1 Summary

1. Autocovariance, Autocorrelation and Partial Autocorrelation

   - Autocorrelation: the correlation between a time series and a time-shifted version of the time series
   - Autocovariance function (acvf)

     $$\gamma_k = E\left[(X_t - \mu_X)(X_{t+k} - \mu_X)\right]$$

     Note that for $k = 0$ we have

     $$\begin{aligned} \gamma_0 &= E\left[(X_t - \mu_X)(X_{t+0} - \mu_X)\right] \\ &= E\left[(X_t - \mu_X)(X_t - \mu_X)\right] \\ &= E\left[(X_t - \mu_X)^2\right] \\ &= \sigma^2 \end{aligned}$$

     That is, the autocovariance at lag $k = 0$ is σ^2
   - Autocorrelation function (acf)

     $$\rho_k = \frac{\gamma_k}{\sigma^2}$$

     For $k = 0$, $\gamma_0 = \sigma^2$ so that

     $$\rho_0 = \frac{\gamma_0}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$$

     That is, the autocorrelation at lag $k = 0$ is always 1
   - Autocorrelation graph (the correlogram)
     - x-axis: lags
     - y-axis: sample autocorrelation
     - If $\rho_k = 0$ then r_k is approximately Normal with mean $-\frac{1}{n}$ and variance $\frac{1}{n}$. The dotted lines on an autocorrelation graph are

       $$-\frac{1}{n} \pm 2 * \frac{1}{\sqrt{n}}$$

       If sample autocorrelations fall outside the graph, significant autocorrelation in that you would reject $H_0 : \rho_k = 0$
   - Partial autocorrelation
     - Removes the effect of correlations at shorter lags

       $$COR\left[X_t, X_{t+k} | X_{t+1} \ldots X_{t+k-1}\right]$$

     - If a process is AR(p), the autocorrelation at lag p is the p^{th} coefficient and helps to identify the model order

2. Stationary
   - Let X be a random process. X is stationary if the joint distribution of $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ is the same as the joint distribution of $X_{t_1-k}, X_{t_2-k}, \ldots, X_{t_n-k}$ for all choices of $t_1, t_2, \ldots, t_n$ and k
   - The joint probability distribution does not change over time
3. Weakly Stationary
   - The mean and variance functions, $\mu(t)$ and $\sigma^2(t)$ are constant
   - The autocovariance only depends on the time shift: $\gamma_{t,t+k} = \gamma_k$
4. Models for data generation
   - White Noise: $w_t \sim N(0, \sigma^2)$
   - MA(q): $x_t = w_t + \beta_1 w_{t-1} + \ldots + \beta_q w_{t-q}$
   - AR(p): $x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_p x_{t-p} + w_t$
   - Random Walk: $x_t = x_{t-1} + w_t$
   - Random Walk with Drift: $x_t = \mu + x_{t-1} + w_t$
5. Smoothing
   - Regression Models/Smoothers: $x_t = m_t + s_t + w_t$
     - Polynomial: $m_t = \sum_{i=0}^{p} \beta_i t^i$
     - Harmonic seasonal (periodic): $s_t = \sum_{i=1}^{\lfloor \frac{s}{2} \rfloor} \left(s_i sin \left(2\pi i \frac{t}{s} \right) + c_i cos \left(2\pi i \frac{t}{s} \right) \right)$
       * s is the number of periods
       * i is changing the frequency
       * s_i and c_i are the coefficients to be estimated, think about them as the β coefficients from the indicator variable representation
       * Note only half the waves have to be estimated
     - s_i can also be seasonal indicators
   - Kernel Smoothers: locally-weighted averaging
   - LOWESS: locally-weighted polynomial regression
   - Spline: polynomial regression on disjoint time buckets, penalty for complexity

2 Approval Ratings

The fiel *Bush.csv* contains the approval ratings for President Bush from 2001-2004.

1. Read the file and convert the *Approval* column to a *ts* object.
2. Make a times series plot. Is there any trend or seasonality? Are there any sudden shocks?
3. Make an ACF plot. Describe how the approvals depend on one another. If these data were an MA process, what does this suggest the MA order should be?
4. Make a PACF plot. Describe how the approvals depend on one another. If these data were an AR process, what does this suggest the AR order should be?
5. Obtain two smooth fits using a kernel smoother. Plot the fits and the original series on the same graph.
6. Obtain two smooth fits using a lowess smoother. Plot the fits and the original series on the same graph.
7. Obtain two smooth fits using a spline smoother. Plot the fits and the original series on the same graph.
8. Obtain two smooth fits using a regression smoother that include seasonal indicators or a harmonic component. Plot the fits and the original series on the same graph.
9. Put all four approaches on one graph sheet? Which do you like the most and why?