Demand Forecasting

Logistics and Supply Chain Analytics

Jiahua Wu

382 Business School
j.wu@imperial.ac.uk

Imperial College Business School

Imperial means
Intelligent Business

Outline

- How do companies forecast?
- Three objective forecasting approaches
  - Univariate time series methods
    - ARIMA model
    - Holt-Winters exponential smoothing
  - Econometric models
- Summary

Outline

- How do companies forecast?
- Three objective forecasting approaches
  - Univariate time series methods
    - ARIMA model
    - Holt-Winters exponential smoothing
  - Econometric models
- Summary

Why do Companies Forecast?

Demand forecasting supports corporate-wide planning activities.

Levels of Forecast	Purposes
Strategic (years)	Business planning Capacity planning
Tactical (quarterly)	Marketing/Sales planning Workforce planning
Operational (days/weeks)	Short-term capacity planning Inventory planning

Forecasting Methods

- Subjective
  - Judgemental: sales force surveys, jury of executive opinion
  - Experimental: customer surveys, focus group sessions, test marketing
- Objective
  - Time series: use prior history to predict the future
  - Causal-effect: figure out cause-effect relationships, and use forecast of cause to predict effect
  - Life cycle: use the sales curve of similar products or product lines to predict sales of the focal product
- Often times, firms use a combination of the above approaches

Forecasting Methods

- Subjective
  - Judgemental: sales force surveys, jury of executive opinion
  - Experimental: customer surveys, focus group sessions, test marketing
- Objective
  - Time series: use prior history to predict the future
  - Causal-effect: figure out cause-effect relationships, and use forecast of cause to predict effect
  - Life cycle: use the sales curve of similar products or product lines to predict sales of the focal product
- Often times, firms use a combination of the above approaches

Outline

- How do companies forecast?
- Three objective forecasting approaches
  - Univariate time series methods
    - ARIMA model
    - Holt-Winters exponential smoothing
  - Econometric models
- Summary

Terminology

- A time series is a set of data collected at successive (discrete) time points, i.e., X_t
- Time series can be viewed as stochastic processes (SP)
  - SP is a random variable indexed by time
- A time series X_t is (weakly) stationary if
  1. $E(X_t) = \mu$, where μ is constant
  2. $Cov(X_t, X_{t+k}) = \gamma_k$, where γ_k is independent of t
- The sequence $\gamma_k = Cov(X_t, X_{t+k})$ is the auto-covariance function
  - Values of auto-covariances depend on the units of measurement of X_t

Terminology

- It is thus more convenient to use auto-correlation function (ACF), which is defined as the auto-covariances divided by the variance

$$\rho_k = corr(X_t, X_{t+k}) = \gamma_k/\gamma_0$$

- Sample counterpart
  - Sample auto-covariance function is

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})$$

  - Sample auto-correlation function is $\hat{\rho}_k = \hat{\gamma}_k/\hat{\gamma}_0$

An Example: A White Noise Process

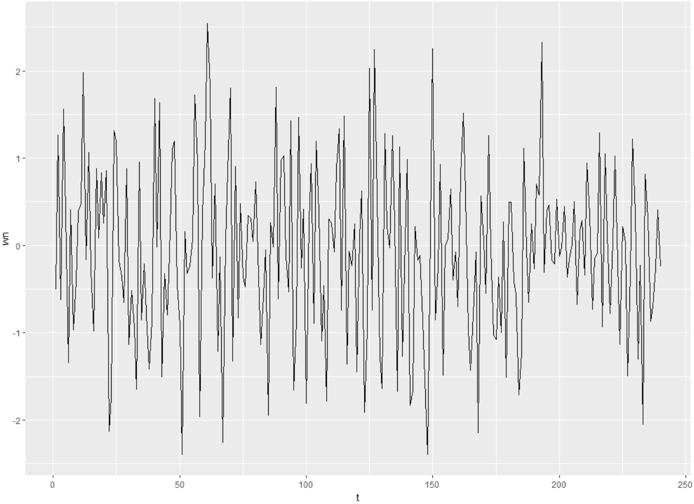
- A white noise process is one with (virtually) no discernible structure. A definition of a white noise process is
  - $E(X_t) = 0$
  - $Var(X_t) = \sigma^2$
  - $\gamma_k = 0$ for any $k > 0$
- Thus the auto-correlation function is zero apart from a single peak of 1 at $k = 0$
- $\hat{\rho}_k, \forall k > 0$ will not be exactly equal to zero, so how do we know whether $\rho_k = 0$?

An Example: A White Noise Process

- We can perform a two-sided test for $H_0 : \rho_k = 0, \forall k > 0$
  - Under the null, $\hat{\rho}_k$ is approximately $Normal(0, 1/n)$, where n is the sample size
  - Typical choice of significance level: 5%
  - We reject H_0 if $|\hat{\rho}_k| > c = \frac{1.96}{\sqrt{n}}$
- Conclusion: If the sample auto-correlation coefficient $\hat{\rho}_k$ falls outside the region $\left[-\frac{1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}}\right]$ for any value of k, then we reject the null hypothesis that the true value of the coefficient at lag k is zero

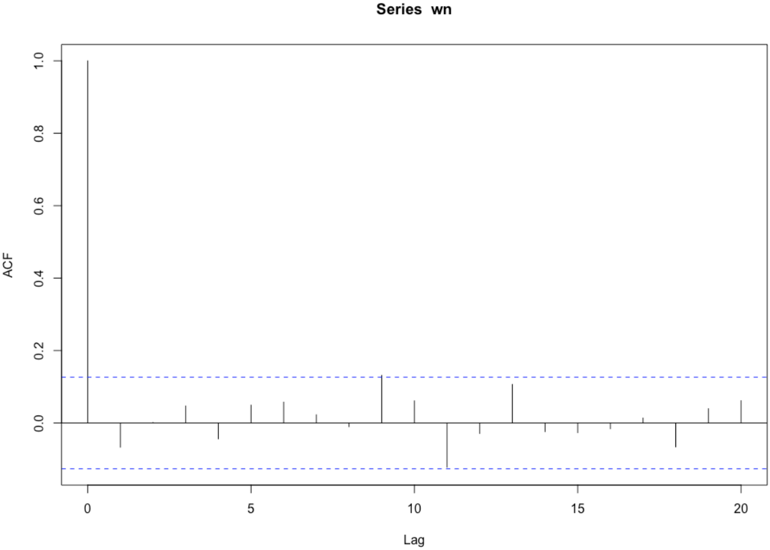
An Example: A White Noise Process

```
> wn <- arima.sim(list(order = c(0,0,0)), 240)
> t <- 1:240
> ggplot(aes(x = t, y = wn)) + geom_line()
```



An Example: A White Noise Process

```
> acf(wn, lag.max=20)
```



Moving Average Models

- In general, ARIMA model has two components: autoregressive (AR) component and moving average (MA) component
- Let $u_t, t = 1, 2, \ldots$ be a sequence of independently and identically distributed (iid) random variables with $E(u_t) = 0$ and $Var(u_t) = \sigma^2$
- The moving average model of order q, or MA(q), is defined to be

$$X_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$$

Moving Average Models

- Define B as the backward shift operator, where $BX_t = X_{t-1}$
- Using the backward shift operator, the moving average model can be re-written as

$$X_t = \mu + \theta(B)u_t,$$

where $\theta(B) \equiv 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$

Properties of Moving Average Models

- $E(X_t) = \mu$
- $Var(X_t) = \gamma_0 = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2$
- Covariances

$$\gamma_k = \begin{cases} (\theta_k + \theta_{k+1}\theta_1 + \cdots + \theta_q\theta_{q-k})\sigma^2, & \text{for } k = 1, 2, \ldots, q \\ 0 & \text{for } k > q \end{cases}$$

ACF of Moving Average Models

- ACF of the moving average model of order q is given by

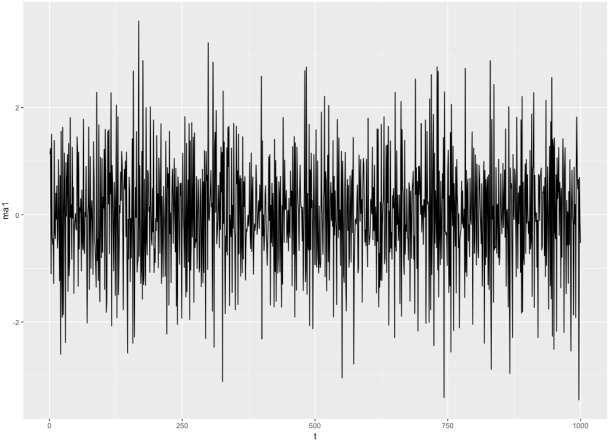
$$\rho_k = \begin{cases} 1 & \text{if } k = 0, \\ \frac{\sum_{i=0}^{q-k} \theta_i \theta_{i+k}}{1+\theta_1^2+\cdots+\theta_q^2} & \text{if } 1 \leq k \leq q, \\ 0 & \text{if } k > q. \end{cases}$$

- For an MA(q) model, its ACF vanishes after lag q

An Example: MA(1)

- Model: $X_t = -0.5u_{t-1} + u_t$

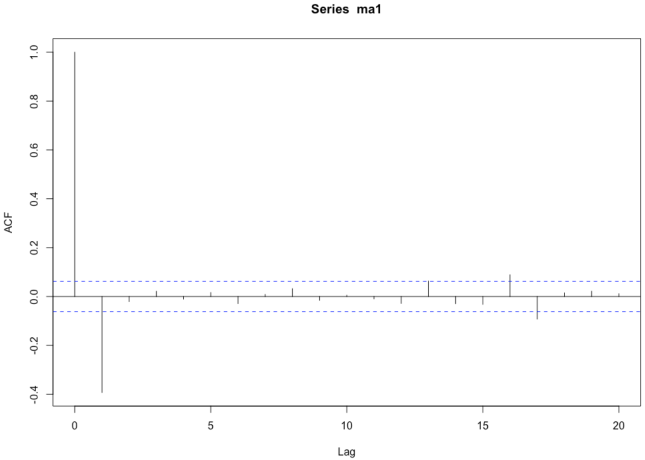
```
> ma1 <- arima.sim(list(ma = c(-0.5)), n = 1000)
> t <- 1:1000
> ggplot(data = ma1, aes(x = t, y = ma1)) + geom_line()
```



An Example: MA(1)

- Model: $X_t = -0.5u_{t-1} + u_t$

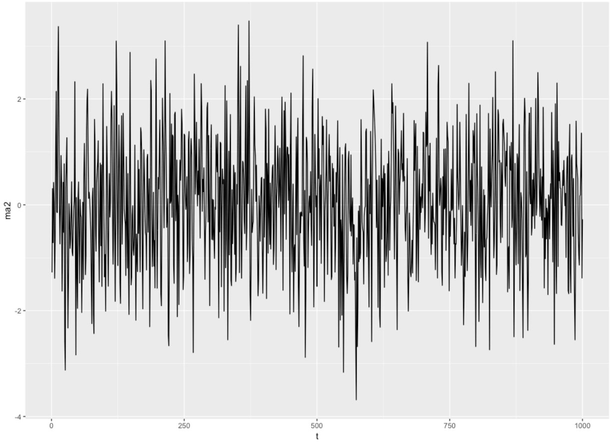
```
> acf(ma1, lag.max=20)
```



An Example: MA(2)

- Model: $X_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$

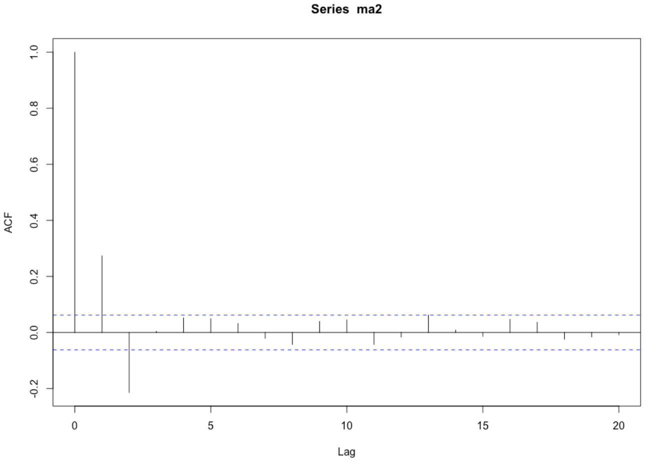
```
> ma2 <- arima.sim(list(ma = c(0.5, -0.25)), n = 1000)
> t <- 1:1000
> ggplot(data = ma2, aes(x = t, y = ma2)) + geom_line()
```



An Example: MA(2)

- Model: $X_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$

```
> acf(ma2, lag.max=20)
```



Autoregressive Models

- The autoregressive model of order p, or AR(p), is of the form

$$X_t = \mu + u_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p}$$

- Using the backward shift operator, the autoregressive model can be re-written as

$$\phi(B) X_t = \mu + u_t,$$

where $\phi(B) \equiv 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$

The Stationary Condition for an AR Model

- The condition for stationarity of a general $AR(p)$ model is that the roots of

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

  all lie outside the unit circle

- Example: Is $X_t = X_{t-1} + u_t$ stationary?
  - The root is 1, so non-stationary
  - It is a unit root process

Autoregressive Models

- ACF of the autoregressive model of order p is given by Yule-Walker equations

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}$$

  - For AR(1), i.e., $X_t = \phi X_{t-1} + u_t$, its ACF is given by $\rho_k = \phi^k$
  - If the AR model is stationary, the auto-correlation function decays exponentially to zero

- ACF alone tells us little about the order of dependence for AR
- We need partial auto-correlation function (PACF), which behaves like ACF for MA models

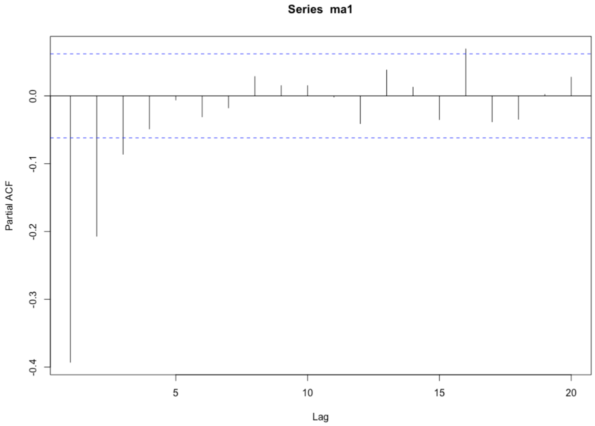
Partial Auto-Correlation Function (PACF)

- Measure the correlation between an observation k periods ago and current observation, after controlling for observations at intermediate lags (i.e., all lags $< k$)
  - That is, PACF with lag k measures the correlation between X_t and X_{t-k} after removing the effects of X_{t-k+1}, X_{t-k+2}, ... X_{t-1}
- For an AR(p) model, its PACF vanishes after lag p
- For an MA(q) model, its PACF decays exponentially

An Example: MA(1)

- Model: $X_t = -0.5u_{t-1} + u_t$

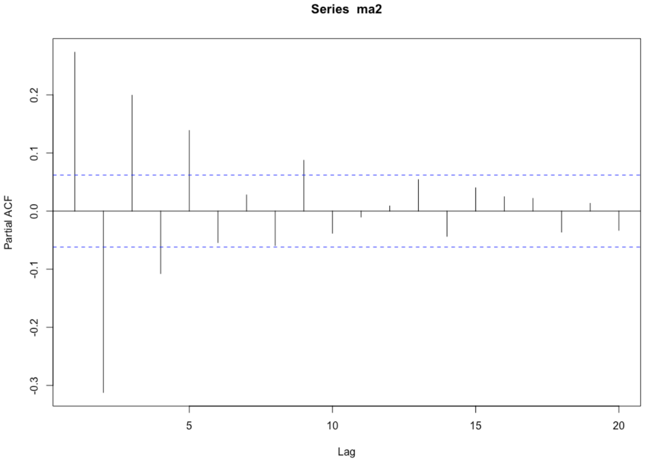
```
> pacf(ma1, lag.max = 20)
```



An Example: MA(2)

- Model: $X_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$

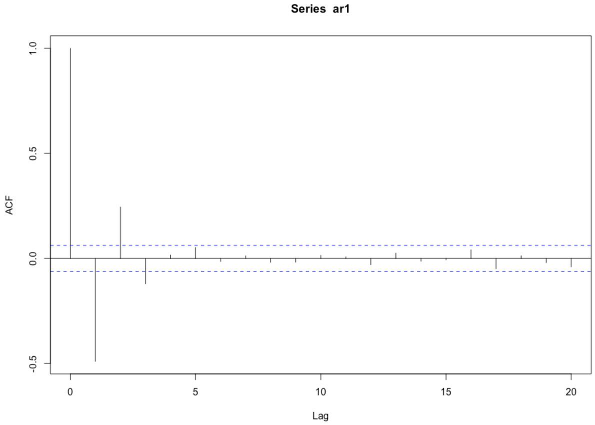
```
> pacf(ma2, lag.max=20)
```



An Example: AR(1)

- Model: $X_t = -0.5X_{t-1} + u_t$

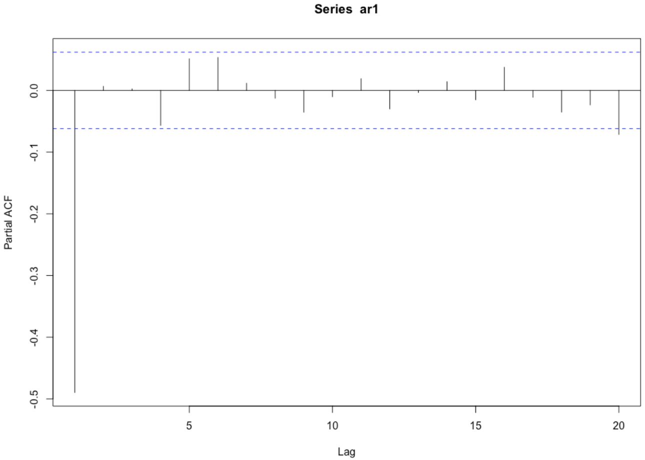
```
> ar1 <- arima.sim(list(ar = c(-0.5)), n = 1000)
> acf(ar1, lag.max = 20)
```



An Example: AR(1)

- Model: $X_t = -0.5X_{t-1} + u_t$

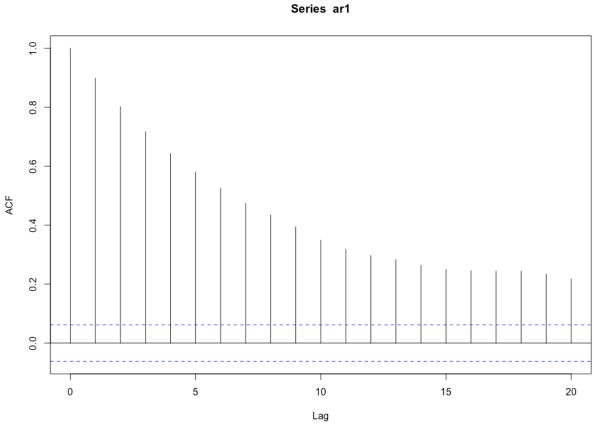
```
> pacf(ar1, lag.max=20)
```



An Example: AR(1)

- Model: $X_t = 0.9X_{t-1} + u_t$

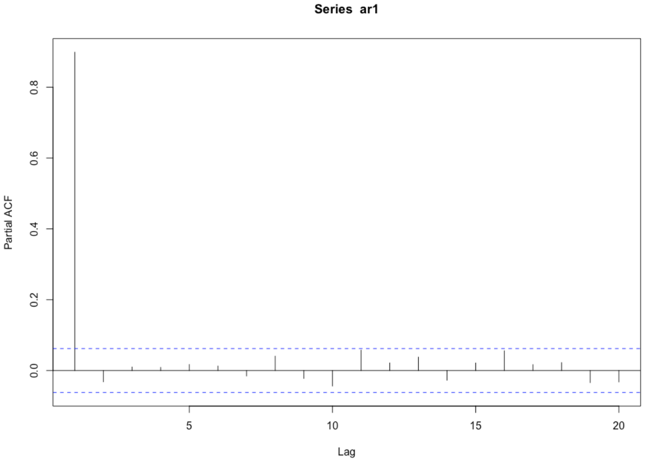
```
> ar1 <- arima.sim(list(ar = c(0.9)), n = 1000)
> acf(ar1, lag.max = 20)
```



An Example: AR(1)

- Model: $X_t = 0.9X_{t-1} + u_t$

```
> pacf(ar1, lag.max=20)
```



Summary of ACF and PACF for AR and MA Processes

- An autoregressive (AR) process has
  - an exponentially decaying ACF
  - number of spikes of PACF = AR order
- A moving average (MA) process has
  - number of spikes of ACF = MA order
  - an exponentially decaying PACF

Autoregressive Moving Average Models (ARMA)

- The autoregressive moving average model of orders p and q, or ARMA(p,q), is of the form

  $$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \mu + u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q},$$

  or

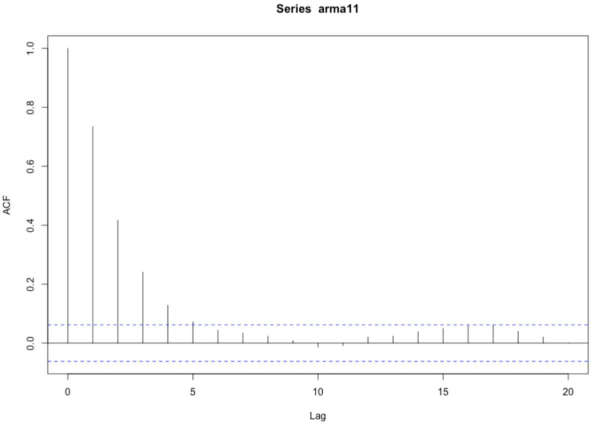
  $$\phi(B)X_t = \mu + \theta(B)u_t$$

- The process reduces to AR(p) if $q = 0$, or to MA(q) if $p = 0$
- ACF/PACF for an ARMA process will display combinations of behavior derived from the AR and MA parts

An Example: ARMA(1,1)

- Model: $X_t = 0.5X_{t-1} + 0.5u_{t-1} + u_t$

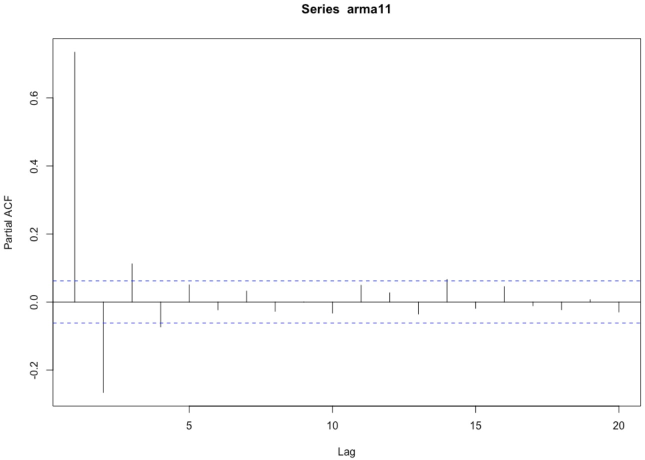
```
> arma11 <- arima.sim(list(ar = c(0.5), ma = c(0.5)), n = 1000)
> acf(arma11, lag.max = 20)
```



An Example: ARMA(1,1)

- Model: $X_t = 0.5X_{t-1} + 0.5u_{t-1} + u_t$

```
> pacf(arma11, lag.max = 20)
```



Time Series Data in Practice

- Most time series are NOT stationary
- Trends in time series
  - Many time series have a common tendency of growing or shrinking over time
- Seasonality in time series
  - Seasonal patterns may be caused by predictable annual events
    - Thanksgiving sales in the US, and Boxing day sales in Canada, UK and Australia
    - Ski sales in winter

Classical Decomposition of Time Series

- One standard method of describing a time series is that of classical decomposition
  - The series is decomposed into three elements
    - Trend (T_t): long term movements in the mean
    - Seasonal effects (S_t): cyclical fluctuations related to calendar or business cycle
    - Microscopic part (M_t): other random or unsystematic fluctuations
  - ARMA model is used to model microscopic part M_t
  - Trend and seasonal components are eliminated by differencing

Elimination of Trends and Seasonal Components by Differencing

- Differencing
  - First order differencing: $\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$
    - $\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t))$
    - Polynomials of ∇ are manipulated in precisely the same way as polynomial functions of real variables
  - Second order differencing

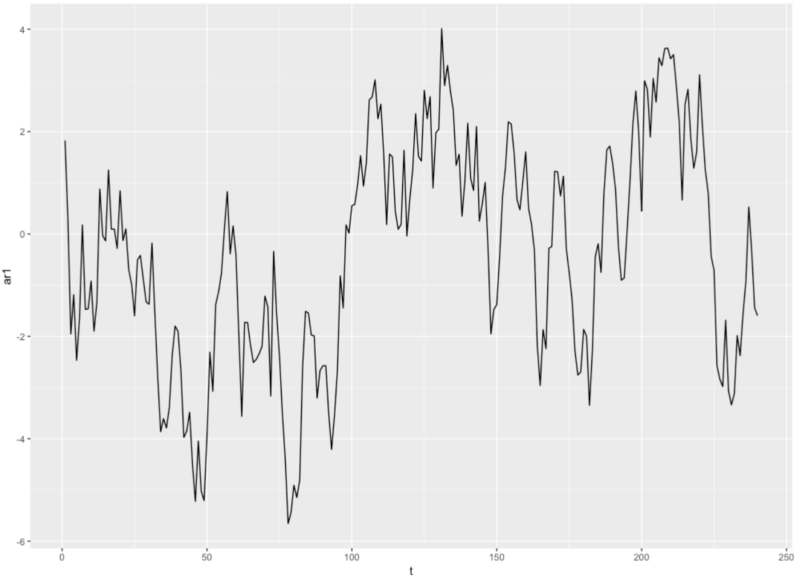
$$\nabla^2 X_t = (1 - B)^2 X_t = (1 - 2B + B^2)X_t = X_t - 2X_{t-1} + X_{t-2}$$

- Rationale behind differencing
  - If $X_t = T_t + M_t$ with $T_t = a + bt$, then T_t is eliminated in the new series $Y_t = X_t - X_{t-1}$
  - If $X_t = S_t + M_t$ with seasonality of period p, we can eliminate the seasonal component with $Y_t = X_t - X_{t-p}$

ARIMA Models

- If the original time series is not stationary, we can look at the first/second order differences
- X_t is said to be ARIMA(p, d, q) process if $\nabla^d X_t$ is an ARMA(p, q) process.
    - Typically, d is a small integer (≤ 2)

Is This Time Series Stationary?



Dickey-Fuller Test

- Using AR(1) model as an example. The test is for the null hypothesis that $\phi = 1$ in

$$X_t = \phi X_{t-1} + u_t$$

against the one-sided alternative $\phi < 1$

- So we have

$$H_0: \text{ series contains a unit root}$$

vs

$$H_1: \text{ series is stationary}$$

Dickey-Fuller Test

- Detailed steps
  1. Take the first order difference and get ∇X_t
  2. Run the regression $\nabla X_t = \psi X_{t-1} + u_t$
     - A test of $\phi = 1$ is equivalent to a test of $\psi = 0$ (since $\phi - 1 = \psi$)
  3. Test statistics is defined as

$$\text{test statistic} = \frac{\hat{\psi}}{se(\hat{\psi})}$$

  4. Reject null hypothesis of a unit root if the test statistic is less than critical values

Augmented Dickey Fuller (ADF) Test

- The test above is only valid if u_t is white noise
- u_t will be auto-correlated if there was auto-correlation in the dependent variable of the regression (∇X_t)
- The solution is to "augment" the test using p lags of the dependent variable. The alternative model is now given by

$$\nabla X_t = \psi X_{t-1} + \sum_{i=1}^{p} \alpha_i \nabla X_{t-i} + u_t$$

- The optimal number of lags can be chosen based on information criteria

Phillips-Perron Test

- Phillips and Perron have developed a more comprehensive theory of unit root stationary
- The tests are similar to ADF tests, but they incorporate a nonparametric correction to the DF procedure to allow for auto-correlated residuals
- The test usually gives the same conclusions as the ADF test, but calculation of the test statistics is more complicated

Criticism of ADF and PP tests

- The tests are poor at deciding when a root is close to the non-stationary boundary
- If the true data generating process is

$$X_t = 0.95X_{t-1} + u_t$$

  then the null hypothesis of a unit root should be rejected
- However, ADF and PP tests usually fail to reject null
- One way to get around this is to use a stationarity test

Stationarity Tests

- In any stationarity test, we have

$$H_0 : X_t \text{ is stationary}$$

vs

$$H_1 : X_t \text{ is not stationary}$$

So that by default under the null the data is stationary

- One such stationarity test is the KPSS test
- We can compare the result of KPSS test with ADF/PP procedure to see if we obtain the same conclusion

Stationarity tests in R

- ADF test: `adf.test()`
- PP test: `pp.test()`
- KPSS test: `kpss.test()`
- A useful function `ndiffs()`: determines the number of first differences required

Fitting ARIMA Models: The Box-Jenkins Procedure

- The Box-Jenkins procedure is concerned with fitting an ARIMA model to data
- It has three parts: identification, estimation and verification
- Identification
  - The data may require pre-processing to make it stationary
  - To achieve stationarity we may do any of the following
    - Rescale it (for instance, by a log or exponential transform)
    - Remove deterministic components
    - Difference it
  - We recognise stationarity by the observation that the autocorrelations decay to zero exponentially fast

Fitting ARIMA Models: The Box-Jenkins Procedure

- Identification
  - Once the series is stationary, we can try to fit an ARMA(p, q) model
  - Selection of p and q based on ACF and PACF

	AR(p)	MA(q)	ARMA(p, q)
ACF	tails off	cuts off after lag q	tails off
PACF	cuts off after lag p	tails off	tails off

  - A rule of thumb is that sample ACF and PACF values are negligible when they lie between $\pm 1.96/\sqrt{n}$
  - More rigorous measures involving information criteria
- Estimation
  - Using the maximum likelihood estimators

Fitting ARIMA Models: The Box-Jenkins Procedure

- Verification: check whether the model fits the data using residuals analysis
  - Calculate the residuals from the model and plot them. The graph should give no indication of a non-zero mean or non-constant variance
  - Plot the sample ACF of the residuals. No more than two or three out of 40 shall fall outside the bounds $\pm 1.96/\sqrt{n}$
  - The same applies to sample PACF of the residuals
  - Tests for randomness of the residuals: Ljung-Box, McLeod-Li, turning points, difference-sign, rank test, Jarque-Bera, and etc.

Forecasting with ARIMA Models: MA Model

- An $MA(q)$ only has memory of q
- Eg. say we have estimated an $MA(3)$ model

$$
\begin{aligned}
X_t &= \mu + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \theta_3 u_{t-3} + u_t \\
X_{t+1} &= \mu + \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2} + u_{t+1} \\
X_{t+2} &= \mu + \theta_1 u_{t+1} + \theta_2 u_t + \theta_3 u_{t-1} + u_{t+2} \\
X_{t+3} &= \mu + \theta_1 u_{t+2} + \theta_2 u_{t+1} + \theta_3 u_t + u_{t+3}
\end{aligned}
$$

- We are at time t and we want to forecast $1, 2, \ldots, s$ steps ahead

Forecasting with ARIMA Models: MA Model

- We know $X_t, X_{t-1}, \ldots$ and $u_t, u_{t-1}, \ldots$

$$
\begin{aligned}
f_{t,1} &= E(X_{t+1}|\Omega_t) \\
&= E(\mu + \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2} + u_{t+1}|\Omega_t) \\
&= \mu + \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2}
\end{aligned}
$$

$$
\begin{aligned}
f_{t,2} &= E(X_{t+2}|\Omega_t) \\
&= E(\mu + \theta_1 u_{t+1} + \theta_2 u_t + \theta_3 u_{t-1} + u_{t+2}|\Omega_t) \\
&= \mu + \theta_2 u_t + \theta_3 u_{t-1}
\end{aligned}
$$

$$
\begin{aligned}
f_{t,3} &= E(X_{t+3}|\Omega_t) \\
&= E(\mu + \theta_1 u_{t+2} + \theta_2 u_{t+1} + \theta_3 u_t + u_{t+3}|\Omega_t) \\
&= \mu + \theta_3 u_t
\end{aligned}
$$

Forecasting with ARIMA Models: MA Model

$$
\begin{aligned}
f_{t,4} &= E(X_{t+4}|\Omega_t) = \mu \\
f_{t,s} &= E(X_{t+s}|\Omega_t) = \mu, \forall s \geq 4
\end{aligned}
$$

Forecasting with ARIMA Models: AR Model

- Say we have estimated an $AR(2)$

$$
\begin{aligned}
X_t &= \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + u_t \\
X_{t+1} &= \mu + \phi_1 X_t + \phi_2 X_{t-1} + u_{t+1} \\
X_{t+2} &= \mu + \phi_1 X_{t+1} + \phi_2 X_t + u_{t+2} \\
X_{t+3} &= \mu + \phi_1 X_{t+2} + \phi_2 X_{t+1} + u_{t+3}
\end{aligned}
$$

- We are at time t and we want to forecast $1, 2, \ldots, s$ steps ahead

$$
\begin{aligned}
f_{t,1} &= E(X_{t+1}|\Omega_t) \\
&= E(\mu + \phi_1 X_t + \phi_2 X_{t-1} + u_{t+1}|\Omega_t) \\
&= \mu + \phi_1 X_t + \phi_2 X_{t-1}
\end{aligned}
$$

Forecasting with ARIMA Models: AR Model

$$\begin{aligned} f_{t,2} &= E(X_{t+2}|\Omega_t) \\ &= E(\mu + \phi_1 X_{t+1} + \phi_2 X_t + u_{t+2}|\Omega_t) \\ &= \mu + \phi_1 f_{t,1} + \phi_2 X_t \end{aligned}$$

$$\begin{aligned} f_{t,3} &= E(X_{t+3}|\Omega_t) \\ &= E(\mu + \phi_1 X_{t+2} + \phi_2 X_{t+1} + u_{t+3}|\Omega_t) \\ &= \mu + \phi_1 f_{t,2} + \phi_2 f_{t,1} \end{aligned}$$

- We can see immediately that

$$f_{t,s} = \mu + \phi_1 f_{t,s-1} + \phi_2 f_{t,s-2}, \forall s \geq 3$$

- Can easily generate *ARIMA*(p, q) forecasts in the same way

Measures of Forecast Error

- Forecast error for Period $t + s$ is given by

$$E_{t+s} = f_{t,s} - X_{t+s}$$

- Common measures of forecast error
  - Mean squared error: $MSE_n = \frac{1}{n} \sum_{s=1}^{n} E_{t+s}^2$
    - penalizes large errors much more significantly than small errors
    - use MSE if the cost of a large error is much larger than the gains from very accurate forecasts
  - Mean absolute deviation: $MAD_n = \frac{1}{n} \sum_{s=1}^{n} |E_{t+s}|$
    - an appropriate choice if the cost of a forecast error is proportional to the size of the error
  - Mean absolute percentage error: $MAPE_n = \frac{\sum_{s=1}^{n} \left| \frac{E_{t+s}}{X_{t+s}} \right|}{n} \cdot 100$
    - a good measure when the underlying forecast has significant seasonality and demand varies considerably from one period to the next

Seasonal ARIMA Models

- A seasonal ARIMA model is written as

$$\text{ARIMA}(p, d, q) \times (P, D, Q)_s,$$

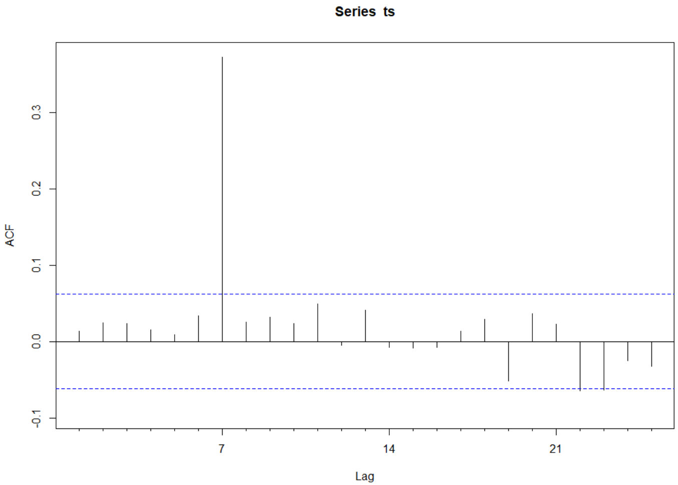
  where

  - (p, d, q): represents the non-seasonal part of the model
  - $(P, D, Q)_s$: represents the seasonal part of the model; s is the periodicity

- The seasonal part of an AR or MA model will be seen in the seasonal lags of the PACF and ACF
  - When $(P, D, Q)_s = (0, 0, 1)_7$
    - A single significant spike at lag 7 in the ACF
    - The PACF shows exponential decay in the seasonal lags

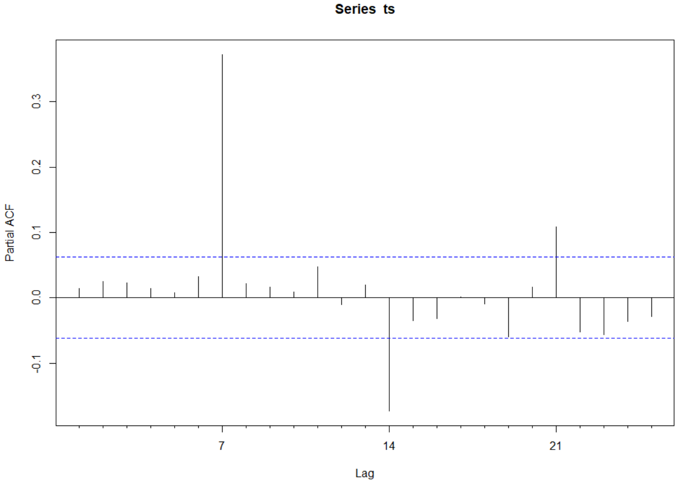
An Example: ARIMA$(0,0,0) \times (0,0,1)_7$

- Model: $X_t = 0.5u_{t-7} + u_t$



An Example: ARIMA$(0,0,0) \times (0,0,1)_7$

- Model: $X_t = 0.5X_{t-7} + u_t$



ARIMA in R

- Model selection with `auto.arima`
  - `auto.arima(x, d=NA, D=NA, max.p=5, max.q=5, max.P=2, max.Q=2, max.order=5, max.d=2, max.D=1, ic=c("aicc", "aic", "bic"), stepwise=TRUE, trace=FALSE, test=c("kpss","adf","pp"),…)`
- Algorithm
  1. Determine d using KPSS tests
  2. Choose p and q by minimizing AICc
     - Initial model candidates: ARIMA$(2, d, 2)$, ARIMA$(0, d, 0)$, ARIMA$(1, d, 0)$ and ARIMA$(0, d, 1)$
     - Variations are considered: add or minus p and/or q by 1
  3. Repeat step 2 until no lower AICc can be found

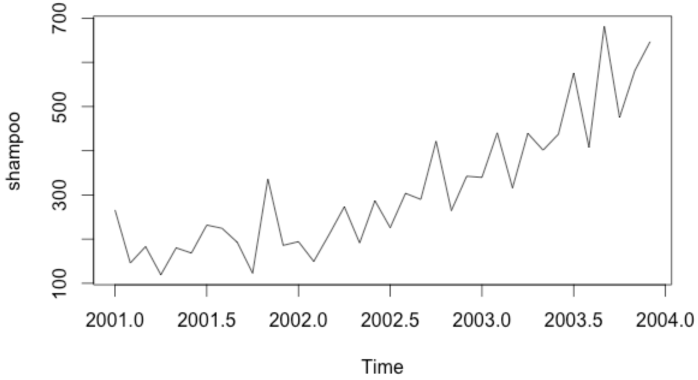
ARIMA in R

- Model estimation with `arima` and `Arima`
  - `arima/Arima(x, order = c(0L, 0L, 0L), seasonal = list(order = c(0L, 0L, 0L), period = NA),…)`
  - Key difference: `Arima` allows for a nonzero constant being included in the model for the first differenced data, i.e.,

$$\phi(B)X_t = \mu + \theta(B)u_t$$

An Example

- A dataset contains monthly sales for shampoo at a retailer store for January 2001-December 2003 (`shampoo.csv`)



Outline

- How do companies forecast?
- Three objective forecasting approaches
  - Univariate time series methods
    - ARIMA model
    - Holt-Winters exponential smoothing
  - Econometric models
- Summary

Classical Decomposition of Time Series

- Recall that under the classical decomposition, the time series can be decomposed into three elements
  - Trend (T_t): long term movements in the mean
  - Seasonal effects (S_t): cyclical fluctuations related to calendar or business cycle
  - Microscopic part (M_t): other random or unsystematic fluctuations
- An alternative approach is to model each of these three elements separately, and then combine them
  - either additively: $X_t = T_t + S_t + M_t$
  - or multiplicatively: $X_t = T_t S_t + M_t$

Holt-Winters Exponential Smoothing

- Notations
  - $X_t =$ actual demand observed in Period t
  - Two parts to characterize the trend in data: intercept (level) and slope (trend)
    - $\hat{L}_t =$ estimate of level for Period t
    - $\hat{T}_t =$ estimate of trend for Period t
  - $\hat{S}_t =$ estimate of seasonal factor for Period t

Simple Exponential Smoothing

- When there is no trend or seasonality, Holt-Winters model reduces to simple exponential smoothing

$$\text{Systematic component of demand} = \text{level}$$

- The initial estimate of level $\hat{L}_0$ is taken to be the average of all historical data

$$\hat{L}_0 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- After observing the demand X_{t+1} for Period $t+1$, we revise the estimate of the level as

$$\hat{L}_{t+1} = \alpha X_{t+1} + (1-\alpha)\hat{L}_t,$$

where α is a smoothing constant for the level, $0 < \alpha < 1$

- The current forecast for all future periods is

$$f_{t,h} = \hat{L}_t, \text{ for any } h > 0$$

Simple Exponential Smoothing

- We can also express the level in a given period as

$$\begin{aligned}\hat{L}_{t+1} &= \alpha X_{t+1} + (1-\alpha)\hat{L}_t \\ &= \sum_{n=0}^{t} \alpha(1-\alpha)^n X_{t+1-n} + (1-\alpha)^t \hat{L}_0\end{aligned}$$

  - The current estimate of the level is a weighted average of all of the past observations of demand
  - It assigns a set of exponentially declining weights to past data (i.e., recent observations weighted higher than older observations)
    - A higher value of α corresponds to a forecast that is more responsive to recent observations
    - A lower value of α represents a more stable forecast that is less responsive to recent observations

Trend-Corrected Exponential Smoothing (Holt's Model)

- The trend-corrected exponential smoothing is appropriate when demand has a trend but no seasonality

$$\text{Systematic component of demand} = \text{level} + \text{trend}$$

- The initial estimate of level and trend is obtained by running a linear regression between demand X_t and time period t, i.e., $X_t = at + b$
  - b: measures the estimate of demand at Period $t = 0$, and is our estimate of $\hat{L}_0$
  - a: measures the rate of change in demand per period, and is our estimate of $\hat{T}_0$

Trend-Corrected Exponential Smoothing (Holt's Model)

- In Period t, given estimates of $\hat{L}_t$ and $\hat{T}_t$, the forecast for future periods is expressed as

$$f_{t,h} = \hat{L}_t + h\hat{T}_t, \text{ for any } h > 0$$

- After observing the demand X_{t+1} for Period $t+1$, we revise the estimate of the level and trend as follows

$$\begin{aligned} \hat{L}_{t+1} &= \alpha X_{t+1} + (1-\alpha)(\hat{L}_t + \hat{T}_t) \\ \hat{T}_{t+1} &= \beta(\hat{L}_{t+1} - \hat{L}_t) + (1-\beta)\hat{T}_t \end{aligned}$$

where α is a smoothing constant for the level, $0 < \alpha < 1$, and β is a smoothing constant for the trend, $0 < \beta < 1$

Trend- And Seasonality-Corrected Exponential Smoothing (Winter's Model)

- We discuss one case to illustrate the concepts of Winter's Model
  - The systematic component has the multiplicative form - can easily be modified for the other case
  - Demand is seasonal with periodicity p
- This method is appropriate when the systematic component of demand has a level, a trend and a seasonal factor, i.e.,

  Systematic component of demand $=$ (level$+$trend)$\cdot$seasonal factor

Trend- And Seasonality-Corrected Exponential Smoothing (Winter's Model)

- Initial estimates are obtained as follows
  1. The deseasonalized demand $\bar{X}_t$ for Period t is given by

$$\bar{X}_t = \begin{cases} \left[X_{t-(p/2)} + X_{t+(p/2)} + \sum_{i=t+1-(p/2)}^{t-1+(p/2)} 2X_i \right] /2p & \text{if } p \text{ is even} \\ \sum_{i=t-[(p-1)/2]}^{t+[(p-1)/2]} X_i/p & \text{if } p \text{ is odd} \end{cases}$$

  2. Regress $\bar{X}_t$ on t, and obtain $\hat{L}_0$ and $\hat{T}_0$ as in the Holt's model
  3. Seasonal factor for Period t is given by $\bar{S}_t = X_t/\bar{X}_t$
  4. Given r seasonal cycles in the data, for all periods of the form $pt + i$, $1 \leq i \leq p$, the seasonal factor is obtained as $\hat{S}_i = \frac{\sum_{j=0}^{r-1} \bar{S}_{jp+i}}{r}$

Trend- And Seasonality-Corrected Exponential Smoothing (Winter's Model)

- In Period t, given estimates of level $\hat{L}_t$, trend $\hat{T}_t$ and seasonal factors, $\hat{S}_t, \dots, \hat{S}_{t+p-1}$, the forecast for future periods is given by
$$f_{t,h} = (\hat{L}_t + h\hat{T}_t)\hat{S}_{t+h}, \text{ for any } h > 0$$
- After observing the demand X_{t+1} for Period $t+1$, we revise the estimate of the level, trend, and seasonal factors as follows
$$\begin{aligned} \hat{L}_{t+1} &= \alpha(X_{t+1}/\hat{S}_{t+1}) + (1-\alpha)(\hat{L}_t + \hat{T}_t) \\ \hat{T}_{t+1} &= \beta(\hat{L}_{t+1} - \hat{L}_t) + (1-\beta)\hat{T}_t \\ \hat{S}_{t+p+1} &= \gamma(X_{t+1}/\hat{L}_{t+1}) + (1-\gamma)\hat{S}_{t+1} \end{aligned}$$
where α is a smoothing constant for the level, $0 < \alpha < 1$; β is a smoothing constant for the trend, $0 < \beta < 1$; and γ is a smoothing constant for the seasonal factor, $0 < \gamma < 1$.

A Framework for Holt-Winters Exponential Smoothing

- Initialize: Compute initial estimate of the level ($\hat{L}_0$), trend ($\hat{T}_0$), and seasonal factors ($\hat{S}_1, \ldots, \hat{S}_p$) from the given data
- For each Period t
  - Forecast: Given the estimates in Period t, forecast demand for future periods using $f_{t,h} = (\hat{L}_t + h \cdot \hat{T}_t)\hat{S}_{t+h}$
    - The first forecast is for Period 1 and is made with the estimates of level, trend, and seasonal factor at Period 0
  - Estimate error: Record the actual demand X_{t+1} for Period $t+1$ and compute the error E_{t+1} in the forecast for Period $t+1$ as

$$E_{t+1} = f_{t,1} - X_{t+1}$$

  - Update estimates of level, trend and seasonality with X_{t+1}
- Optimize: Find the smoothing constants such that the selected measure of forecast errors is minimized

Holt-Winters in R

```
HoltWinters(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())
```

- `x`: An object of class `ts`
- `alpha`, `beta`, `gamma`, `seasonal`: Holt-Winters model specification
- `start.periods`: start periods used in the autodetection of start values
- `l.start`, `b.start`, `s.start`: start values for level, trend and seasonal factors
- `optim.start`: the starting values for the optimizer

Holt-Winters in R

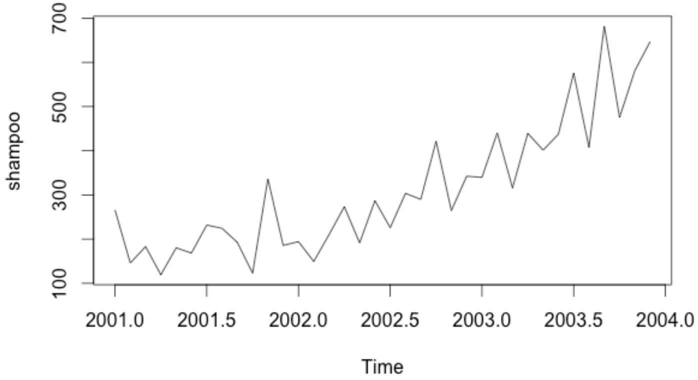
- Another function for Holt-Winters algorithm: `ets(y, model="ZZZ",…)`
  - `model="ZZZ"`: the first letter denotes error type, the second letter denotes the trend type, and the third letter denotes the season type
  - Parameters
    - "N" = none
    - "A" = additive
    - "M" = multiplicative
    - "Z" = automatically selected
    - Eg., "AAA" indicates additive Holt-Winters' model with trend and seasonality

Holt-Winters in R

- Difference between `HoltWinters()` and `ets()`
  - `HoltWinters()`
    - uses heuristic values for the initial states
    - estimates the smoothing parameters by minimizing MSE
  - `ets()`
    - estimates both the initial states and smoothing parameters by maximizing the likelihood function
    - provides a larger model class
- The author claims that `ets()` is more reliable; however, not widely tested

An Example

- A dataset contains monthly sales for shampoo at a retailer store for January 2001-December 2003 (`shampoo.csv`)



Outline

- How do companies forecast?
- Three objective forecasting approaches
  - Univariate time series methods
    - ARIMA model
    - Holt-Winters exponential smoothing
  - Econometric models
- Summary

Examples of Time Series Regression

- Static models
  - A static model describes the relationship among contemporaneous variables
  - Eg. A simple sales model: $sales_t = \beta_0 + \beta_1 price_t + u_t$ may be used to study the relationship between sales and price
- Finite distributed lag (FDL) models
  - A FDL model allows the lags of one or more variables to affect the dependent variable

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \ldots + \delta_q z_{t-q} + u_t,$$

which is called an FDL of order q.

Finite Distributed Lag Models

- FDL(q) model

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \ldots + \delta_q z_{t-q} + u_t.$$

- The partial effect of z_{t-j} on y_t is δ_j (holding everything else fixed)
- δ_0 is called the impact propensity/multiplier
- When there is a permanent one-unit shift in z at t, i.e.,

$$\Delta z_s = 0 \text{ for } s < t \quad \text{and} \quad \Delta z_s = 1 \text{ for } s \geq t,$$

  the eventual effect on y is $(\delta_0 + \delta_1 + \ldots + \delta_q)$, known as the long-run propensity/multiplier

Using Time Trend in Regression Analysis

- In econometrics models, time trend is generally explicitly included
  - Linear trend

    $$y_t = \alpha_0 + \alpha_1 t + u_t.$$

    Holding other factors fixed,

    $$\Delta y_t = y_t - y_{t-1} = \alpha_1.$$

  - Exponential trend

    $$\log(y_t) = \beta_0 + \beta_1 t + u_t.$$

    Holding other factors fixed,

    $$\Delta y_t / y_t \approx \Delta \log(y_t) = \beta_1.$$

Using Time Trend in Regression Analysis

- Consider the model

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 t + u_t,$$

  where the time trend captures unobserved factors that grow (or shrink) over time

  - β_1 is the effect of x_{t1} on y_t, holding other factors (including trend) fixed
  - In a time series regression with trending variables, R^2 is often very high, which is the consequence of trending, not necessarily the explanatory power of (x_{t1}, x_{t2})

Using Seasonal Dummies in Regression Analysis

- A model of monthly series with both trend and seasonality

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 t + \beta_4 t^2 \\ &\quad + \delta_1 feb_t + \delta_2 mar_t + \ldots + \delta_{11} dec_t + u_t, \end{aligned}$$

where β_1 is the effect of x_{t1} on y_t, holding other factors (including trend and seasonality) fixed

Forecasting with Econometric Models

- One-step-ahead forecast
  - Models with only lagged independent variables

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + u_t$$

    - Exactly the same as prediction in univariate time series models
    - $f_{t,1} = E(y_{t+1}|\Omega_t) = \beta_0 + \beta_1 y_t + \beta_2 x_t$
  - Models with contemporaneous variables

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

    - Need to know x_t in the future time period
- Multiple-step-ahead forecast
  - Need to forecast both dependent and independent variables
  - Vector autoregressive (VAR) model

When to use Econometric Models?

- Appropriate when demand is highly correlated with some environmental factors
  - Model is built to relate the independent exogenous factors to the demand
- Examples
  - Diapers/Nappies $\sim$ birth rates lagged by 1 year
  - Promotional items $\sim$ marketing promotions & ads
  - Umbrellas/Fuel $\sim$ weather, temperature, etc.

Outline

- How do companies forecast?
- Three objective forecasting approaches
  - Univariate time series methods
    - ARIMA model
    - Holt-Winters exponential smoothing
  - Econometric models
- Summary

Forecasting Methods

Products need to be segmented to help identify the types of forecasting methods needed

Product Segment	**Common Methods**
New products	Life cycle (diffusion models)
Mature products	Time series
Promoted and event-based products	Cause-effect
Slow-moving or sporadic	Poisson Croston's: `crost()` in R

Miscellaneous Forecasting Issues

- Characteristics of forecasts
  - Long-term forecasts are less accurate than short-term forecasts
  - Aggregate forecasts are more accurate than disaggregate forecasts
    - Top down vs. bottom up forecasting?
- Data issues
  - Sales data is not demand data
  - Historical data might not exist
  - Demand visibility can be skewed by level of echelon
    - Collaborative planning, forecasting, and replenishment (CPFR)