Analyzing Time Series with Long-Range Dependencies 2: Using Exponential Smoothing to Model Complex Periodic Patterns

Matthijs Koopmans – *Mercy College*

Poster Presented at the Annual Modern Modeling Methods (M³) Conference, Storrs, CT. May 22, 2018.

Abstract

Traditional seasonal time series models estimate short-range regularities (e. g., days of the week), but are not equipped to address long-range dependencies, nested seasonal cycles, or modeling periods including fractions, such as the 365.25 days per year in the Gregorian calendar (De Livera, Hyndman, & Snyder, 2012). The *forecast* package (ibid.) in *R* makes such analyses possible. This poster shows how with two datasets: 1. Daily high school attendance rates in one New York City High School (2009-2014), modeling autocorrelations over a school year of 177 days, and 2. Daily recordings of births to teens in the state of Texas (1964-1999), requiring a model that estimates weekly as well as annual dependencies. R script for *forecast* is provided.

Rationale for the Study

Until recently, the estimation of long-range regularities in time series was cumbersome at best. This poster illustrates:

- how the Trigonometric Box-Cox ARMA Trend Seasonal (TBATS) model addresses this problem, and
- how the *forecast* package in *R* implements this model to analyze long-range dependencies statistically.

Datasets

- Daily high school attendance rates in one New York City high school (School 2) from 2009 to 2014;
- Daily recordings of births to teens in the state of Texas from 1964 to 1999 (Hamilton *et al.*, 1997).

Plan of the Analysis

- Initial exploration of the data, including stationarity tests;
- Outlier removal (attendance data only);
- Estimating short-range processes and long-range irregularity;
- Conventional ARIMA estimates with *d* =1 and weekly; seasonal estimates (teen birth data only);
- Estimation of long-range regularity with TBATS;
- Analysis of residuals.

The TBATS Model

The Trigonometric Box-Cox ARMA Trend Seasonal (TBATS) model can be expressed as follows:

$$Y_t^{(\omega)} = (Y_t^{\omega} - 1)/\omega \quad (1a)$$

$$Y_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-1}^{(i)} + d_t \quad (1b)$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \delta_t \quad (1c)$$

$$b_t = (1-\phi) b + \phi b_{t-1} + \beta d_t \quad (1d)$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)} \quad (1e)$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} cos\lambda_j^{(i)} + s_{j,t-1}^{*(i)} sin\lambda_j^{(i)} + \gamma_1^{(i)} d_t \quad (1f)$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} sin\lambda_j^{(i)} + s_{j,t-1}^{*(i)} cos\lambda_j^{(i)} + \gamma_2^{(i)} d_t \quad (1g)$$

$$d_t = \sum_{i=1}^{p} \varphi_i d_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (1h)$$

The Box-Cox transformation (1a) stabilizes the variance by ω; ℓ_t estimates the local level at *t*, b_t the short-range trend at *t,* and *b* the long-range trend across the series; d_t represents the ARMA (*p*, *q*) process; $s_{j,t}^{(i)}$ models the seasonal component as a Fourier series with $\lambda_j^{(i)} = 2\pi j/m_i$ with m_i representing the seasonal period; $s_{j,t}^{(i)}$ captures the level variance at the i^{th} seasonal cycle, and $s_{j,t}^{*(i)}$ models the change in seasonal variability over time; α, β, $\gamma_1^{(i)}$ and $\gamma_2^{(i)}$ are smoothing parameters, and ε_t is Gaussian white noise with zero mean and constant variance σ^2 (De Livera, Hyndman, & Snyder, 2012).

Results

Table 1.
Summary Statistics and Stationarity Tests:
Daily Attendance in School 2 (N = 885)

Summary Statistics	Uncontaminated Series
Mean	91.65
Standard Deviation	2.63
Minimum	82.60
First Quartile	90.28
Median	91.84
Third Quartile	93.56
Maximum	97.47
Stationarity Tests	
Augmented Dickey Fuller Test	-6.07* Lag Order = 9
KPSS Test Level Trend	 4.32* 0.66* Lag Order = 6

* p < .01. A rejection of the null hypothesis implies stationarity in all three tests.

Table 2.
Model Fitting with TBATS:
Attendance in School 2 (Uncontaminated)

Model	ARMA (p, q)	Period	σ^2	AIC	LB Test
I	--	--	5.14	7460.77	34.45
II	(0, 1)	--	5.04	7446.82	21.55
III	**(0, 1)**	**177**	**4.12**	**7383.55**	**25.82**

P < .05
Ljung-Box (LB) Portmanteau tested under a χ^2 distribution at df = 12
Preferred Model in Boldface

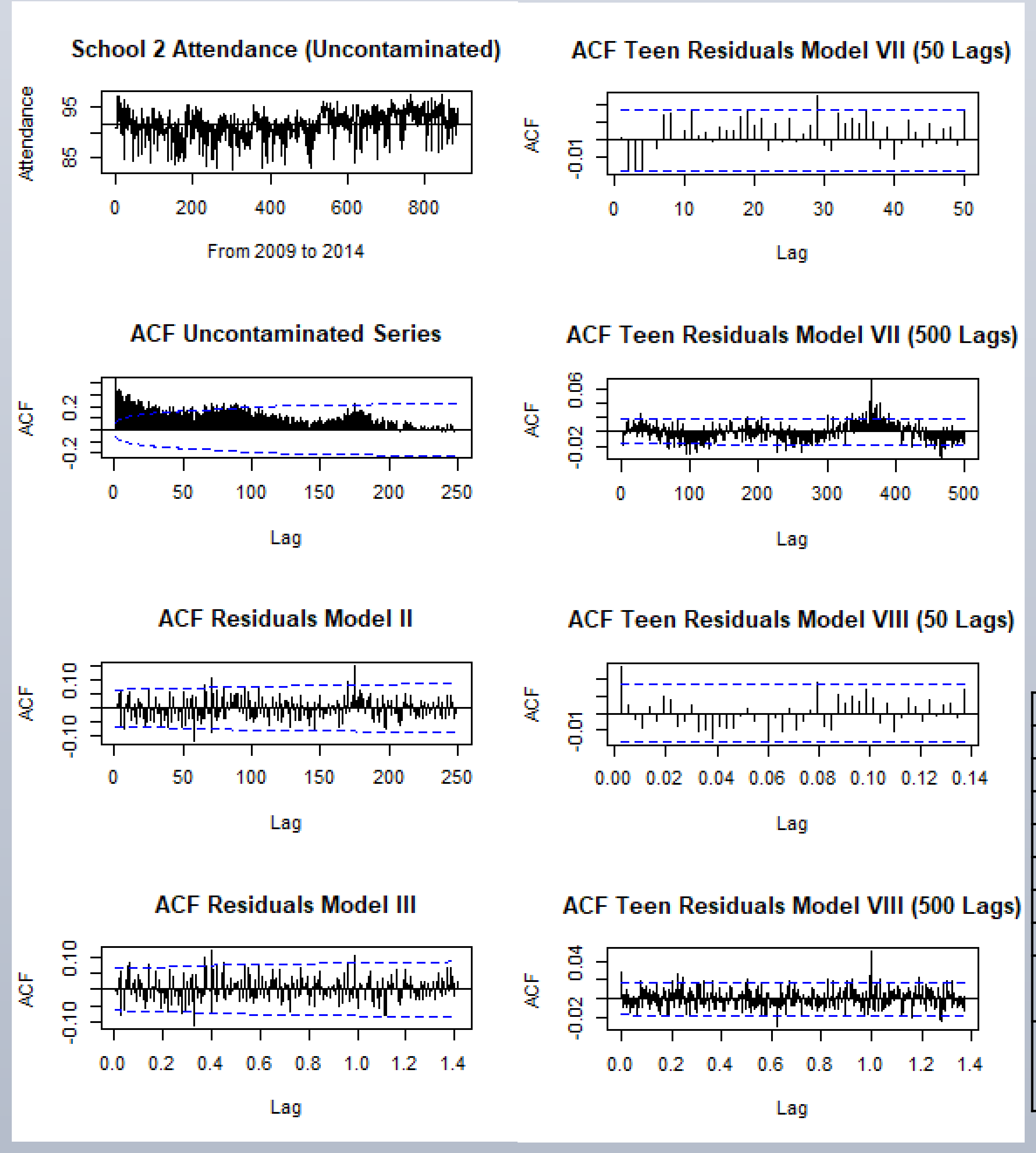


Table 3.
Summary Statistics and Stationarity Tests:
Texas Teen Births Data (N = 13,149)

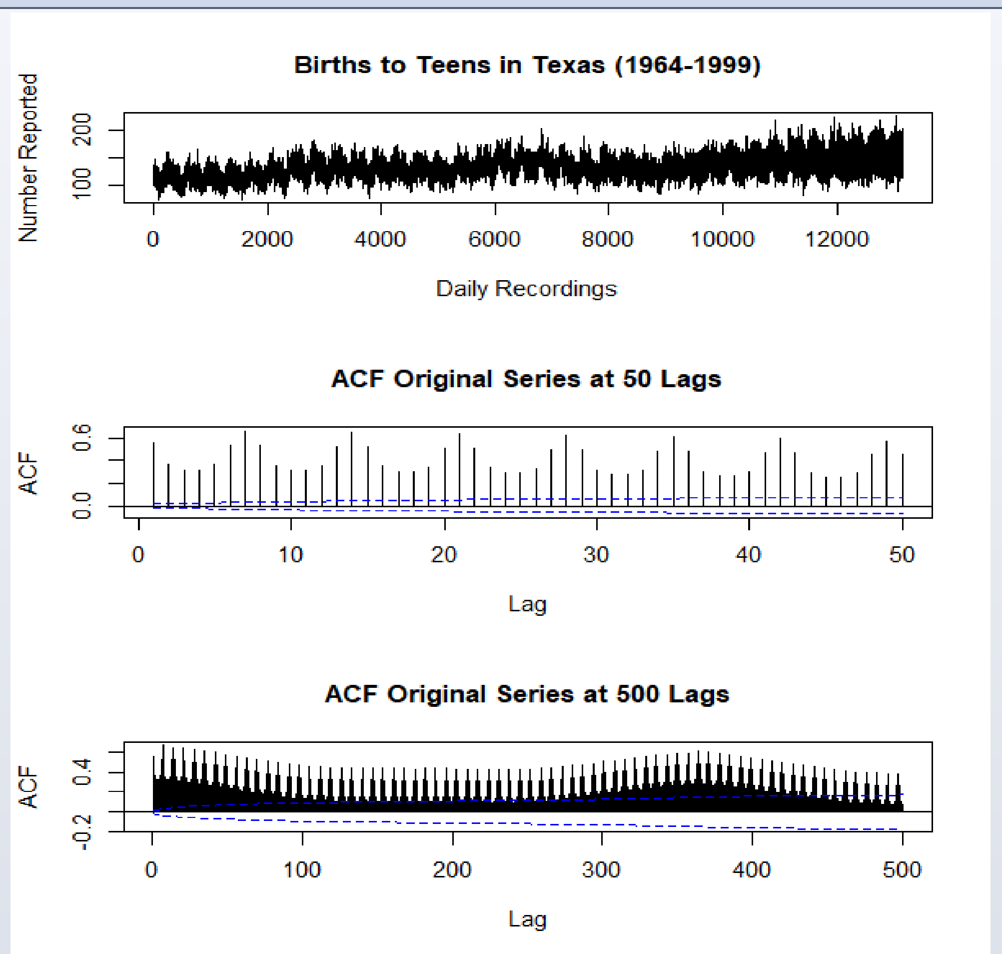
Summary Statistics	
Mean	132.2
Standard Deviation	21.04
Minimum	73.0
First Quartile	117.0
Median	131.0
Third Quartile	145.0
Maximum	226.0
Stationarity Tests	
Augmented Dickey Fuller Test	-8.25* Lag Order = 23
KPSS Test Level Trend	 24.23* 1.08* Lag Order = 26

* p < .01. A rejection of the null hypothesis implies stationarity in all three tests.

Table 4.
Model Fitting with Fractional Differencing and TBATS:
Texas Teen Births Data

Model	Specification	σ^2	LB
I	(0, 0, 0)	442.50	29,486.00*
II	(1, 0, 1)	264.40	6,719.10*
III	(0, *d*, 0)	269.96	4,989.70*
IV	(1, *d*, 1)	266.53	3,690.90*
V	$(1, 0, 1) X (0, 0, 1)_7$	234.30	2,004.80*
VI	$(1, 0, 1) X (0, 1, 1)_7$	161.80	42.56*
VII	$(1, 1, 1) X (0, 1, 1)_7$	162.50	20.87*
VIII	Residuals Model VII Seasonal Period = 365.25	158.36	17.70

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



References

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2012). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, *106*, 1513-1527.

Hamilton, P., Pollock, J. E., Mitchell, D. A., Vincenzi, A. E., & West, B. J. (1997). The application of nonlinear dynamics to nursing research. *Nonlinear Dynamics, Psychology, and Life Sciences*, *1*, 237-261.

R Script for Modeling Daily Attendance (School 2)

```
>library(forecast)  ##call the forecast package
>attach(school 2)   ##call the dataset for school 2
>summery(rate)      ##generate attendance summary statistics
>tsoutliers(rate,iterate = 2)  ##identify outliers
>urate<-tsclean(rate,replace.missing = TRUE)
                    ##replace outliers
>urate.m1<-tbats(urate,use.box.cox = T, use.trend = F,
   use.damped.trend = F, use.arma.errors = F)
>urate.m1           ##fit model 1 and generate output
>checkresiduals(urate.m1)
>urate.m2<-tbats(urate, use.box.cox = T, use.trend = F,
   use.damped.trend = F, use.arma.errors = T)
>urate.m2           ##fit model 2 and generate output
>checkresiduals(urate.m2)
>urate.msts<-msts(urate, seasonal.periods=177)
                    ##adjustment for annual cycle
>urate.m3<-tbats(urate.msts, use.box.cox=T,
   use.trend=F, use.damped.trend=F, use.arma.errors = T)
>urate.m3           ##fit model 3 and generate output
>checkresiduals(urate.m3)
```