Time Series Data Analysis Using R

- Introduction to R
- Getting Started - Using RStudio IDE
  - R 3.3.x
  - RStudio 1.0.xxx
- On Line Data Resources
  - quantmod
  - Quandl

Time Series Data

- Economic Time Series Data
  - GDP, CPI, Oil Price, Climate Change
  - Interest Rates, Exchange Rates
- Financial Time Series Data
  - S&P 500, VIX (Fear Index)
  - International Stock Markets
- High Frequency Time Series

$$\{y_t\} = \{\ldots, y_1, y_2, \ldots, y_n, \ldots\}$$

Time Series Data

- Decomposition
  - Additive Components $y_t = m_t + s_t + \varepsilon_t$
  - Multiplicative Components $\log(y_t) = m_t + s_t + \varepsilon_t$
    - Deterministic vs. Stochastic Trend (and/or Seasonality)

$$y_t \text{ or } \log(y_t) = \alpha + \beta t + \gamma x_t + ... + \varepsilon_t$$

- Transformation
  - Stationary vs. Non-stationary Time Series
    - Box-Cox or Log -> Stationarity in the Variance
    - Difference -> Stationarity in the Mean

Time Series Decomposition

- decompse
  - `function (x, type = c("additive", "multiplicative"), filter = NULL)`
- stl
  - `function (x, s.window, s.degree = 0, t.window = NULL, t.degree = 1, l.window = nextodd(period), l.degree = t.degree, s.jump = ceiling(s.window/10), t.jump = ceiling(t.window/10), l.jump = ceiling(l.window/10), robust = FALSE, inner = if (robust) 1 else 2, outer = if (robust) 15 else 0, na.action = na.fail)`

Strategies for Time Series Analysis

- Data Exploration
  - Using Graphs
- Hypothesis Testing
  - Normality
  - Stationarity
  - Serial Correlation
    - Durbin-Watson
    - Box-Pierce / Ljung-Box
    - ACF/PACF
- Time Series Smoothing
  - Exponential Smoothing
  - Structural Time Series
- Model Estimation
  - Regression Model
  - ARIMA Model
  - Dynamic Linear Model

Time Series Forecasting

- One-Step Ahead Forecasts

$$\hat{y}_{t+1|t} = E(y_{t+1} \mid I_i), \quad I_t = \{..., y_1, ..., y_t\}$$

- h-Step Ahead Forecasts

$$\hat{y}_{n+h|n} = E(y_{n+h} \mid I_n), \quad I_n = \{y_1, ..., y_n\}, h = 1, 2, ...$$

- Forecast Evaluation
  - Training (Estimation) vs. Testing (Forecast)
  - Cross Validation: Rolling Forecasts

Time Series Forecasting

- Simple Forecast
  - Random Walk
    - Naive Forecast $\hat{y}_{t+1|t} = y_t, \quad t = 1, 2, \ldots, n$
    - Naïve Seasonal Forecast $\hat{y}_{n+h|n} = y_n, \quad h = 1, 2, \ldots$

$$\hat{y}_{t+1|t} = y_{t+1-p}, \qquad t = p, p+1, \ldots, n$$

$$\hat{y}_{n+h|n} = y_{n+h-pk}, \quad h = 1, 2, \ldots; k = [(h-1)/p] + 1$$

  - Random Walk with Drift

$$\hat{y}_{n+h|n} = y_n + \frac{h}{n-1}\sum_{t=2}^{n}(y_t - y_{t-1})$$

Time Series Forecasting

- Regression-Based Forecast
  - Deterministic Trend and Seasonal Forecasts

$$y_t = \alpha + \beta t + [\gamma t^2 + \sum_{s=2}^{p} \delta_s D_{s,t}] + \varepsilon_t$$

$$\hat{y}_{t+1|t} = \hat{\alpha} + \hat{\beta}(t+1), \quad t = 1, 2, \ldots, n, \ldots$$

  - Stochastic Trend Forecasts

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t, \quad |\rho| < 1$$

$$\hat{y}_{t+1|t} = \hat{\alpha} + \hat{\rho} y_{t-1}, \quad t = 1, 2, \ldots, n, \ldots$$

$$(if \;\; y_{n+h} \; is \; known \;\; for \; h = 1, 2, \ldots)$$

Time Series Forecasting

- One-Step Ahead Forecast Error

$$\hat{\varepsilon}_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t}, \quad t = p, p+1, \ldots, n \; (p = seasonal \;\; period)$$

- Forecast Error Statistics

$$MAE = mean_{t=p,\ldots,n}(|\, \hat{\varepsilon}_{t+1|t} \,|)$$

$$MAPE = 100 \; mean_{t=p,\ldots,n}(|\, \hat{\varepsilon}_{t+1|t} \,/\, y_{t+1} \,|)$$

$$MSE = mean_{t=p,\ldots,n}(\hat{\varepsilon}^2_{t+1|t})$$

$$RMSE = \sqrt{mean_{t=p,\ldots,n}(\hat{\varepsilon}^2_{t+1|t})}$$

$$RMSPE = 100\sqrt{mean_{t=p,\ldots,n}[(\hat{\varepsilon}_{t+1|t} \,/\, y_{t+1})^2]}$$

Time Series Forecasting

- h-Step Ahead Forecast Error (if y_{n+h} is known)

$$\hat{\varepsilon}_{n+h|n} = y_{n+h} - \hat{y}_{n+h|n}, \quad h = 1, 2, \ldots$$

- Forecast Error Statistics

$$MAE = mean_h(|\hat{\varepsilon}_{n+h|n}|)$$

$$MAPE = 100\ mean_h(|\hat{\varepsilon}_{n+h|n} / y_{n+h}|)$$

$$MSE = mean_h(\hat{\varepsilon}^2_{n+h|n})$$

$$RMSE = \sqrt{mean_h(\hat{\varepsilon}^2_{n+h|n})}$$

$$RMSPE = 100\sqrt{mean_h[(\hat{\varepsilon}_{n+h|n} / y_{n+h})^2]}$$

Time Series Forecasting

- Using R Package `forecast`
  - rwf
    - `function (y, h = 10, drift = FALSE, level = c(80, 95), fan = FALSE, lambda = NULL, biasadj = FALSE, x = y)`
  - accuracy
    - `function (f, x, test = NULL, d = NULL, D = NULL)`

Time Series Smoothing

- Exponential Smoothing

$$y_t = m_t + s_t + \varepsilon_t \ or \ \log(y_t) = m_t + s_t + \varepsilon_t$$

$$trend: m_t = a_t + b_t \ (level + slope)$$

$$seasonal: s_t \quad random: \varepsilon_t$$

  – Simple Exponential Smoothing (EWMA)

$$b_t = 0, s_t = 0$$

  – Holt Exponential Smoothing

$$s_t = 0$$

  – Holt-Winters Exponential Smoothing

Time Series Smoothing

- Simple Exponential Smoothing (EWMA)
  - Forecast Equation

$$\hat{y}_{t+1|t} = a_t, \quad t = 1, 2, \ldots, n$$

$$\hat{y}_{n+h|n} = a_n, \quad h = 1, 2, \ldots$$

  - Smoothing Equation

$$a_t = \alpha y_t + (1-\alpha) a_{t-1}, \quad 0 < \alpha < 1$$

$$= \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \ldots$$

Time Series Smoothing

- Simple Exponential Smoothing (EWMA)
  - State Space Representation

$$y_t = a_{t-1} + e_t \quad (\textit{observation equation})$$

$$a_t = a_{t-1} + \alpha e_t \quad (\textit{state equation})$$

  - Model Estimation

$$\hat{\alpha} = \min_{\alpha} \arg \sum_{t=1}^{n} e_t^2, \textit{ with initials } a_0$$

$$\textit{where} \quad e_t = y_t - a_{t-1}, \quad t = 1, 2, \ldots, n$$

Time Series Smoothing

- Holt Exponential Smoothing
  - Forecast Equation

$$\hat{y}_{t+1|t} = a_t + b_t, \quad t = 1, 2, \ldots, n$$

$$\hat{y}_{n+h|n} = a_n + hb_n, \quad h = 1, 2, \ldots$$

  - Smoothing Equation

$$a_t = \alpha y_t + (1-\alpha)(a_{t-1} + b_{t-1}), \quad 0 < \alpha < 1$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}, \quad 0 < \beta < 1$$

Time Series Smoothing

- Holt Exponential Smoothing
  - State Space Representation

$$y_t = a_{t-1} + b_{t-1} + e_t$$

$$a_t = a_{t-1} + b_{t-1} + \alpha e_t$$

$$b_t = b_{t-1} + \alpha\beta e_t$$

  - Model Estimation

$$(\hat{\alpha}, \hat{\beta}) = \min_{(\alpha,\beta)} \arg \sum_{t=1}^{n} e_t^2, \; with\ initials\ (a_0, b_0)$$

$$where \quad e_t = y_t - (a_{t-1} + b_{t-1})$$

Time Series Smoothing

- Holt-Winters Exponential Smoothing
  - Forecast Equation

$$\hat{y}_{t+1|t} = a_t + b_t + s_{t+1-p}, \quad t = 1, 2, \ldots, n$$

$$\hat{y}_{n+h|n} = a_n + hb_n + s_{n+1-p+[(h-1)\,\mathrm{mod}\,p]}, \quad h = 1, 2, \ldots$$

$$\left(\hat{y}_{n+h|n} = a_n + hb_n + s_{n+h-p}, \quad h = 1, 2, \ldots \leq p\right)$$

  - Smoothing Equation

$$a_t = \alpha(y_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}), \quad 0 < \alpha < 1$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}, \quad 0 < \beta < 1$$

$$s_t = \gamma(y_t - a_t) + (1-\gamma)s_{t-p}, \quad 0 < \gamma < 1$$

Time Series Smoothing

- Holt-Winters Exponential Smoothing
  - State Space Representation

$$y_t = a_{t-1} + b_{t-1} + s_{t-p} + e_t$$

$$a_t = a_{t-1} + b_{t-1} + \alpha e_t$$

$$b_t = b_{t-1} + \alpha\beta e_t$$

$$s_t = s_{t-p} + (1-\alpha)\gamma e_t$$

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \min_{(\alpha,\beta,\gamma)} \arg \sum_{t=1}^{n} e_t^2, \textit{ with initials } (a_0, b_0, s_0, \ldots, s_{2-p}, s_{1-p})$$

$$\textit{where} \quad e_t = y_t - (a_{t-1} + b_{t-1} + s_{t-p})$$

Time Series Smoothing

- One-Step Ahead Forecast at t = p, p+1,…

$$\hat{y}_{t+1|t} = a_t + b_t + s_{t+1-p}$$

$$\hat{y}_{p+1|p} = a_p + b_p + s_1$$

$$\hat{y}_{p+2|p+1} = a_{p+1} + b_{p+1} + s_2$$

$$\ldots$$

$$\hat{y}_{2p|2p-1} = a_{2p-1} + b_{2p-1} + s_p$$

$$\hat{y}_{2p+1|2p} = a_{2p} + b_{2p} + s_{p+1}$$

$$\ldots$$

$$\hat{y}_{n|n-1} = a_{n-1} + b_{n-1} + s_{n+1-p}$$

Initialization

$$\hat{y}_{1|0} = a_0 + b_0 + s_{1-p}$$

$$\hat{y}_{2|1} = a_1 + b_1 + s_{2-p}$$

$$\ldots$$

$$\hat{y}_{p|p-1} = a_{p-1} + b_{p-1} + s_0$$

Forecast Error

$$\hat{\varepsilon}_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t}$$

$$t = p, p+1, \ldots, n$$

Time Series Smoothing

- h-Step Ahead Forecast at t = n

$$\hat{y}_{n+h|n} = a_n + hb_n + s_{n+h-p}, \quad h = 1, 2, \ldots \le p$$

$$\hat{y}_{n+h|n} = a_n + hb_n + s_{n+[(h-1) \bmod p]-p}, \quad h > p$$

$$Note: [(h-1) \bmod p] = 0, 1, 2, \ldots p \;\; for \; h = 1, 2, \ldots$$

- Forecast Error (if y_{n+h} is known)

$$\hat{\varepsilon}_{n+h|n} = y_{n+h} - \hat{y}_{n+h|n}$$

$$\hat{y}_{n+h|n} = a_n + hb_n + s_{n+1-p+[(h-1)\bmod p]}$$

$$h = 1: \quad \hat{y}_{n+1|n} = a_n + b_n + s_{n+1-p}$$

$$h = 2: \quad \hat{y}_{n+2|n} = a_n + 2b_n + s_{n+2-p}$$

...

$$h = p: \quad \hat{y}_{n+p|n} = a_n + pb_n + s_n$$

$$h = p+1: \hat{y}_{n+p+1|n} = a_n + (p+1)b_n + s_{n+1-p}$$

$$h = p+2: \hat{y}_{n+p+1|n} = a_n + (p+2)b_n + s_{n+2-p}$$

...

$$n = 2p: \quad \hat{y}_{n+h|n} = a_n + 2pb_n + s_n$$

...

$$\hat{y}_{2p|2p-1} = a_{2p-1} + b_{2p-1} + s_p$$

$$\hat{y}_{2p+1|2p} = a_{2p} + b_{2p} + s_{p+1}$$

...

Time Series Smoothing

- HoltWinters
  - `function (x, alpha = NULL, beta = NULL, gamma = NULL, seasonal = c("additive", "multiplicative"), start.periods = 2, l.start = NULL, b.start = NULL, s.start = NULL, optim.start = c(alpha = 0.3, beta = 0.1, gamma = 0.1), optim.control = list())`
- Predict.HoltWinters
  - `predict(object, n.ahead = 1, prediction.interval = FALSE, level = 0.95, ...)`

Time Series Smoothing

- Using R Package `forecast`:
- ets
  - `function (y, model = "ZZZ", damped = NULL, alpha = NULL, beta = NULL, gamma = NULL, phi = NULL, additive.only = FALSE, lambda = NULL, biasadj = FALSE, lower = c(rep(1e-04, 3), 0.8), upper = c(rep(0.9999, 3), 0.98), opt.crit = c("lik", "amse", "mse", "sigma", "mae"), nmse = 3, bounds = c("both", "usual", "admissible"), ic = c("aicc", "aic", "bic"), restrict = TRUE, allow.multiplicative.trend = FALSE, use.initial.values = FALSE, ...)`

Time Series Smoothing

- Using R Package `forecast`:
- forecast
  - `function(object, h=ifelse(frequency(object$x)>1, 2*frequency(object$x),10), level=c(80,95), fan=FALSE, lambda=NULL, biasadj=FALSE,...)`

Structural Time Series Model

- Linear Gaussian State Space Model
  - Measurement (Observation) Equation

$$y_t = a_t + s_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

  - State Equations

$$a_t = a_{t-1} + b_{t-1} + \xi_t \qquad \xi_t \sim N(0, \sigma_\xi^2)$$

$$b_t = b_{t-1} + \varsigma_t \qquad \varsigma_t \sim N(0, \sigma_\varsigma^2)$$

$$s_t = -s_{t-1} - \ldots - s_{t-p+1} + \omega_t \qquad \omega_t \sim N(0, \sigma_\omega^2)$$

Structural Time Series Model

- Special Cases
  - Non-Seasonal (or Local Linear Trend) Model

$$y_t = a_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

$$a_t = a_{t-1} + b_{t-1} + \xi_t \qquad \xi_t \sim N(0, \sigma_\xi^2)$$

$$b_t = b_{t-1} + \varsigma_t \qquad \varsigma_t \sim N(0, \sigma_\varsigma^2)$$

  - Local Level Model

$$y_t = a_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

$$a_t = a_{t-1} + \xi_t \qquad \xi_t \sim N(0, \sigma_\xi^2)$$

Structural Time Series Model

- Linear Gaussian State Space Model
  - Matrix Representation

$$y_t = Z\theta_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

$$\theta_t = T\theta_{t-1} + R\eta_t \quad \eta_t \sim N(0, V)$$

$$\theta_0 \sim N(C_0, P_0)$$

$$\eta_t = \begin{bmatrix} \xi_t \\ \varsigma_t \\ \omega_t \end{bmatrix}, V = \begin{bmatrix} \sigma_\xi^2 & 0 & 0 \\ 0 & \sigma_\varsigma^2 & 0 \\ 0 & 0 & \sigma_\omega^2 \end{bmatrix}$$

Assuming $p = 4$:

$$\theta_t = \begin{bmatrix} a_t \\ b_t \\ s_t \\ s_{t-1} \\ s_{t-2} \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Structural Time Series Model

- StrucTS
  - `function (x, type = c("level", "trend", "BSM"), init = NULL, fixed = NULL, optim.control = NULL)`

Example 1

- GDP and GDP Growth
  - GDP Quarterly Time Series
    - Trend, Seasonality
  - GDP Growth
    - Simple vs Compound Growth Rate
    - Quarterly Growth vs Annual Growth
  - Time Series Decomposition
  - Smoothing and Forecasting

Example 2

- Standard and Poor's 500 Stock Index
  - High Frequency Daily Index
  - Monthly Index Time Series
    - Trend, Seasonality
  - Monthly Log-Return and Volatility
  - Time Series Decomposition
  - Exponential Smoothing and Forecasting