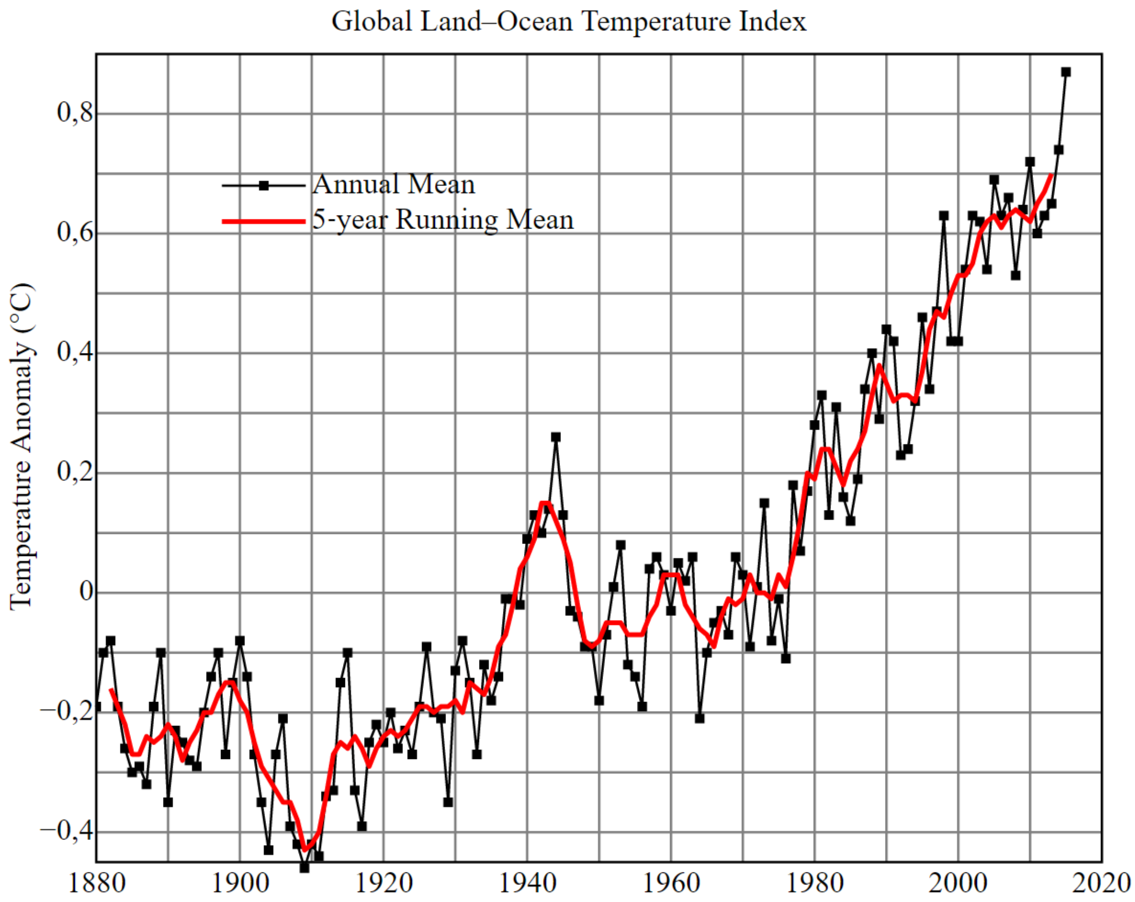
Time Series



STA303/STA1002: Methods of Data Analysis II, Summer 2016

Michael Guerzhoy

Time Series

- Measurements at regular time intervals:
  - $f(t = 1), f(t = 2), f(t = 3), f(t = 4), \ldots$
- The measurements are *not* independent: since the time series graph is very likely smooth, if $f(t = 200)$ is large, $f(t = 201)$ is likely large, but that tells us less about $f(t = 400)$
- Examples: t is year, f(t) is unemployment in year t. t is second, f(t) is the price of some stock at time t…

Case Study: Unemployment in the US, 1974-2004

- (the data, in R)

AR(1) model

- Response: Y
- $Y_t = \beta_0 + \beta_1 Y_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2)$ iid

AR(1) model – Model Checking

- It seems that the fake-data simulations capture a lot of what the actual data is like, but one thing to notice is that the real data is a lot smoother than the fake data
- Quantify that:
  - Count the number of times that an increasing trend switches to a decreasing trend in the real data (answer: 23)
  - In 1000 simulations, also count the number of switches
  - In 97% of the simulated datasets, there more than 23 switches
    - ⇒The model is not necessarily appropriate

AR(1) Model – Model Checking

- Note that the “smoothness” is just one property of the data – there are *lots* of ways in which the model could be wrong
- Just because we haven’t rejected the hypothesis that the model is correct, doesn’t mean that the model is correct

Using the AR(1) Assumption

- Response: Y_t Covariates: $X_{t,1}, X_{t,2}$
- If the AR(1) model is appropriate and $Y_t \approx \gamma_0 + \gamma_1 Y_{t-1}$, can transform the variables as follows:

$$V_t = Y_t - rY_{t-1}$$
$$U_{t,1} = X_{t,1} - rX_{t-1,1}$$
$$U_{t,2} = X_{t,2} - rX_{t-1,2}$$

Where $r = \frac{c_1}{c_0}, c_1 = \frac{1}{n-1}\Sigma_t res_t res_{t-1}, c_0 = \frac{1}{n-1}\Sigma_t res_t^2$

res_t is the t-th residual when regressing Y on X

Then, can try to fit the model

$$\mathrm{V}_t = U\beta + \epsilon$$

- Intuition: the *differences* between subsequent Y's can be independent even if the Y's are not dependent

Global Warming Acceleration

- (data, in R)

Is global warming accelerating?

- Naïve model:
  - $temp_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon$
- Idea: if $\beta_2 > 0$, we can fit a parabola to the data. If $\beta_2 = 0$, there is at most a linear data – the warming is not accelerating
- Problem: the residuals will not be independent
  - E.g., maybe there was a warm period 1950-1960 due to unusual solar activity, or a volcano going off. Then all the residuals for 1950-1960 will be positive

Is global warming accelerating?

- Assume the AR(1) model holds for the temperatures (check!)
- Now, use

$$V_t = Y_t - rY_{t-1}$$
$$U_{t,1} = t - r(t-1)$$
$$U_{t,2} = t^2 - r(t-1)^2$$

More intuition

- We transform the responses so that the correlation between adjacent covariates becomes 0
- Transform the predictors the same way we transform the response in order to run a regression like what we wanted to, but without the residuals being correlated