Analyzing Time Series with Long-Range Dependencies 1: Using Fractional Differencing to Detect Irregularity

Matthijs Koopmans – *Mercy College*

Poster Presented at the Annual Modern Modeling Methods (M^3) Conference, Storrs, CT. May 22, 2018.

Abstract

The purpose of this poster is to show an application of fractional differencing (Beran, 1994), a recently developed technique to analyze long-range irregular patterns in time series data, using two illustrative datasets: 1. The annual discharge volume of the River Nile from 622 through 1245 AD (a 'classic' dataset), and 2. Daily attendance rates in one New York City high school from 2010-2014. The main features of fractional differencing are explicated, and it is shown how irregularity manifests itself in the time series and diagnostic autocorrelation function plots of these data. The differencing parameter *d* is generated and interpreted for both datasets.

Rationale for the Study

- Irregularity in time series is suggestive of complex adaptive processes in the system of interest, with unpredictably recurring cycles.
- As an extension to conventional time series analysis, fractional differencing approaches are well within reach of social science research.

Datasets

- A time series plot of the annual recordings of water discharge levels of the River Nile near Cairo from 622 through 1245AD, a classic dataset that is often cited as an illustration of long-range processes (Beran, 1994).
- Daily attendance rates in one New York City high school (School 1) over a four-year period, starting in the fall of 2010.

Analytical Steps

- Initial exploration of the data and conducting stationarity tests;
- Smoothing of the series to address extreme values (attendance data only);
- Comparison of statistical models, using Akaike's Information Criterion and the Ljung-Box Portmanteau test;
- Analysis of residuals;
- Interpretation of the parameter estimates.

The Fractional Differencing Model

- Assuming a time series trajectory $Y_t, Y_{t-1}, Y_{t-2}, \dots Y_{t-n}$ with $\mu = E(Y_t) = 0$ and a lag operator:

$$B(Y_t) = Y_{t-1};\ B^2(Y_t) = Y_{t-2}; \dots\ B^n(Y_t) = Y_{t-n},$$

the fractional differencing model can be expressed as:

$$(1 + \varphi_1 B + \varphi_2 B^2 + \cdots + \varphi_p B^p)(1 - B)^d Y_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) e_t.$$

- The autoregressive component (at *p* lags) and long-range memory component, respectively, are on the left; the moving average component (at *q* lags) is on the right.
- We assume a stationary series, with $-0.5 < d < +0.5$.
- The differencing parameter *d* indicates long-range autocorrelation. A positive value denotes *persistence*, a negative value of *d* indicates *anti-persistence* (Sowell, 1992).

Results

Table 1.
Summary Statistics and Stationarity Tests: River Nile Data (N = 623)

Summary Statistics	
Mean	11.48
Standard Deviation	0.89
Minimum	9.35
First Quartile	10.94
Median	11.48
Third Quartile	12.05
Maximum	14.66
Stationarity Tests	
Augmented Dickey Fuller Test	-5.03* Lag Order = 8
KPSS Test	
Level	1.91*
Trend	0.30* Lag Order = 5
* p < .01. A rejection of the null hypothesis implies stationarity in all three tests.	

Table 2.
Fractional Differencing Model Selection Process: River Nile Data

Model	Specification (*p*, *d*, *q*)	*d*	σ²	AIC	LB Test[B]
I	(0, 0, 0)	--	.79	1724.22	1001.00*
II	(1, 0, 1)	--	.50	1429.04	15.48
III	**(0, *d*, 0)**	**.39**	**.49**	**1413.07**	**7.30**
IV	(1, *d*, 1)	.37	.49	1417.11	5.25

* p < .05
[B] Ljung-Box Portmanteau tested under a χ^2 distribution at df=12
Preferred Model in Boldface

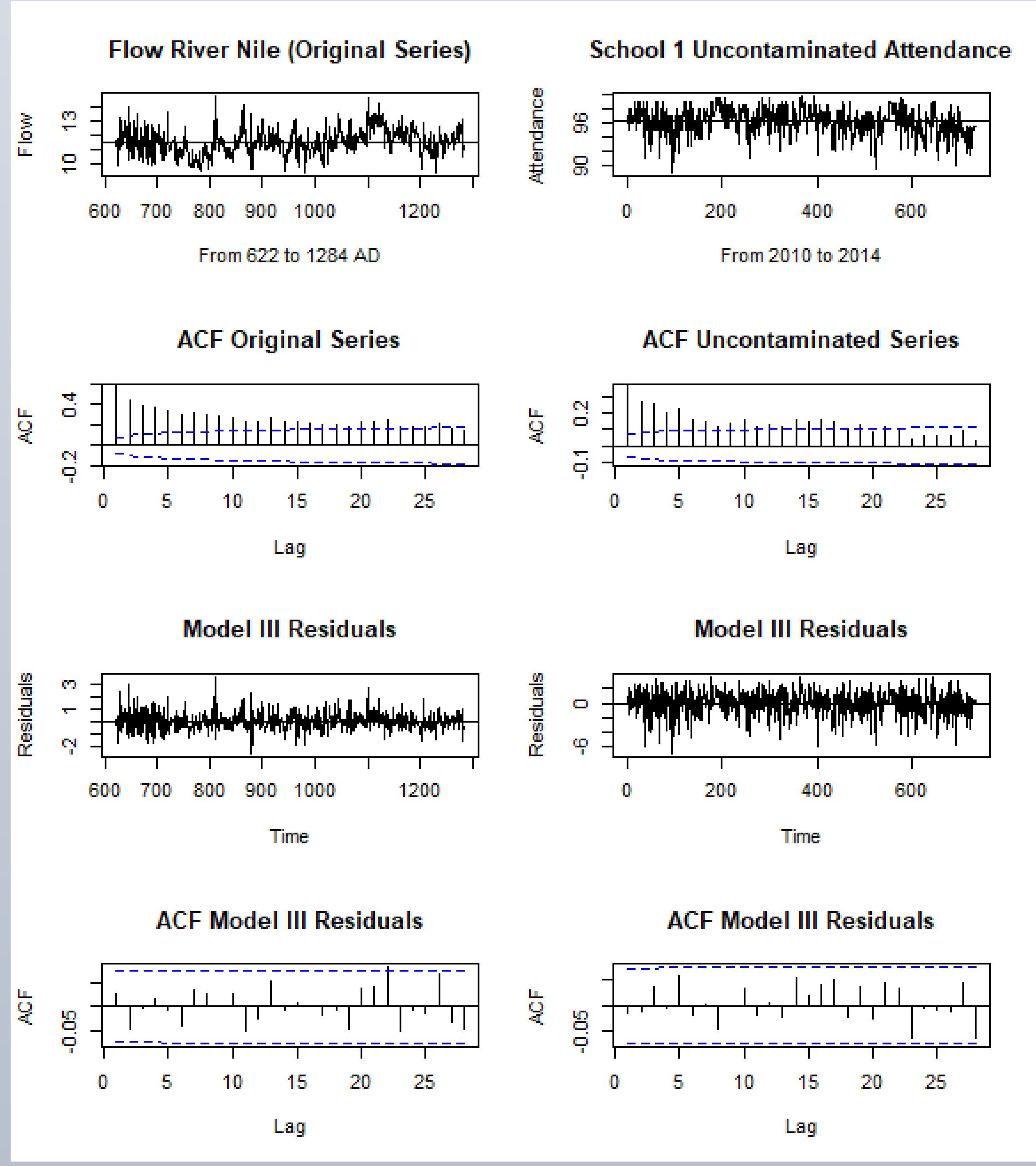

Table 3.
Summary Statistics and Stationarity Tests: Daily Attendance Rates School 1 (N = 735)

Summary Statistics	**Uncontaminated Series**
Mean	96.09
Standard Deviation	1.86
Minimum	88.78
First Quartile	94.95
Median	96.28
Third Quartile	97.31
Maximum	99.63
Stationarity Tests	
Augmented Dickey Fuller Test	-6.00* Lag Order = 9
KPSS Test	
Level	1.37*
Trend	0.35* Lag Order = 6
* p < .01. A rejection of the null hypothesis implies stationarity in all three tests.	

Table 4.
Fractional Differencing Model Selection Process: Daily Attendance Rates in School 1

Model	Specification (*p*, *d*, *q*)	*d*	σ²	AIC	LB Test[B]
I	(0, 0, 0)	--	3.45	2998.74	371.66*
II	(1, 0, 1)	--	2.83	2857.43	7.41
III	**(0, *d*, 0)**	**.27**	**2.81**	**2850.99**	**6.54**
IV	(1, *d*, 1)	.23	2.82	2855.86	6.65

* p < .05
[B] Ljung-Box Portmanteau tested under a χ^2 distribution at df=12
Preferred Model in Boldface

Summary of Results

- Persistence is moderate in the attendance data (*d* = .27), substantial in the River Nile data (*d* = .39);
- The fractional differencing results suggest irregularity in both data sets.

Discussion

- Although short-range models also describe these data well, the models relying exclusively on the long-range parameter describe the data best in both cases indicating long-range dependencies.
- The significance of the differencing parameters points to a complex adaptive process in these systems that warrants further examination.

References

Beran, J. (1994). *Statistics for long-memory processes*. New York: Chapman-Hall.

Sowell, F. (1992). Modeling long-run behavior with the fractional ARFIMA model. *Journal of Monetary Economics*, *29*, 277-302.

Appendix