Preamble: Stationarity

When modeling time series, we pursue stationarity in the residuals, and usually one goes further and wants the residuals to be white noise. Why is this desirable? Here are some answers:

Stationarity: structure is stable over time - well-controlled.

- Differencing/variance-stabilizing transform: want result to be stationary
- Sets up use of stationary models like (S)ARIMA (will be covered later).

White noise: stationary, and its structure is "uninteresting".

- When modeling, usually want residuals to be this.
- "Uninteresting" because trivial covariance structure, 0 mean - no trend left to model.

Two simple examples of periodograms

1. (a) Let $y_t = 5\cos(2\pi(0.2)t) + 3\sin(2\pi(0.4)t)$, $t = 0, 1, \ldots, 49$. Plot the periodogram.
   Hint: For this exercise, you should see if you can calculate the periodogram from `fft`. *If you get stuck, You might find the function* `periodogram()` *in the TSA package or the* `pgram()` *function from in class helpful.*
   (b) Let $x_t = 5\cos(2\pi(0.23)t) + 3\sin(2\pi(0.37)t)$, $t = 0, 1, \ldots, 49$. Plot the periodogram for $\omega_j = 1/50, \ldots, 25/50$.
   (c) What is the difference between the two?

Indicator Variables

In class we called this a "nonparametric" seasonality method. We can pursue it that way, by simply taking the average of each $y_t, y_{t+d}, y_{t+2d}, \ldots$ for some period d. But it will be simpler for us to use indicator functions. For logical condition c,

$$I(c) = \begin{cases} 1 & \text{if } c \text{ is true} \\ 0 & \text{if } c \text{ is false} \end{cases}.$$

For example, if we wanted to indicate which month is January,

$$I(January) = \begin{cases} 1 & \text{if the month is January} \\ 0 & \text{if } c \text{ is false} \end{cases}.$$

Here we'll look at problem 2.1 in the book, slightly adjusted. For the Johnson & Johnson data (`JohnsonJohnson` in `R`), time t is in quarters $(1960.00, 1960.25, \ldots)$ so one unit of time is a year. Here we'll look at log-transformed earnings, i.e., $\log(Y_t)$.

2. (a) Fit the regression model

$$\log(Y_t) = \beta t + \alpha_1 Q_1(t) + \alpha_2 Q_2(t) + \alpha_3 Q_3(t) + \alpha_4 Q_4(t) + W_t.$$

   where βt is the linear trend, and $Q_i(t) = 1$ if time t corresponds to quarter i = 1, 2, 3, 4, and zero otherwise. The $Q_i(t)$'s are indicator variables. We will assume for now that w_t is a Gaussian white noise sequence.
   (b) If the model is correct, what is the estimated average annual increase in the logged earnings per share?
   (c) If the model is correct, does the average logged earnings rate increase or decrease from the third quarter to the fourth quarter?
   (d) What happens if you do or do not include an intercept term in the model in part 1?
   (e) Graph the data and superimpose the fitted values on the graph. Examine the residuals and state your conclusions. Does it appear that the model fits the data well (do the residuals look white)?

Birth Rate Data

The dataset *birth* (in the R package `astsa`) contains "monthly live births (adjusted) in thousands for the United States, 1948-1979."

3. Modeling trend and seasonality
   (a) Make a time series plot of the data, and note the sinusoidal patterns.
   (b) Make a periodogram of the data. Which spike is the tallest?
   (c) Use least squares to fit a model that estimates the trend using a sinusoid with frequency $1/n$.
   (d) Create a different model that fits the trend using a quartic (4th-order) polynomial.
   (e) Add both models' trend lines to the time series plot. How similar are they? What will be different about their forecasts?
   (f) Choose one of the models. How do the residuals look?
   (g) Add seasonality to your chosen linear model of the signal, in the form of indicator variables. The command `factor(time %% 12)` can be helpful.
   (h) Remake the time series plot with this new model's curve added, and also plot the residuals. How is this model doing (specifically, do the residuals look stationary)? What's keeping it from being (even) better?
4. Differencing
   (a) Now instead of modeling the trend and seasonality, we're removing their effects and trying to jump straight to a stationary-looking set of data. It typically makes more sense here to first take a seasonal difference, as that should eliminate both seasonal effects and linear trends. Thus, plot $\nabla_d Y_t$, the seasonal difference of the data.
   (b) Does this look stationary? If not, try some differences of your own choice. .
   (c) In lecture we mentioned $\nabla\nabla_d Y_t$ and $\nabla_d^2 Y_t$ both removed quadratic trends as well as seasonality. Which do you think looks better in this case?
   (d) Does differencing do a better job pursuing stationarity than our models above (i.e. does the differenced data look more stationary than the part 1 model's residuals?)

Airline Passengers

The dataset *AirPassengers* (native to R—can be called without loading) contains "monthly totals of international airline passengers, 1949 to 1960."

5. Variance stabilizing transform
   (a) Make a time series plot of the data, note the linear trend, and for now let's say we don't see the heteroscedasticity.
   (b) Create a model with a linear trend and monthly indicators.
   (c) Add the model's fitted values to the time series plot. Is there heteroscedasticity?
   (d) We will consider both the square root and log transforms. Plot both the log(AirPassengers) and sqrt(AirPassengers). Which looks more homoscedastic?
   (e) Add your chosen variance stabilizing transform to your model with linear trend and monthly seasonality.
   (f) Create two plots: the time series data with the model's fitted values, and the residuals. Have we fixed the i.e. are the residuals stable)?