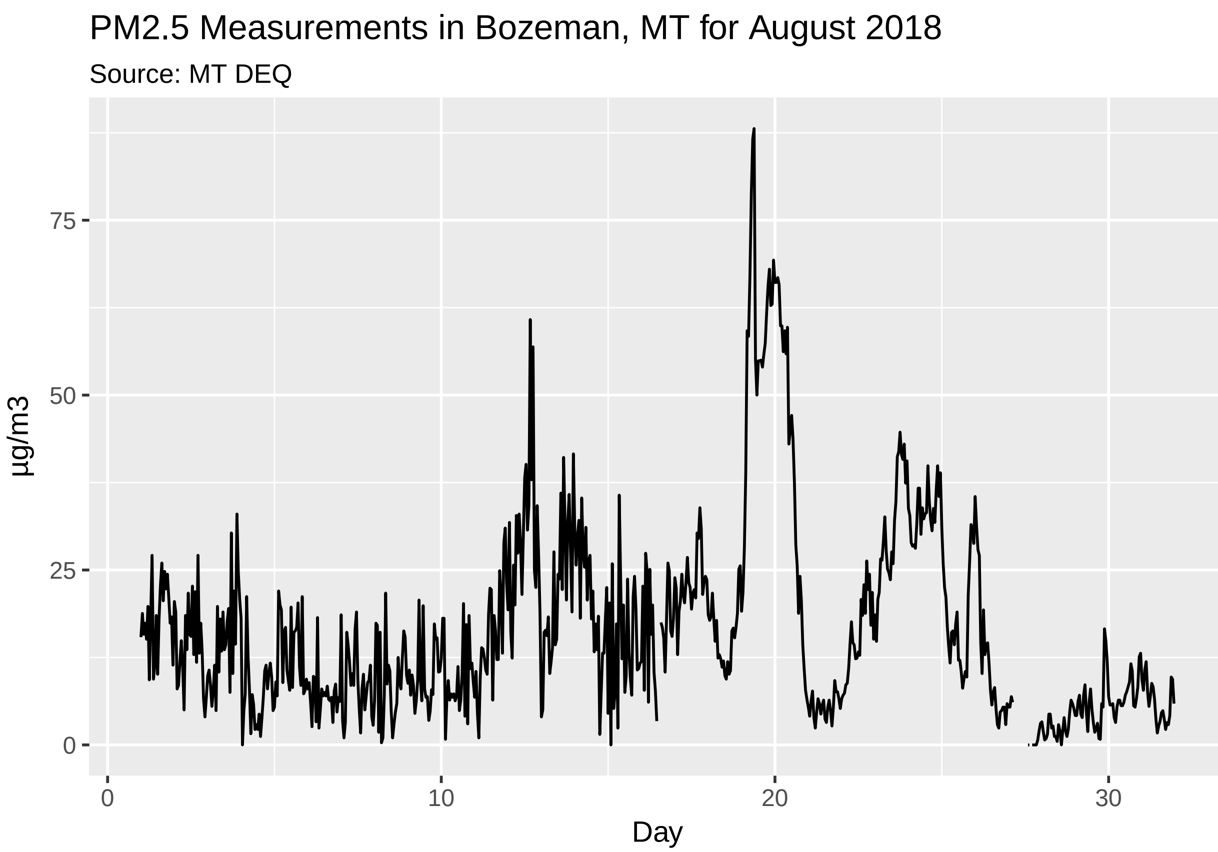
STAT 436 / 536 - Lecture 5

September 12, 2018

Forecasting Strategies

Consider an hourly time series of particulate matter less than 2.5 micrometers from Bozeman, MT in August of 2018.



- What information would be useful to predict the PM2.5 measurments on September 1? How about the PM2.5 measurements for tomorrow?

  - *Make sure to think about this as an interval rather than a point estimate*

  - *One strategy would be to identify the other variables that might be influential to air pollution in Bozeman. For instance, we might consider fire growth in the Pacific Northwest the prior day and the wind direction.*

  - *Another strategy might be to estimate trends and patterns based on the existing time series*

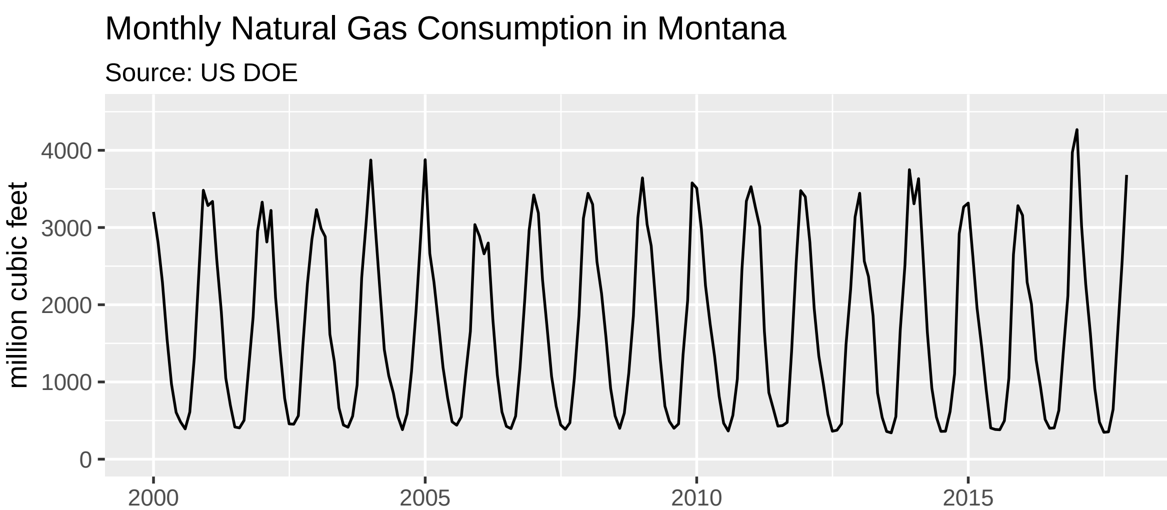
  - *On September 1st, the hourly measurements are between 2.3 and 10.4.*

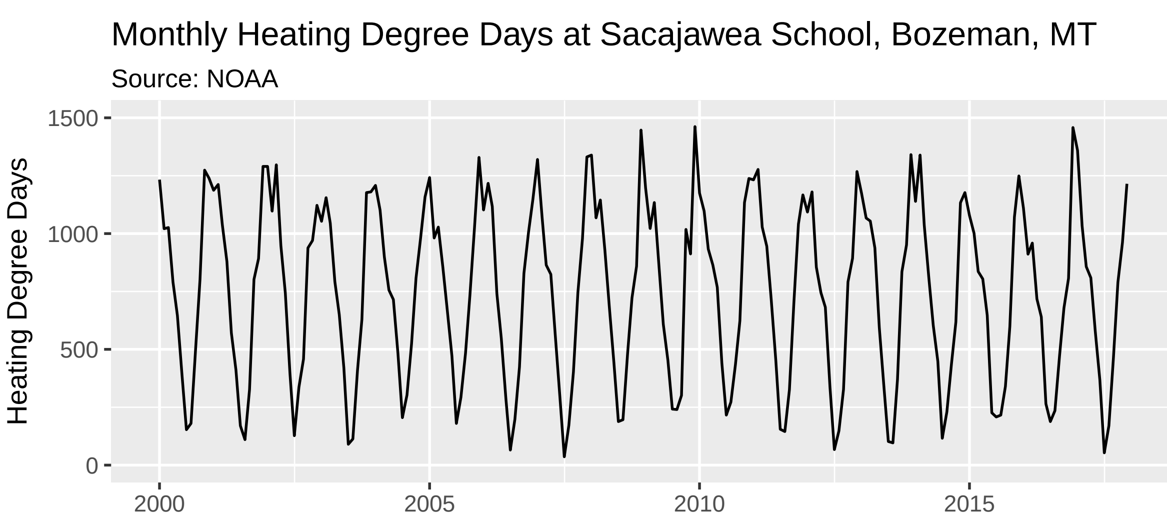
Leading Variables and Associated Variables

- In some situations another variable may be useful informing our forecasts. A common example of this is in the housing industry where building approvals would be a *leading indicator* of building activity.

- Consider natural gas consumption and heating degree days (HDD). *HDD = the difference between the daily average temperature and 65 degrees Fahrenheit.*

- What do you assume is the relationship between the heating degree days during a given month and the reported natural gas consumption?





Cross-Correlation

- Suppose two times series models, say x and y, are stationary with a constant mean and variance. The variables, x and y, may be serially correlated, perhaps with different time lags.

- The cross covariance function can be written as: $\gamma_k(x, y) = E[(x_{t+k} - \mu_x)(y_t - \mu_y)]$.

- Q: Is the cross-correlation function symmetric? In other words, does $\gamma_k(x,y) = \gamma_k(y,x)$?

- The sample cross-correlation function (ccf) can be written as $\rho_k(x,y) = \frac{\gamma(x,y)}{\sigma_x \sigma_y}$.

- Similarly to single variable autocorrelation calculations, sample based computations can be made:
  - Sample cross covariance: $c_k(x,y) = \frac{1}{n}\sum_{t=1}^{n-k}(x_{t+k} - \bar{x})(y_t - \bar{y})$
  - Sample cross correlation: $r_k(x,y) = \frac{c_k(x,y)}{\sqrt{c_0(x,x)c_0(y,y)}}$

- Note the cross correlation and cross covariance functions are defined on the de-seasonalized and de-trended time series, so it will not capture similar patterns in those structures. See HW2 Q1 for a demonstration.

```
acf(ts.union(ts.weather,ts.consumption))
```

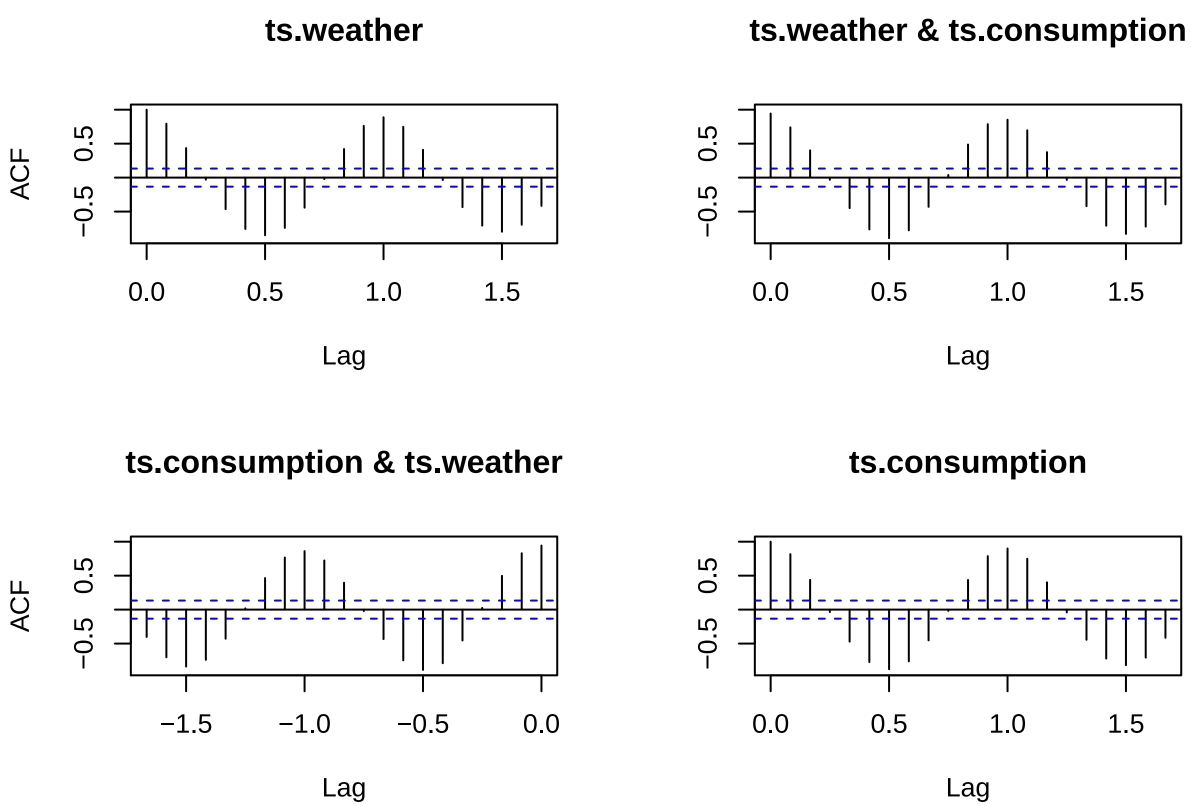


Figure 1: No Decomposition of the Time Series

```
hdd.random <- decompose(ts.weather)$random %>% window(start=c(2001,1), end = c(2016,12))
gas.random <- decompose(ts.consumption)$random %>% window(start=c(2001,1), end = c(2016,12))
acf(ts.union(hdd.random,gas.random))
```

- *While there may not be strong autocorrelation structure between the two variables, that does not mean one is not useful for the other. In this case, it means that part of the model is useful for modeling the mean structure, but does not give correlated errors.*

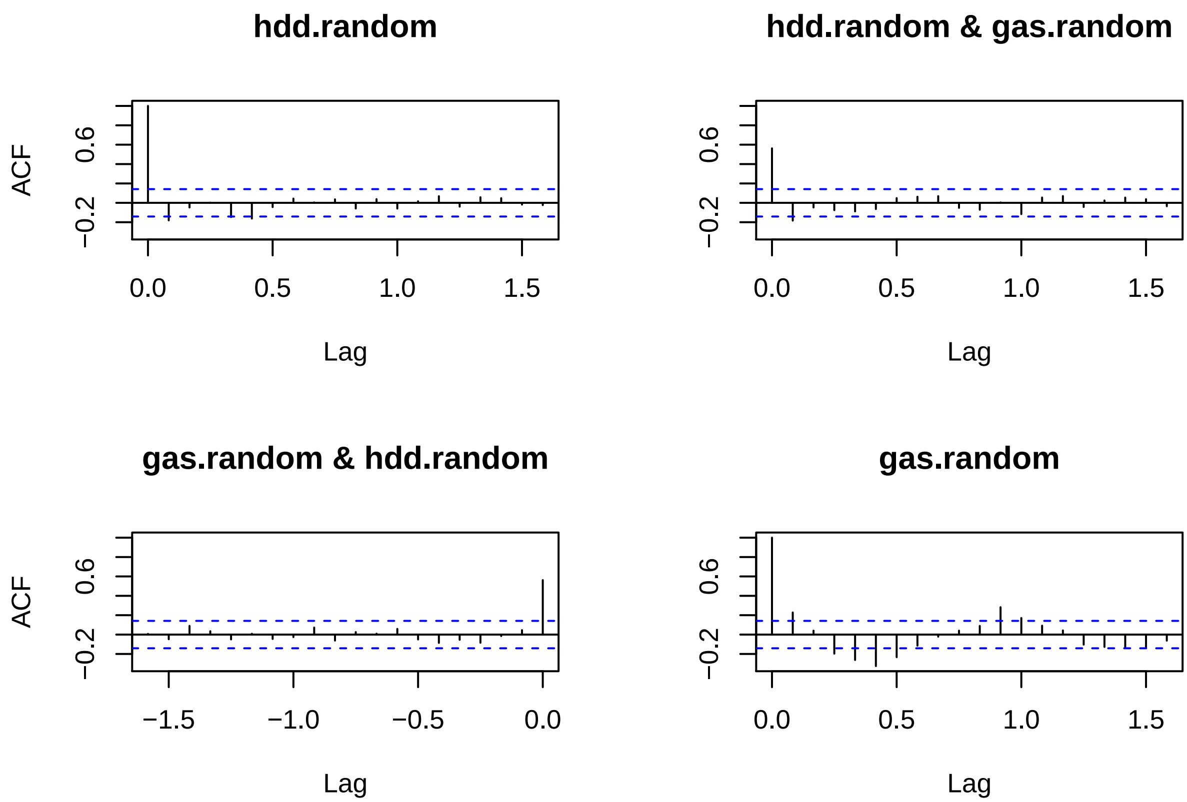


Figure 2: Using the decomposed random component