STAT 436 / 536 - Lecture 14: Key

State Space Models

- Most statistical analysis focuses on static parameter estimation; in other words, all of the values of the parameters are considered fixed across time. This is true for regression models, but what about ARIMA models and exponentially weighted moving average models?

- The inherent assumption in these models is that the relationships are constant. Sometimes there are fundamental processes that might change in time such that the relationship between the variables and the response are not constant. Consider the relationship between housing prices and square footage.

- State-space models and more specifically, Dynamic Linear Models (DLMs) can model parameter values with dynamics. Furthermore, these models also contain all of the static frameworks that we have previously studied.

  A state space model is typically expressed using two levels, the observation equation and state equation:

1. *Observation Equation:* The observation equation is focused on the data generating mechanism for the data you actually observe.

2. *State or Evolution Equation:* This captures the dynamics of a unobservable (latent) variable. The state vector contains the relevant information for summarizing the past behavior and is used to model the future as a function of the past.

   Consider the example:

$$\begin{aligned} y_t &= \theta_t + \epsilon_t \quad \text{(Observation Equation)} \\ \theta_t &= \theta_{t-1} + w_t \quad \text{(Evolution Equation)}, \end{aligned}$$

   where $\epsilon_t \sim N(0, \sigma^2)$ and $w_t \sim N(0, \sigma^2_w)$. Note that ν, σ^2, and σ^2_w are all assumed known, but they can easily be estimated in an MCMC procedure.

Dynamic Linear Models

- Now consider the following representation:

$$\begin{aligned} x_t &= \tilde{F}^T \tilde{\theta}_t + v_t \\ \tilde{\theta}_t &= G \tilde{\theta_{t-1}} + w_t, \end{aligned}$$

  where $\tilde{\theta}_t$ is known as the state vector for time t, $\tilde{F}$ is vector of known constants or regressors, v_t is observation noise, G is the state evolution matrix which is typically assumed known, and w_t is the innovation noise.

- Dynamic linear models are a broad class of models that incoporate many existing models we have seen in this class.

- What do we have if: G is an identity matrix, and $w_t = 0 \; \forall t$? Thus $\tilde{\theta}_t = \tilde{\theta}_{t-1} = \tilde{\theta}$, so this can be re-written as

$$x_t = \tilde{F}^T \tilde{\theta} + v_t$$

  , what if $F = X$ and $\tilde{\theta} = \tilde{\beta}$? This is just standard regression.

\- Recall an autoregressive model can be written as

$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + w_t$$

\- How can we write an AR model in the DLM framework? Starting from the last model, let $\tilde{F} = (x_{t-1}, \ldots x_{t-p})$ and $\tilde{\theta} = (\alpha_1, \ldots, \alpha_p)$ Thus $\tilde{\theta}_t = \tilde{\theta}_{t-1} = \tilde{\theta}$, so this can be re-written as

$$x_t = \tilde{F}^T \tilde{\theta} + v_t.$$

\- Now what would it mean if w_t is non-zero in the two previous models? *There would be time-varying coefficients*

Exponential Smoothing

- The exponential smoothing techniques (Holt-Winters) can be re-imagined in a state-space framework. In particular there is an entire book on this subject. *Forecasting with Exponential Smoothing: The State Space Approach* by Hyndman, Koehler, Ord and Snynder is a Spring book with a free e-link through the library.

- In this book the exponential smoothing techniques are referred to by the following three components:
  - T (Trend): Long-term direction of the series
  - S (Seasonal): Pattern that repeats with a known periodicity
  - E (Error): Unpredictable component of the series.

- The methods are even generalized beyond what we have seen. For example consider the trend component, which is a combination of the level term (mean or intercept) and the growth term (slope). An additional term ($0 < \phi < 1$) is introduced to fit a dampening pattern. Let T_h be the forecast trent over the next h time periods.

1. $T_h = l$ No trend

2. $T_h = l + bh$ Additive

3. $T_h = l + (\phi + \phi^2 + \cdots + \phi^h)b$ Additive, damped

4. $T_h = lb^h$ Multiplicative

5. $T_h = lb^{(\phi+\phi^2+\cdots+\phi^h)}$ Multiplicative, damped

- Furthermore, these can be generalized and combined with three levels of the seasonal component (None, Additive, Multiplicative) to create a set of 15 models. These 15 models would incorporate all of the exponential smoothing models we have seen thus far: No trend and no seasonal component would be the simple exponential smoothing method, additive and additive is the Holt-Winters 3-parameter model.

- Recall that the exponential smoothing techniques do not have a probabilistic interpretation (For creating confidence intervals), but rather only produce point estimates:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$$

or

$$\begin{aligned}
\hat{y}_{t+h} &= l_t + b_t h \\
l_t &= \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \\
b_t &= \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}
\end{aligned}$$

- We now consider the state-space representations for a one-step ahead prediction with Holt's Linear Method:

$$\begin{aligned}
\hat{y}_{t+1} &= l_t + b_t + \epsilon_t \\
l_t &= l_{t-1} + b_{t-1}\alpha\epsilon_t \\
b_t &= b_{t-1} + \beta(l_t - l_{t-1} - b_{t-1}) = b_{t-1} + \alpha\beta\epsilon_t
\end{aligned}$$

where $\epsilon_t = y_t - \mu_t$.

- Using standard notation, let

$$\begin{aligned}
y_t &= [1\ \ 1]x_{t-1} + \epsilon_t, \\
x_t &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_{t-1} + \begin{bmatrix} \alpha \\ \alpha\beta \end{bmatrix} \epsilon_t,
\end{aligned}$$

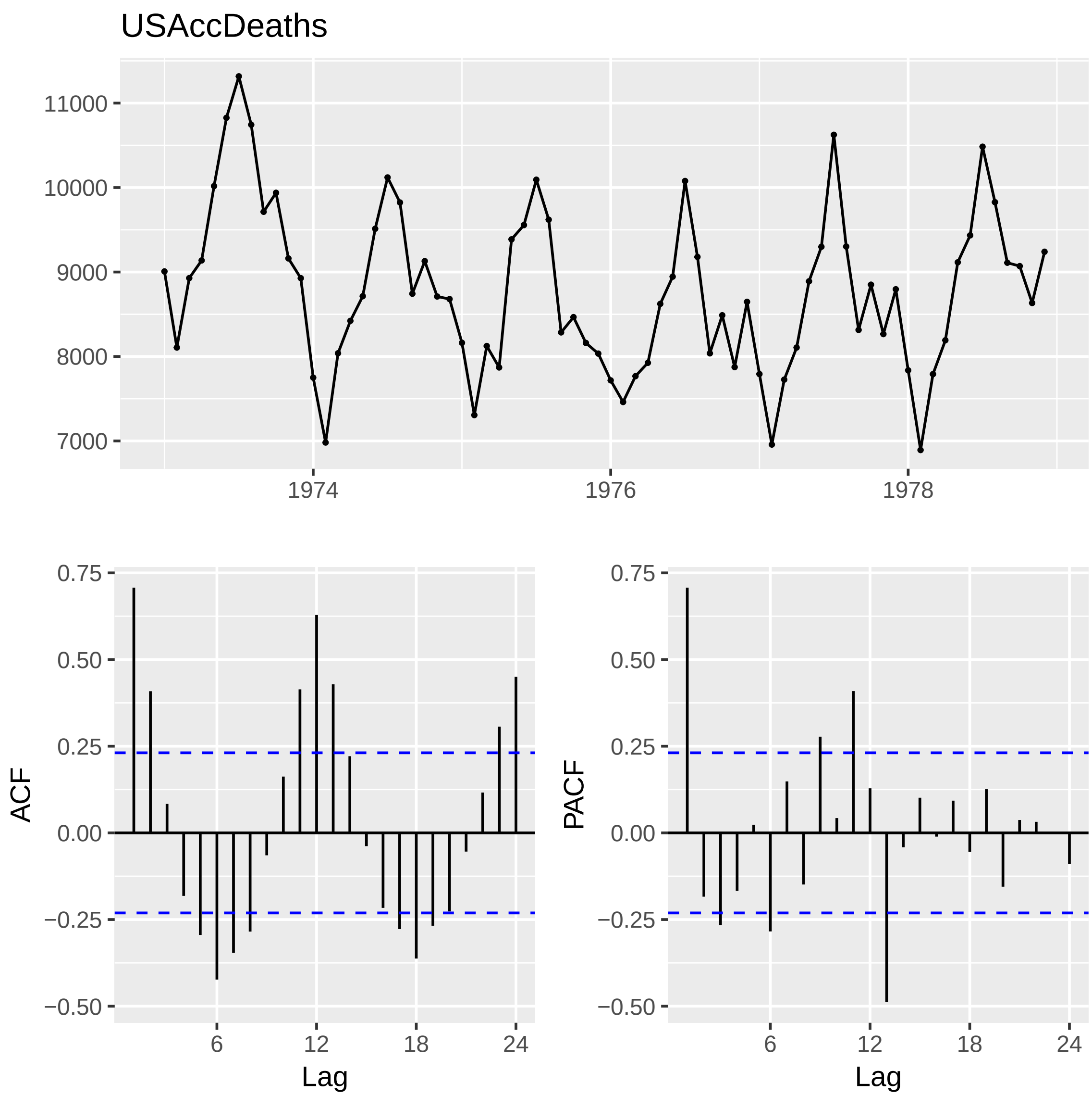
where the state vector $x_t = (l_t, b_t)$.

- Once a distribution is specified for ϵ_t, the model is fully specified and can be estimated. Typically, $\epsilon_t \sim N(0, \sigma^2)$

State Space Model Estimation

- The major benefit of the state-space model framework in this situation is a likelihood-based approach that permits uncertainty calculation and likelihood-based model comparison.
- The state-space framework requires initial values for the state components, say x_0. From a Bayesian perspective, this is a prior distribution.
- With only a single source of error, $\sigma_t \sim N(0, \sigma^2)$, Kalman filtering equations are not required in this setting.
- The `forecast` package contains the `ets` function for fitting smoothing models with a state-space framework.
- The `ets` function permits specifying a model directly using: A (additive), M (multiplicative), and N (none). Additionally, the model can be automatically selected.

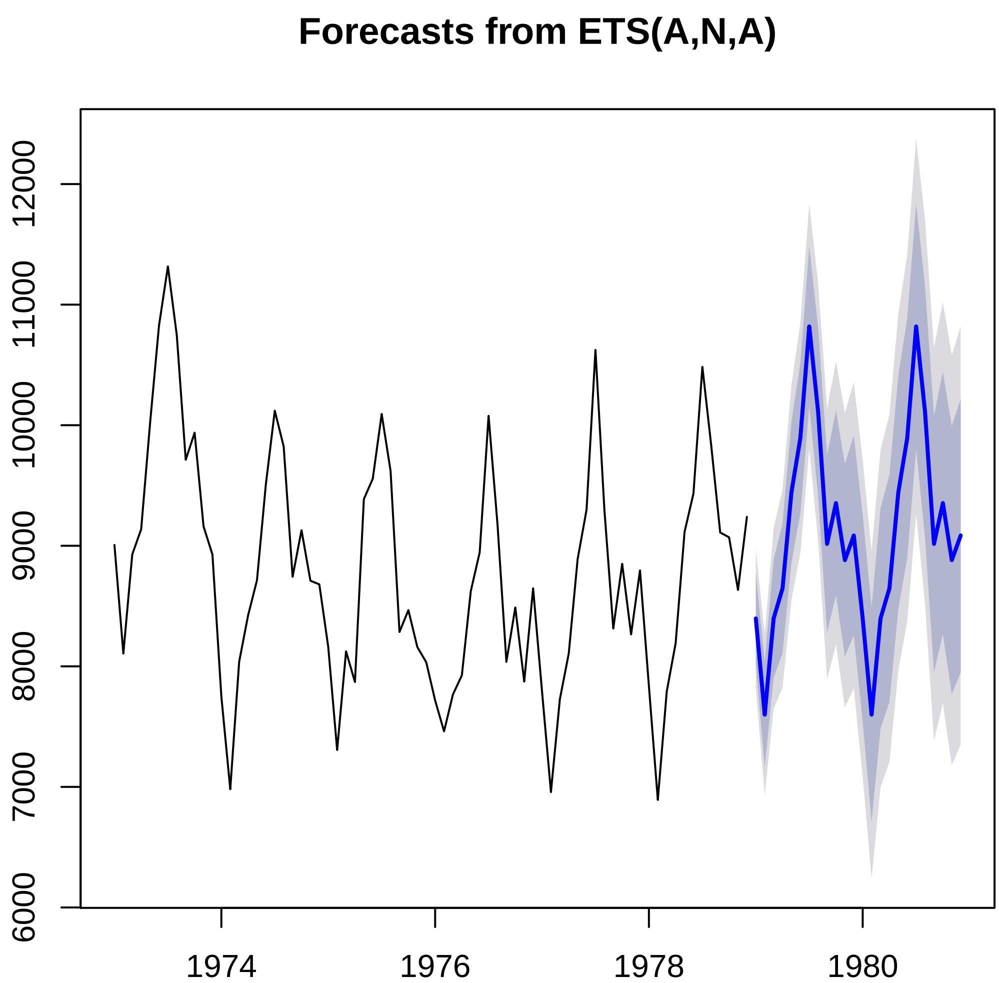
```
library(forecast)
ggtsdisplay(USAccDeaths)
```




```
fit <- ets(USAccDeaths); fit
```

```
## ETS(A,N,A)
##
## Call:
##  ets(y = USAccDeaths)
##
##   Smoothing parameters:
##     alpha = 0.5946
##     gamma = 0.002
##
##   Initial states:
##     l = 9248.3628
##     s = -51.3449 -255.3528 218.2901 -121.771 970.7387 1683.237
##            756.092 306.4212 -489.5627 -739.9004 -1537.792 -739.0552
##
##   sigma:  292.6907
##
##      AIC     AICc      BIC
## 1140.145 1148.716 1174.295
```

```
plot(forecast(fit))
```



- Look at the Bozeman Air Quality data for July 3 - 31, 2018 and fit an ets model.

```r
library(readr)
library(dplyr)
library(rvest)
scrape_BZNPM_Jul <- function(days){
  # Scrapes hourly PM2.5 readings in Bozeman for July 2018
  # inputs: sequence of days
  # outputs: data frame that contains day, hour, and hourly average PM2.5 concentration.
  smoke.df <- data.frame(day = NULL, hour = NULL, conc = NULL)

  for (d in days){
    bzn_aq <- read_html(paste("http://svc.mt.gov/deq/todaysair/AirDataDisplay.aspx?siteAcronym=BH&targe
    daily.smoke <- cbind(rep(d,24),html_table(bzn_aq)[[1]][,1:2])
    colnames(daily.smoke) <- c('day','hour','conc')
    daily.smoke$conc[daily.smoke$conc == 'DU'] <- 'NA'
    daily.smoke$conc <- as.numeric(daily.smoke$conc)
    smoke.df <- bind_rows(smoke.df,daily.smoke)
  }
  return(smoke.df)
}
air.quality <- scrape_BZNPM_Jul(3:31)
```

Exponential Smoothing Models with Regression Parameters

- One drawback of the exponential smoothing techniques is that regressor variables typically cannot be included in the model.

- Regressor variables can take two forms:
  - Explanatory variables such as the number of cups of coffee sold the previous day, or the weather the day before.
  - Intervention variables, which typically take the form of an indicator variables, for things like special holidays or extreme weather events. In the taxi data set some of the lowest use days were major snow storms in NYC.

- The general state-space model for the exponential smoothing framework can be written as:

$$\begin{aligned} y_t &= wx_t + \epsilon_t \\ x_t &= Fx_{t-1} + g\epsilon_t \end{aligned}$$

  now consider the extension

$$\begin{aligned} y_t &= wx_t + z\theta + \epsilon_t \\ x_t &= Fx_{t-1} + g\epsilon_t, \end{aligned}$$

  where z are regression variables and θ are the associated covariates. The `ets` function does not permit variables, but instead users are encouraged to consider the DLM implementation of the ARIMA model framework.