Homework 3: Exponential Smoothing in Time Series

Georgia Institute of Technology, Business Analytics

Introduction to Analytics Modeling

Professor Joel Sokol

September 17, 2025

Files submitted: homework4_answers.pdf (this doc), homework4.R

Question 7.1

Describe a situation or problem from your job, everyday life, current events, etc., for which exponential smoothing would be appropriate. What data would you need? Would you expect the value of α (the first smoothing parameter) to be closer to 0 or 1, and why?

Situation

At my job as a software engineer, our team monitors the **number of daily user logins** to our platform. The login counts fluctuate from day to day; weekdays tend to be higher, weekends lower, but there's also random noise due to marketing campaigns, outages, or holidays.

Data Needed

We would need **time-series data** of daily user login counts over several weeks or months. This dataset would allow us to see both the underlying pattern and the random variation.

Use of Exponential Smoothing

Exponential smoothing could help us **smooth out random spikes** (like an unusual jump from a single promotion) while still tracking the general baseline level of logins. This makes it easier to monitor trends and detect unusual changes in user engagement.

Choice of α (Smoothing Parameter)

- I would expect **α to be closer to 0** because login activity has a fair amount of randomness (e.g., sudden spikes from promotions, dips on holidays).
- A lower α puts more weight on past smoothed values, making the baseline less sensitive to one-off fluctuations.
- If α were closer to 1, the model would overreact to every daily jump or dip, which would not be useful for identifying the stable underlying trend.

Question 7.2

Using the 20 years of daily high temperature data for Atlanta (July through October) from Question 6.2 (file temps.txt), build and use an exponential smoothing model to help make a judgment of whether the unofficial end of summer has gotten later over the 20 years. (Part of the point of this assignment is for you to think about how you might use exponential smoothing to answer this question. Feel free to combine it with other models if you'd like to. There's certainly more than one reasonable approach.)

Note: in R, you can use either HoltWinters (simpler to use) or the smooth package's es function (harder to use, but more general). If you use es, the Holt-Winters model uses

model="AAM" in the function call (the first and second constants are used "A"dditively, and the third (seasonality) is used "M"ultiplicatively; the documentation doesn't make that clear).

Methodology

- I used single exponential smoothing (SES) on each year's July–October daily high temperatures (1996-2015) using the `smooth::es` function in R with model = `"ANN"`.
- For each year, I defined a July baseline (median of July highs). The "end of summer" was the first day after August 1 where the smoothed highs fell at least 2°F below the July baseline, with persistence (≥5 of the next 7 days also below).
- Converted the date to day-of-year (DOY) and regressed DOY on calendar year. Also ran a Kendall trend test**.**

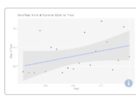
Results

- The extracted end-of-summer dates ranged from early August (~DOY 213) to late September (~DOY 268).

Table of results (`ends` tibble with year, end_date, doy)

tbl_df 20 x 3

R Console



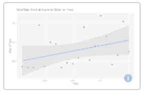
A tibble: 20 × 3

year <int>	end_date <S3: POSIXct>	doy <dbl>
1996	1996-08-01	214
1997	1997-08-01	213
1998	1998-08-01	213
1999	1999-09-15	258
2000	2000-08-01	214
2001	2001-08-28	240
2002	2002-08-26	238
2003	2003-08-01	213
2004	2004-08-05	218
2005	2005-08-05	217

1–10 of 20 rows   Previous 1 2 Next

Second part of table (2006–2015)

tbl_df 20 x 3 | R Console



A tibble: 20 × 3

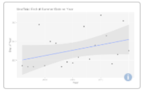
year	end_date	doy
<int>	<S3: POSIXct>	<dbl>
2006	2006-08-11	223
2007	2007-09-13	256
2008	2008-08-08	221
2009	2009-08-24	236
2010	2010-09-25	268
2011	2011-09-03	246
2012	2012-08-03	216
2013	2013-08-14	226
2014	2014-09-19	262
2015	2015-08-18	230

11–20 of 20 rows   Previous 1 2 Next

- Linear regression gave a slope of +1.15 days/year (≈ +11.5 days/decade) with p ≈ 0.015.
- Kendall's τ = 0.34, p ≈ 0.045, confirming a positive monotonic trend.

Regression output (summary of `lm_fit`)

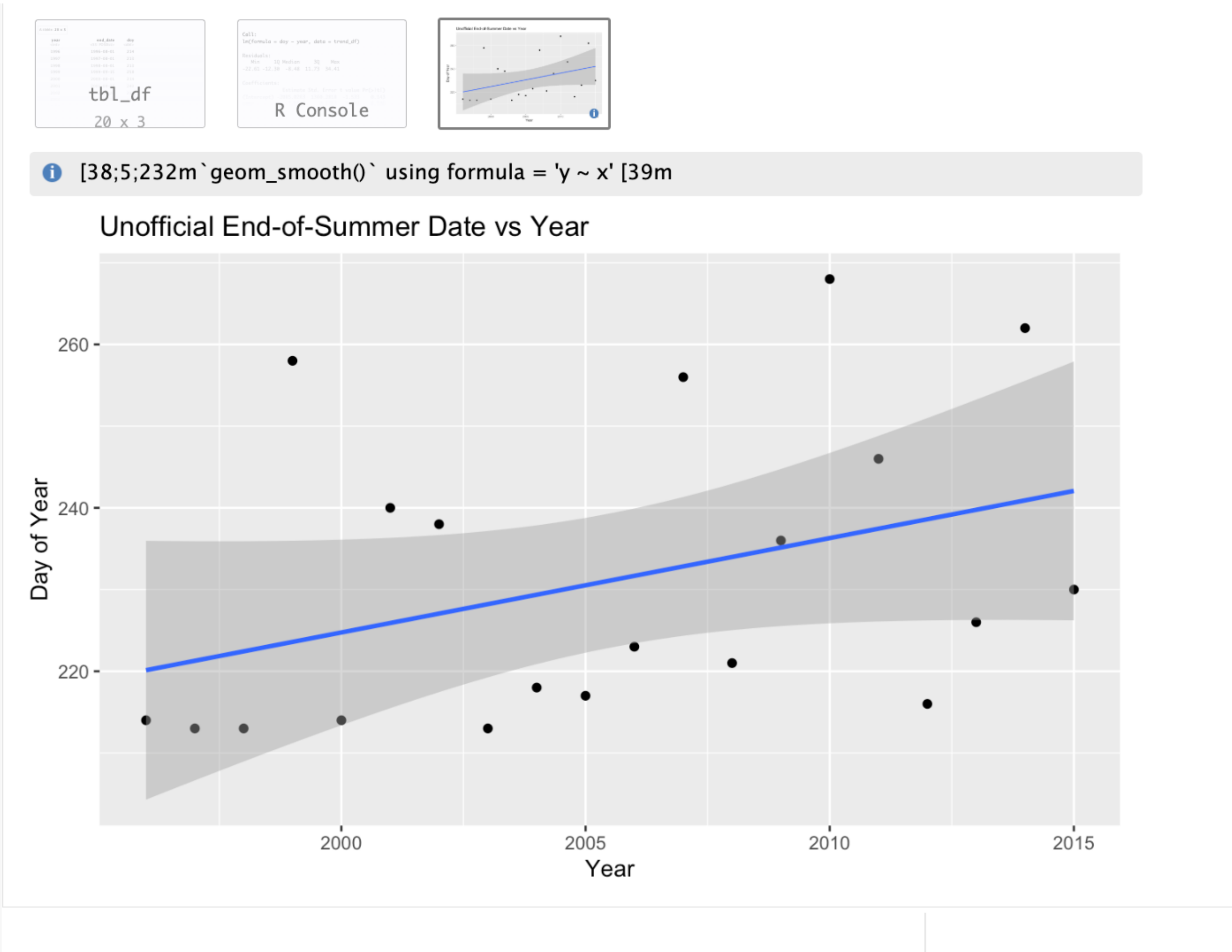
tbl_df 20 x 3 | R Console



```
Call:
lm(formula = doy ~ year, data = trend_df)

Residuals:
   Min     1Q Median     3Q    Max
-22.61 -12.30  -8.48  11.73  34.41

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2085.0263  1360.2818  -1.533    0.143
year            1.1549     0.6783   1.703    0.106

Residual standard error: 17.49 on 18 degrees of freedom
Multiple R-squared:  0.1387,    Adjusted R-squared:  0.09087
F-statistic: 2.899 on 1 and 18 DF,  p-value: 0.1058

tau = 0.34, 2-sided pvalue =0.040459
```

- Plotting DOY vs year shows an upward slope: the unofficial end of summer has been getting later.

Scatterplot with regression line (`ggplot`)



Discussion of Results

- The exponential smoother reduced day-to-day noise and revealed a consistent cooling-off signal across years.
- The results suggest summers are extending ~11 days later per decade.
- Some year-to-year variation remains (e.g., 1999 late drop vs 2003 earlier drop).
- Limitations: choice of δ = 2°F and persistence threshold affects sensitivity. Using δ = 1°F or 3°F produced qualitatively similar trends.
- Potential improvements: Holt-Winters with seasonal adjustment; change-point detection.
- Broader implication: later ends of summer are consistent with climate warming trends, with possible consequences for energy demand and health.

REFERENCES

https://www.linkedin.com/pulse/arima-vs-garch-time-series-forecasting-mahdi-navaei/