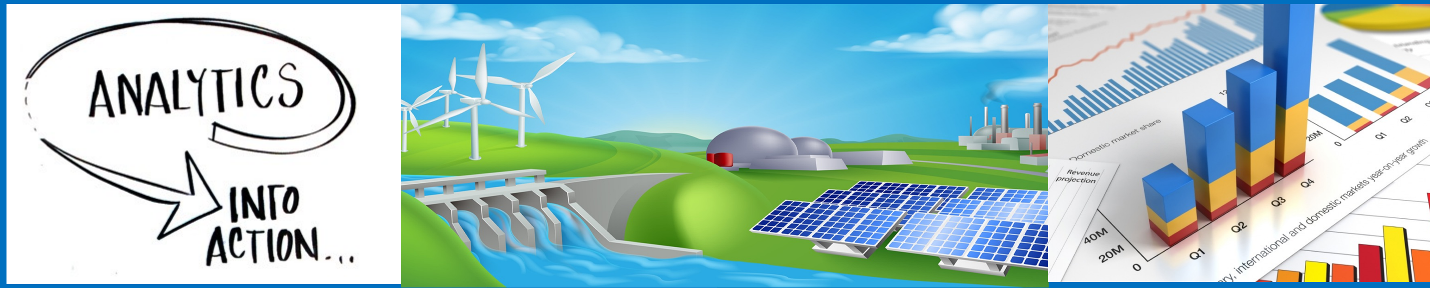


ENV797 – TIME SERIES ANALYSIS FOR ENERGY AND ENVIRONMENT APPLICATIONS
M10 - State Space Models / Bayesian Framework

Prof. Luana Medeiros Marangon Lima, Ph.D.

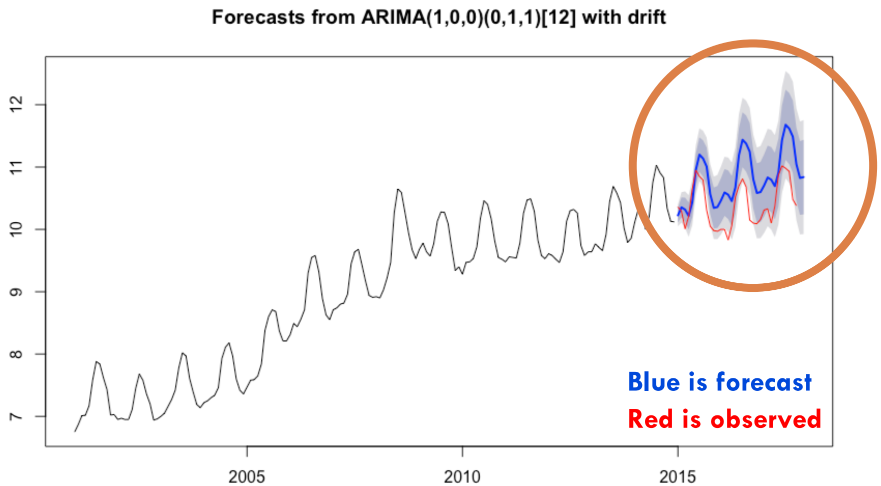
Nicholas School of the Environment - Duke University

Learning Goals

- Discuss state space models in general
- Discuss Bayesian Framework
- Learn about local level, linear trend and basic structure models
- See how models we are familiar with fit under state space framework (linear regression, exponential smoothing)
- Implement state space model in R

What's next?

What's next?



- After you have done the simple model, you start thinking about ways to improve it
- Other forecasting techniques can help make the model more accurate

- Suppose the variable is highly dependent on other factors, such as weather, holidays, time of the day, etc. One could try fitting time series models that allow for inclusion of other predictors using methods such
  - ARMAX (still a linear regression but with exogenous predictors)

$$y_t = \boldsymbol{\beta_1 x_t} + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

More general models

- State Space models
  - Model not only the variable but also the coefficients
  - Bayesian approach to state space models
    - Ex: Dynamic Linear Models

State Space Models

State Space Models

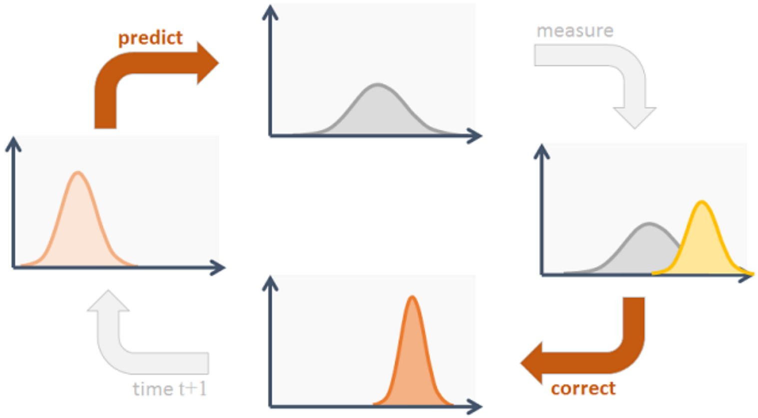
- The State Space Model approach offers a very **general and powerful framework** to operate with time series data
- Models with time-varying parameters can be created
- Classical linear regression is embedded as special case
- **ARIMA-model** class is also a special case

State Space Models (cont'd)

- State Space models allow **decomposition of a time series** into relevant components – trend, cycle, seasonal
- And analyze each in real-time (filter), to infer best historical estimates (smoothing), and to **forecast all components** as well as the original series

State Space Models (cont'd)

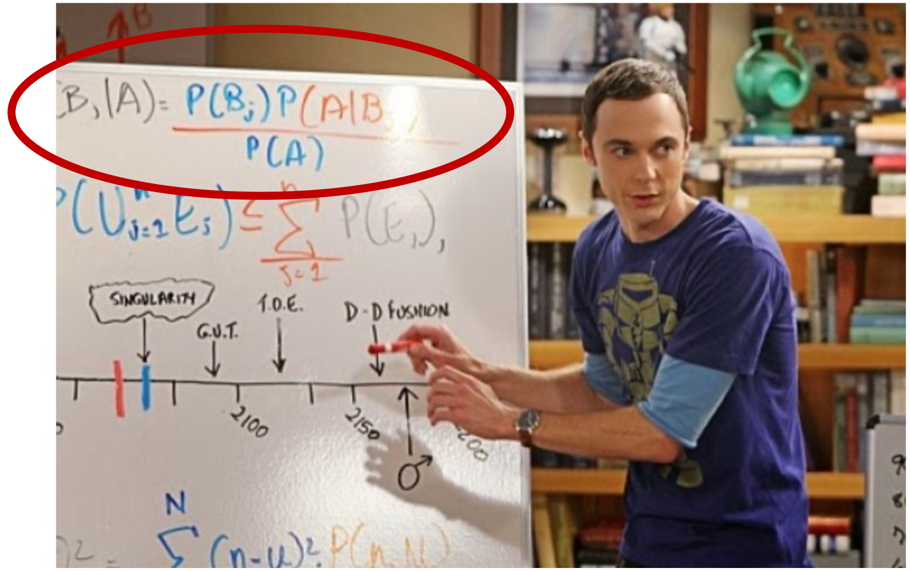
- Main estimation algorithm, Kalman Filter, is set-up as recursive form



- Approach allows to include a priori knowledge through a suitable Bayesian formulation of the initial state vector

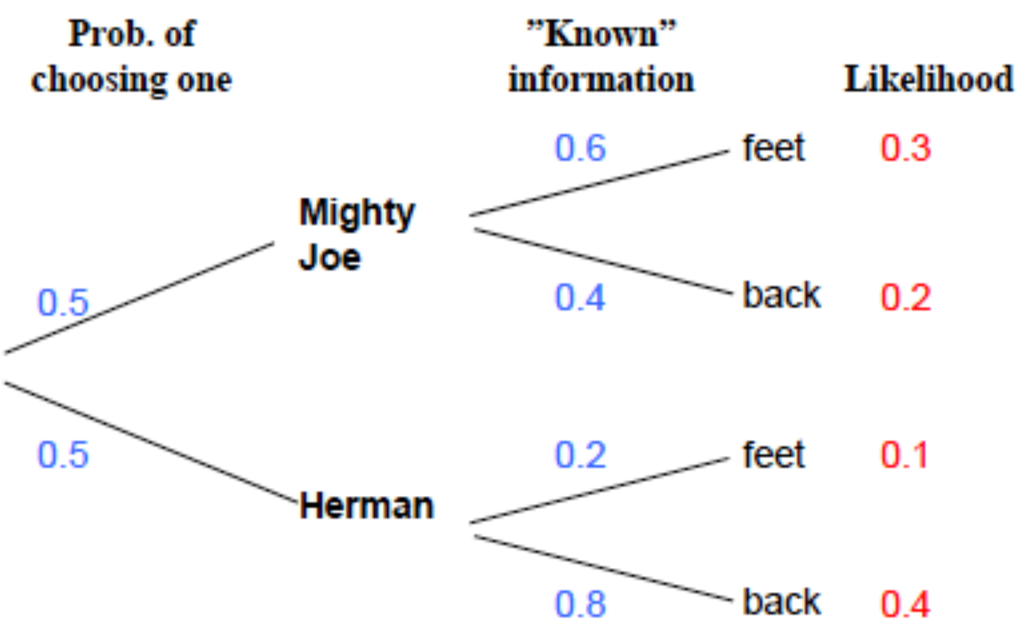
A little background: Bayesian Statistics

- Frequentist versus Bayesian statistic
- Different way of thinking
- Instead of asking: What is the **likelihood of this data point given the model** (frequentist), the Bayesian ask: What is the **likelihood of the model given this data point?**



Simple Example (Bayesian approach)

- Two frogs: Joe ($P(feet) = 0.6$) and Herman ($P(feet) = 0.2$)
- Pick one frog and "jump" it. It lands on its feet. What is the probability it is Joe?



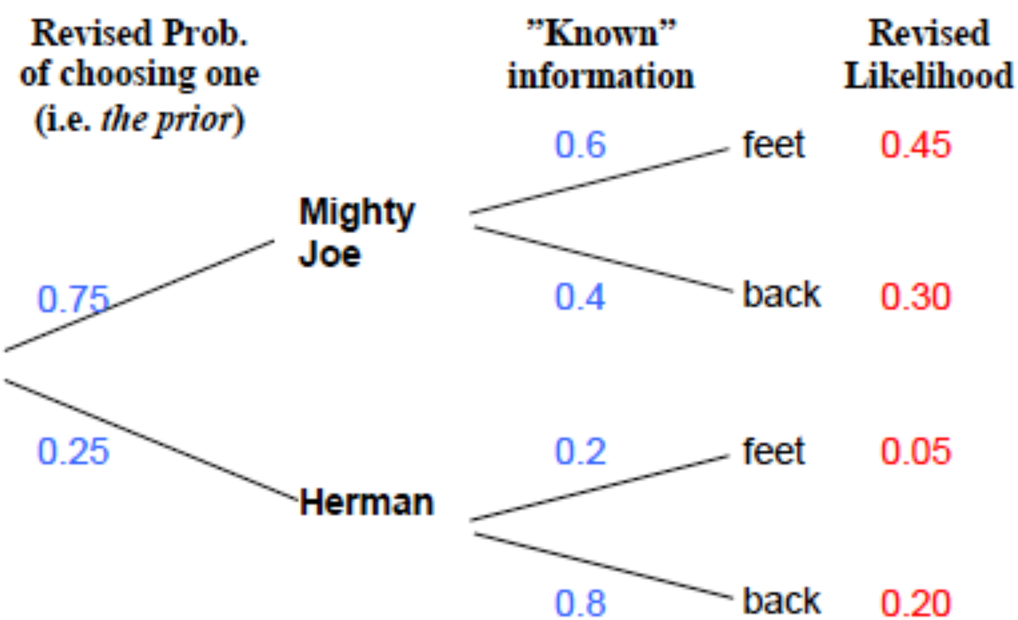
$$P(Joe|feet) = \frac{P(Joe \cap feet)}{P(feet)}$$

$$P(Joe|feet) = \frac{0.3}{0.3 + 0.1}$$

$$P(Joe|feet) = 0.75 \text{ or } 75\%$$

Simple Example (cont'd)

- We "jump" the frog again. It lands on its feet one more time
- How can we update our beliefs of whether or not this is Joe?



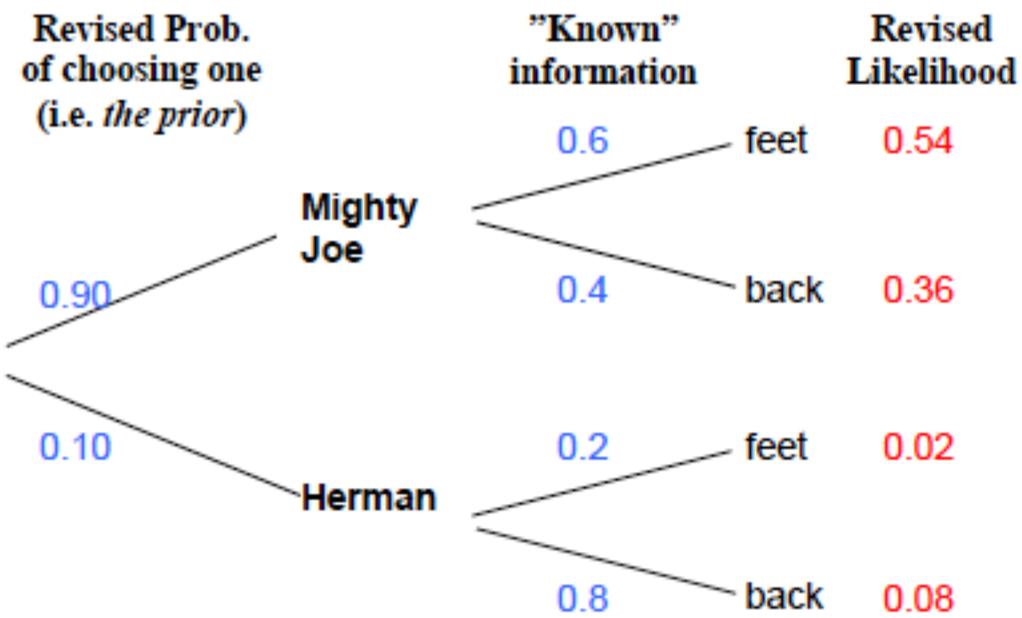
$$P(Joe|feet) = \frac{P(Joe \cap feet)}{P(feet)}$$

$$P(Joe|feet) = \frac{0.45}{0.45 + 0.05}$$

$$P(Joe|feet) = 0.90 \text{ or } 90\%$$

Simple Example (cont'd)

- We "jump" the frog a third time. Now it lands on its back.
- How can we update our beliefs of whether or not this is Joe?



$$P(Joe|back) = \frac{P(Joe \cap back)}{P(back)}$$

$$P(Joe|back) = \frac{0.36}{0.36 + 0.08}$$

$$P(Joe|back) = 0.82 \text{ or } 82\%$$

Simple Example (Frequentist approach)

- Find which distribution you could use to model the problem
  1. Continuous or discrete?
     - Discrete
  2. Possibilities: Binomial, Poisson, Discrete uniform,… Which one to choose?
     - For the frog example: Binomial – 2 possible outcomes, experiment is repeated multiple times
  3. What is the binomial distribution formula?

Simple Example (Frequentist approach)

- Binomial distribution has probability mass function:

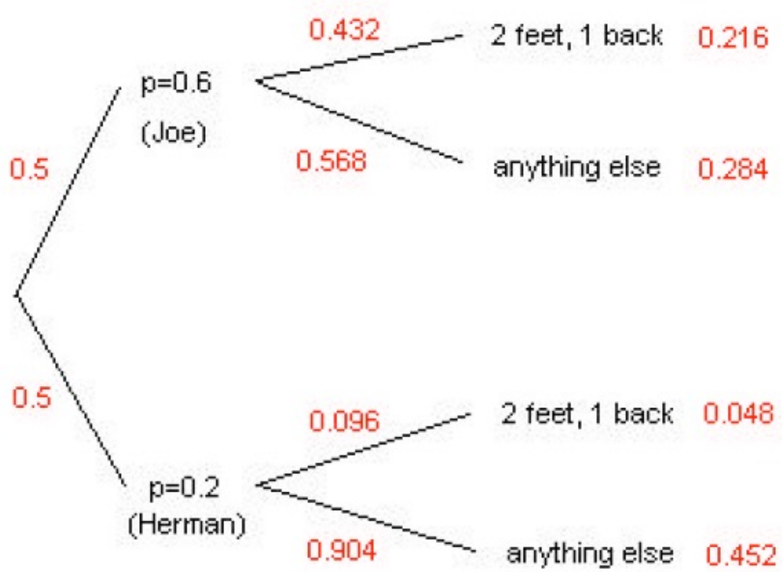
$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$P(2F,1B|Joe)=\binom{3}{2}(0.6)^2(0.4)^1=0.432$$

$$P(2F,1B|Herman)=\binom{3}{2}(0.2)^2(0.8)^1=0.096$$

k is the probability of success, here we assume frog lands on its feet as success, n is the number of trials, p is the probability of success

- Probability tree



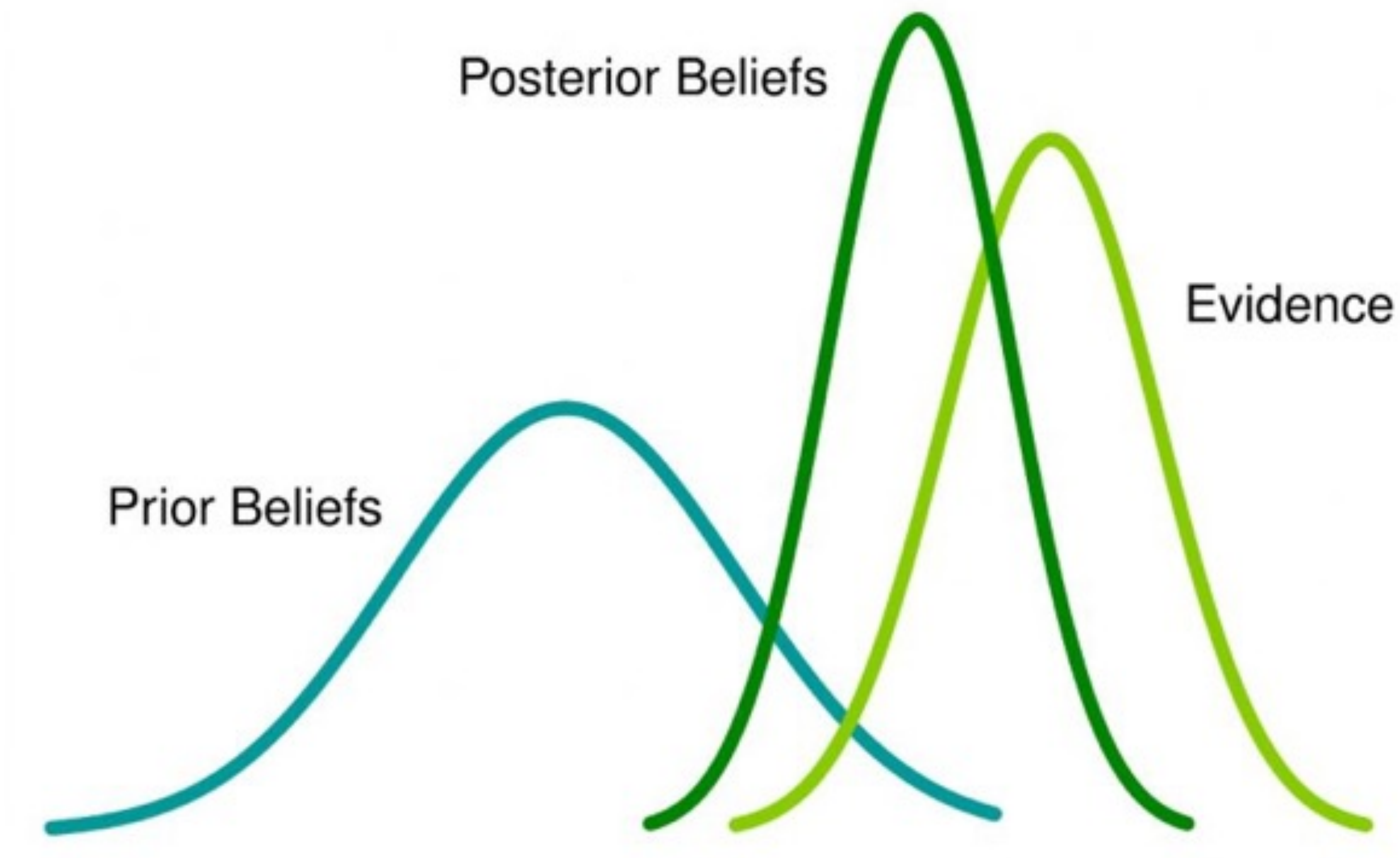
$$P(Joe|2F,1B)=\frac{P(Joe\cap 2F,1B)}{P(2F,1B)}$$

$$P(Joe|2F,1B)=\frac{0.216}{0.216+0.048}$$

$$P(Joe|2F,1B)=0.82 \text{ or } 82\%$$

Same results, just a different approach!

Bayesian Framework



Back to State Space Models

Time Series Decomposition

- Classical decomposition of a time series

  $y_t = \mu_t + \gamma_t + \varepsilon_t, \qquad t = 1, \ldots, n$

  where

  y_t: observation

  μ_t: slowly changing (trend) component

  γ_t: seasonal component

  ε_t: error term (random)

- Component can be determinist (ARIMA) or functions of time (stochastic)

Local Level (LL) Model

- Deterministic Example

  $y_t = \mu + \varepsilon_t$ where ε_t are $i.i.d$ and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ or $\varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2)$

- Stochastic Example

  $y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2)$

  If μ_t is a function of time, we also need a model for μ_t

  $\mu_{t+1} = \mu_t + \eta_t \qquad \eta_t \sim \mathcal{NID}(0, \sigma_\eta^2)$

  - What do we need to specify?

  Parameters: $\mu_1 \sim N(a, P)$ and $\sigma_\varepsilon^2, \sigma_\eta^2$

Local Linear Trend (LLT) Model

- Stochastic Example

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2)$$
$$\mu_{t+1} = \beta_t + \mu_t + \eta_t \qquad \eta_t \sim \mathcal{NID}(0, \sigma_\eta^2)$$
$$\beta_{t+1} = \beta_t + \xi_t \qquad \xi_t \sim \mathcal{NID}(0, \sigma_\xi^2)$$

- What do we need to specify?

Parameters: prior distributions μ_1, β_1 and $\sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\xi^2$

- Special cases:
  - If $\sigma_\eta^2 = \sigma_\xi^2 = 0$, the trend is a straight line with slope β_1 and intercept μ_1
  - If $\sigma_\eta^2 = 0 \; and \; \sigma_\xi^2 > 0$, the trend is a smooth curve

Local Trend with Seasonality Model

- Also known as Basic Structural Model or BSM

$$y_t = \mu_t + \gamma_t + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2)$$
$$\mu_{t+1} = \beta_t + \mu_t + \eta_t \qquad \eta_t \sim \mathcal{NID}(0, \sigma_\eta^2)$$
$$\beta_{t+1} = \beta_t + \xi_t \qquad \xi_t \sim \mathcal{NID}(0, \sigma_\xi^2)$$
$$\gamma_{t+1} = -\sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t \qquad \omega_t \sim \mathcal{NID}(0, \sigma_\omega^2)$$

- s denote number of seasons
- What do we need to specify?

  Parameters: $\text{prior distributions } \mu_1, \beta_1, \gamma_1, \dots, \gamma_s$ and $\sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\xi^2, \sigma_\omega^2$
- Special case
  - If $\sigma_\omega^2 = 0$, the seasonal component is determinist, i.e., dummy coefficient's do not change over time

Exponential Smoothing under SS

- Model equations will be

Observation equation $$y_t = l_t \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2)$$

State equation $$l_t = l_{t-1} + \alpha \varepsilon_t$$

Linear Regression State Space Model

Observation equation

$$Y_t = F_t'\theta_t + v_t$$
$$v_t \sim \mathcal{NID}(0, \sigma_v^2)$$

State equation

$$\theta_{t+1} = G_{t+1}\theta_t + w_{t+1}$$
$$w_{t+1} \sim \mathcal{NID}(0, \sigma_w^2)$$

- What do we need to specify?

$$\{F_t, G_t, \sigma_v^2, \sigma_w^2\}$$

  - Common approach is to make G_t constant and equal to identity matrix (random walk)

- What about priors?
  - Find initial parameters for θ_0 based on initial information

$$\theta_0/D_0 \sim N(m_0, C_0)$$

  - Ex. Use the initial observations and run linear regression to get prior for the mean and variance of coefficients

State Space Forecasting in R

package "smooth"

- Exponential Smoothing in State Space Model
- Similar to the function *ses()* but with time varying parameters

  es(data, model = "ZZZ", persistence = NULL, phi = NULL, initial = c("optimal", "backcasting"), initialSeason = NULL, ic = c("AICc", "AIC", "BIC"), cfType = c("MSE", "MAE", "HAM", "MSEh", "TMSE", "GTMSE", "MSCE"), h = 10, holdout = FALSE, cumulative = FALSE, intervals = c("none", "parametric", "semiparametric", "nonparametric"), level = 0.95, intermittent = c("none", "auto", "fixed", "interval", "probability", "sba", "logistic"), imodel = "MNN", bounds = c("usual", "admissible", "none"), silent = c("all", "graph", "legend", "output", "none"), xreg = NULL, xregDo = c("use", "select"), initialX = NULL, updateX = FALSE, persistenceX = NULL, transitionX = NULL, ...)

- *model* three-character string
  - The first letter denotes the error type ("A", "M" or "Z");
  - the second letter denotes the trend type ("N","A","M" or "Z"); and
  - the third letter denotes the season type ("N","A","M" or "Z").

  "N"=none, "A"=additive, "M"=multiplicative and "Z"=automatically selected

State Space Forecasting in R

- General class

package "stats"

StructTS(x, type = c("level", "trend", "BSM"), init = NULL, fixed = NULL, optim.control = NULL)

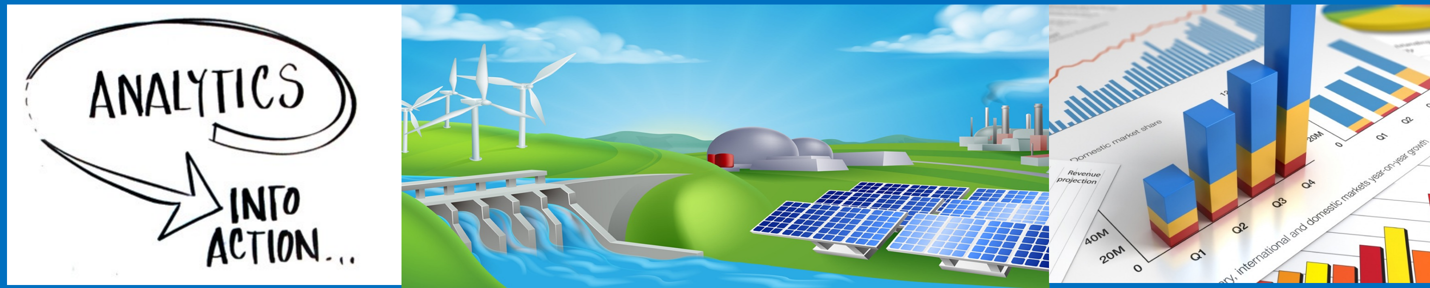
- *Structural time series* models are (linear Gaussian) state-space models for (univariate) time series based on a decomposition of the series into the components
  - *type ="level"* local level model
  - *type ="trend"* local linear trend model
  - *type ="BSM"* basic structural model, i.e., local trend with seasonal component

StrucTS() explained

StrucTS(data, type="BSM", fixed=c(0.1,0.001,NA,NA))

$$y_t = \mu_t + \gamma_t + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2)$$
$$\mu_{t+1} = \beta_t + \mu_t + \eta_t \qquad \eta_t \sim \mathcal{NID}(0, \sigma_\eta^2)$$
$$\beta_{t+1} = \beta_t + \xi_t \qquad \xi_t \sim \mathcal{NID}(0, \sigma_\xi^2)$$
$$\gamma_{t+1} = -\sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t \qquad \omega_t \sim \mathcal{NID}(0, \sigma_\omega^2)$$

- Fixed argument refers to variances
  - ***fixed=c***$(\sigma_\eta^2\,, \sigma_\xi^2, \sigma_\omega^2, \sigma_\varepsilon^2,)$



THANK YOU !

luana.marangon.lima@duke.edu