the way the detrending algorithm is defined. It is assumed that the 0 category is the most likely category before the transition point, and that the most likely category transitions from 0 to 1. The transition point was chosen as either $k = 50, 100,$ or 150. The simulation model is the most favorable for the flipping algorithm, however, simulations from other models, described in Section 4, also show good performance for the algorithm.

For each value of p, q, and k, one thousand sequences of length two hundred were generated. The method of moments estimate of q under the DAR(1) model was calculated. Next mean squared errors (MSE's) for the estimates of q for each of these forty-five combinations of the parameters were calculated. Table 2 gives these MSE's, multiplied by 1000. Only the results for $k = 100$ are shown because the results for $k = 50, 100$ are similar.

The simulation error, on the scale given in the table, is around 1 for the smallest MSE's and 5 for the largest MSE's. The MSE's for the estimation of q for the raw data (not detrended) are given in the top row for each value of p. The MSE's in bold-face type are those from the detrended series. Maximum likelihood estimates were also calculated, but the results are not shown because they are very similar to the ones given here.

From the table one can make several observations. First, non-stationarity has a serious

effect on the estimate of q, particularly when q is small, which would be more typical for real data. Second, flipping the non-stationary series produces much better estimates of q. Further examination of the estimates shows that bias is a serious problem with non-stationarity. For example, if $p = q = 0.1$ and $k = 100$, the average value of $\hat{q}$ was 0.60. With flipping the average value fell to 0.09, an almost total reduction in bias. (For a less extreme example, when $p = 0.1$, $q = 0.7$, and $k = 100$, the average value of q without detrending was 0.88. With flipping it was 0.63). This illustrates the potential value of the algorithm.

4. WEAKLY STATIONARY CATEGORICAL TIME SERIES

The detrending algorithm produces a series which is stationary, but not in the strong sense described in the introduction. The output of the detrending algorithm is a series such that the identity of the most common category is the same over time, although its probability could change. We will term this type of stationarity categorical, or modal, stationarity, and it will be denoted by C(1). In general, one could have a series for which the identity of the $J - 1$ most likely categories remained the same, with all the others changing. This would be C(J) stationarity.

For completeness, we simulated 1000 C(1) stationary categorical time series of length 200 and estimated q without applying the detrending algorithm. We also did the same for a non-stationary series where $p_t = P(Y_t = 1)$ changes linearly with time. That is, $p_t = \beta_0 + \beta_1 t$. By choosing different values of β_0 and β_1 one can control how rapidly p_t changes with t and what range of values are observed across the sequence. Further, we simulated strongly stationary series (which we term distributionally stationary, denoted by D(1)) and obtained estimates of q without applying the detrending algorithm.

The results are given in Table 3. Clearly, trying to estimate q from a nonstationary series results in poor estimates. Distributional stationary series produce good estimates of q, as the DAR model itself is distributional stationary. Estimates of q from categorical stationary series are better than those from nonstationary series. Detrending a nonstationary series to C(1) using the flipping algorithm results in better estimates for q than does starting with a pure C(1) stationary series, except for large values of q.

7

It is interesting that even though the estimation procedure for q assumes strong stationarity, parameter estimates from a weakly stationary series are a great improvement over estimates from series that are completely nonstationary. In turn, parameter estimates from series resulting from the flipping algorithm give estimates that are almost as good as estimates from D(1) stationary series.

5. DETRENDING THE ALL–STAR DATA

Since 1933, Major League Baseball has played an All–Star game, where the best and most popular players from the American and National Leagues play against each other. The All–Star Game was not played in 1945, and two games were played each year between 1959 and 1962. The data were gathered from the Major League Baseball web site (www.mlb.com) for the years 1950 to 2005.

This time period was the focus largely for the simplicity of illustrating a series with only one transition. The result for the 2002 game was omitted because the game resulted in a tie. For the years where two games were played, the winner for that year was taken to be the league that scored the most combined runs against the other. A plot of the data was given in Figure 1(b), and a decadal summary of the data is displayed in Table 4.

The series was detrended using the scheme outlined in the section 3. The detrended series is the opposite of the original series until the year 1985, in other words the detrending algorithm picked 1985 as the year that superiority switched from the National League to the American League.

Assuming a stationary DAR model, the value of q was estimated by the sample lag-one correlation for both the original and detrended series. The estimate for q for the original series is 0.39, and for the detrended series is 0.19. This is a substantial reduction and shows that detrending can have a large impact on estimates of q.

6. DISCUSSION

In this paper, we introduce an algorithm for inducing a type of stationarity in categorical time series data. The algorithm does not really detrend the series in the traditional sense,

8

because the kinds of examples we consider have an abrupt change in the probability that a particular category occurs, rather than a trend in the probability of the occurrence of a category. The algorithm uses mild assumptions on the data, and produces a series which is stationary, but not in the strong sense. We term this weaker form of stationarity "categorical stationarity".

Using a simple strongly stationary model, it was shown via simulation that fitting a strongly stationary model to non-stationary data can result in poor estimates of the correlation parameter, but that the estimate can be dramatically improved by first using the flipping algorithm. Additional simulations show that fitting a model which assumes strong stationarity of the data to a series which is category stationary (without flipping) gives estimates which are less biased than the estimates would have been if the non-stationary data were fit.

It is clear that the most widely accepted definition of stationarity in categorical time series is too strong for some real categorical time series data. Such data typically have changepoints where the probability of the most likely category changes (or the most likely category itself changes). However, many models for categorical time series, such as the DAR and DARMA models, and other methods, such as spectral estimation (Stoffer, 1991; McGee & Ensor, 1998) were developed under the assumption that the series is strongly stationary. Additional work is needed to ascertain whether the use of methods intended for strongly stationary categorical time series would still be valid for categorically stationary categorical time series. Further extensions and improvements to the flipping algorithm are also left for future work.

REFERENCES

Chang, T. J.; Kavvas M. L.; and Delleur J. W. (1984). Modeling of Sequences of Wet and Dry Days by Binary Discrete Autoregressive Moving Average Processes. *Journal of Climate and Applied Meteorology* , 23:1367-1378.

Jacobs, P. A. and Lewis P. W. (1978a). Discrete Time Series Generated by Mixtures I:

Correlational Runs and Properties. *Journal of the Royal Statistical Society, Series B* 40: 94-105.

Jacobs, P. A. and Lewis P. W. (1978b)Discrete Time Series Generated by Mixtures II: Asymptotic Properties. *Journal of the Royal Statistical Society, Series B* , 40: 222-228.

Jacobs, P. A. and Lewis P. W. (1983). Stationary Discrete Autoregressive–Moving Average Time Series Generated by Mixtures. *Journal of Time Series Analysis*, 4:19-36.

McGee, M. and Ensor, K. (1998). Tests for Harmonic Components in the Spectra of Categorical Time Series. *Journal of Time Series Analysis*, 19: 309-324.

Negi, J. G., Tiwari, R. K., and Rao, K. N. N. (1993). Comparison of Walsh and Fourier Spectroscopy of Geomagnetic and Nonsinusoidal Paleoclimatic Time Series. *IEEE Trans. Geosci. Remote Sensing* , 31: 127- 135.

Quinn, W. H., Victor T. N., and Santiago E. A. (1987). El Niño Occurrences Over the Past Four and a Half Centuries. *Journal of Geophysical Research* 92: 14449 - 14461.

Stoffer, D. S., Scher, M. S., Richardson, G. A., Day, N. L., and Coble, P. (1988). A. A Walsh–Fourier Analysis of the Effects of Moderate Maternal Alcohol Consumption on Neonatal Sleep–state Cycling. *Journal of the American Statistical Association*, 83: 954-963.

Stoffer, D. S., Tyler, D. E. , and McDougall, A. J. (1993). Spectral Analysis for Time Series: Scaling and the Spectral Envelope. *Biometrika*, 80: 611-622.

Table 1: Sequences for example

Number flipped, k	Sequence	Number of 1's
0	0 1 0 0 1 0 1 1	4
1	1 1 0 0 1 0 1 1	5
2	1 0 0 0 1 0 1 1	4
3	1 0 1 0 1 0 1 1	5
4	1 0 1 1 1 0 1 1	6
5	1 0 1 1 0 0 1 1	5
6	1 0 1 1 0 1 1 1	6
7	1 0 1 1 0 1 0 1	5
8	1 0 1 1 0 1 0 0	4

Table 2: MSE times 1000 for the methods of moments estimate of q for models which are not detrended (top row) and detrended (bottom row - bold face type).

	True Value of q			
	0.1	0.3	0.5	0.7
$p_t = 0.1$	331	198	99	35
	9	**14**	**18**	**25**
$p_t = 0.3$	26	17	9	4
	5	**6**	**5**	**5**
$p_t = 0.5$	5	4	4	
	6	**5**	**4**	

Table 3: MSE times 1000 for the estimate of q for different degrees of stationarity.

Model		*True Value of q*			
		0.1	0.3	0.5	0.7
non-stationary	p_t varies	90	55	30	11
D(1) stationary	$p_t = 0.5$	4	5	4	3
	$p_t = 0.75$	5	6	6	5
C(1) stationary	p_t varies	15	11	8	5
Detrended to C(1)	p_t varies	9	8	6	7

Table 4: Decadal summary of National League and American League Wins in the All–Star game from 1950 to 2005.

Decade	Wins by NL	Wins by AL
1950's	6	4
1960's	9	1
1970's	9	1
1980's	6	4
1990's	3	7
2000's	0	5

12

Figure 1: El Niño and All-Star Data. Note clear shifts in the probability that each series takes on the value 1.

**El Nino Occurrences**

(a)

All–Star Game Outcome

(b)

13
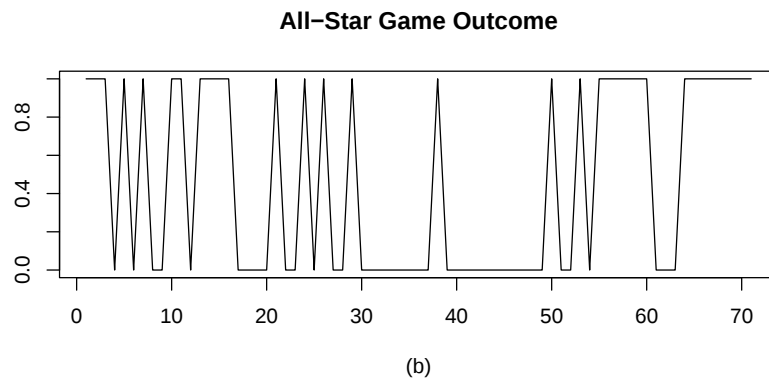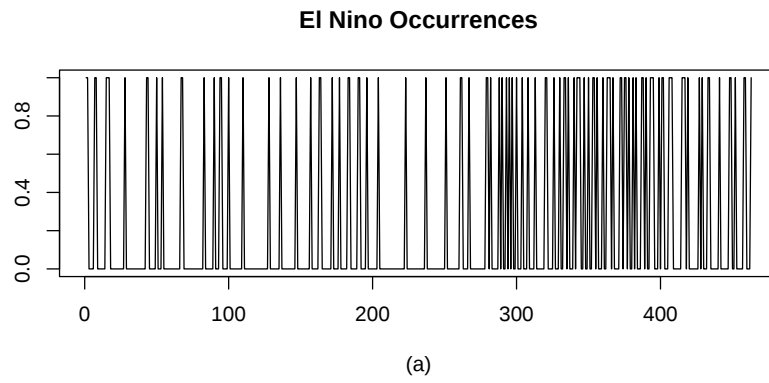
Figure 2: Simulated Stationary DAR(1) models with $p = 0.5$ and $q = 0.1$ (top), $q = 0.5$ (middle), and $q = 0.9$ (bottom).

Stationary DAR(1) Model with p = 0.5 and q = 0.1

Stationary DAR(1) Model with p = 0.5 and q = 0.5

Stationary DAR(1) Model with p = 0.5 and q = 0.9

14