FORECASTING

20/04/15

Quantitative forecasting can only be applied if:
- Information about the past is available
- Information is quantified as numerical data.
- The continuity assumption holds: means that some aspects of the past will continue into the future

Time Series: we make no attempt to discover what variables might affect the quantity to be forecast. We look at the values of the quantity over time (time series), try to discover patterns and make a forecast by extrapolating them into the future.
We could model our inflation rate as:

inflation rate$_{t+1}$ = g(inflation rate, inflation rate$_{t-1}$, inflation rate$_{t-2}$ ... error)

The sequence of values $\{y_t\}_{t=0,1,\dots n}$ over time is called a time series.
Time plot – data over time
Season plot – data for each year.

Time Series Pattern
Decomposed into several components:
1. TREND: a long term increase or decrease occurs
2. SEASONAL: series influenced by seasonal factors. Thus, the series exhibits a behaviour that repeats over a fixed period of time.
3. CYCLICAL: series rises or falls regularly but these are not of fixed period
4. ERROR: corresponds to random fluctuations that cannot be explained by a deterministic pattern

Auto Correlation function
For a time series $y_1, y_2 \dots y_n$, the auto correlation at lag k is

$$r_k = \frac{\sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$$

Where $\bar{y} = \frac{1}{n} \sum_{t=1}^{n}$
- Plotted with ... value r_k on y-axis and lag k on x-axis

- Similarity between observations as a function of time lag between them

Partial Autocorrelation Function [PACF]

- Extension of auto correlation where the dependence on the intermediate elements (those within the lag) is removed
- ~~PAC~~ at lag k is the auto correlation between X_t and X_{t-k} that is not accounted for by lag 1 through k-1.
- 1st order partial autocorrelation will be defined equal to 1st order autocorrelation

2nd is:

$$\frac{\text{Covariance}(x_t, x_{t-2} | x_{t-1})}{\sqrt{\text{var}[x_t | x_{t-1}] \text{var}[x_{t-2} | x_{t-1}]}}$$

3rd is:

$$\frac{\text{Cov}[x_t, x_{t-3} | x_{t-1}, x_{t-2}]}{\sqrt{\text{var}[x_t | x_{t-1}, x_{t-2}] \text{var}[x_{t-3} | x_{t-1}, x_{t-2}]}}$$

and so on

FORECASTING 3

HOLT WINTERS

Single Exponential Smoothing (SES)

- $y_1 \dots y_n$ observed values
- y_n last value
- Forecasts for value of time series at future times $n+1, n+2$ using a model fitted to $y_1 \dots y_n$ are denoted F_{n+1}, F_{n+2}
- Fitted values using model are $F_1, F_2 \dots F_n$
- Residuals or errors are called $y_1 - F_1 \dots y_n - F_n$

SES

- Algorithm for creating forecasts iteratively on the basis of how well one did with previous forecasts
- suppose we make a forecast F_t for value of y_t (not yet observed)
- now we observe y_t and wish to make a forecast F_{t+1}. Do this by taking our old forecast F_t and adjusting it using the error in forecasting y_t:
  $F_{t+1} = F_t + \alpha(y_t - F_t)$ where $0 \leq \alpha \leq 1$
- Nearer α is to 1, then the larger the adjustment.
- Cannot forecast the first term in the series (since $F_1 = F_0 + \alpha(y_0 - F_0)$ and there is no F_0 or y_0). By convention we fix $F_1 = y_1$ and forecast from y_2 on

- Initialise: $F_1 = y_1$ choose $0 < \alpha < 1$
- Forecast: $F_{t+1} = F_t + \alpha(y_t - F_t)$
  Until no more observations are available then
  $F_{n+k} = F_{n+1} \quad \forall k \geq 1$

Exponential smoothing forecasts are a weighted sum of all the previous observations

$$F_{t+1} = F_t + \alpha(y_t - F_t)$$
$$= [F_{t-1} + \alpha(y_{t-1} - F_{t-1})] + \alpha(y_t - [F_{t-1} + \alpha(y_{t-1} - F_{t-1})])$$
$$= \alpha y_t + \alpha(1-\alpha)y_{t-1} + (1-\alpha)^2 F_{t-1}$$

- Suitable for no trend and no seasonality

Double Exponential Smoothing (DES)

- AKA Holt's linear model
- Suitable for ts with a trend but no seasonality
- Trying to fit a line continuously to last point to predict next point
- level L_t and slope b_t

- Initialise: $L_1 = y_1$ $b_1 = y_2 - y_1$ $F_1 = y_1$ $0 \le \alpha \le 1$ $0 \le \beta \le 1$
- Compute and forecast:

$$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + b_{t-1})$$
$$b_t = \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}$$
$$F_{t+1} = L_t + b_t$$

- Until no more observations available:

$$F_{n+k} = L_n + k b_n \quad \forall\, k \ge 1$$

- No forecasts or fitted values computed until y_1 and y_2 observed, let $F_1 = y_1$

Picking Values of α and β

- SSE (Sum of Squared Errors): $= \sum_{t=1}^{n} (y_t - F_t)^2$
- computers compute parameters (α, β) such that SSE is minimised

- RMSE (Root Mean Square Error): $= \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t - F_t)^2} = \sqrt{\frac{SSE}{n}}$

- MAPE (Mean Absolute Percent Error): $= 100 \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - F_t}{y_t} \right|$

20/04/15

FORECASTING

HOLT WINTERS

Holt Winters' Exponential Smoothing With Seasonality

- Data with trend and seasonality
- Two techniques additive and multiplicative
- select parameters with lower SSE/RMSE/MAPE

SHW +

Initialise $L_s = \frac{1}{s} \sum_{i=1}^{s} y_i$

$b_s = \frac{1}{s}\left[\frac{y_{s+1}-y_1}{s} + \frac{y_{s+2}-y_2}{s} + \dots + \frac{y_{2s}-y_s}{s}\right]$

$S_i = y_i - L_s \quad i = 1 \dots s$

$0 \le \alpha \le 1 \quad 0 \le \beta \le 1 \quad 0 \le \gamma \le 1$

Compute for $t > s$:

level $L_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1})$

trend $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$

Seasonal $S_t = \gamma(y_t - L_t) + (1-\gamma)S_{t-s}$

forecast $F_{t+1} = L_t + b_t + S_{t+1-s}$

Subsequent forecasts: $F_{n+k} = L_n + k b_n + S_{n+k-s}$

SHW x Multiplicative

- Initialise $L_s = \frac{1}{s} \sum_{i=1}^{s} y_i$

$b_s = \frac{1}{s}\left[\frac{y_{s+1}-y_1}{s} + \dots + \frac{y_{2s}-y_s}{s}\right]$

$S_i = y_i / L_s, \; i = 1 \dots s$

$0 \le \alpha \le 1 \quad 0 \le \beta \le 1 \quad 0 \le \gamma \le 1$

- Compute for $t > s$: $L_t = \alpha \, y_t / S_{t-s} + (1-\alpha)(L_{t-1} + b_{t-1})$

trend $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$

Seasonal $S_t = \gamma \, y_t / L_t + (1-\gamma)S_{t-s}$

forecast $F_{t+1} = (L_t + b_t) S_{t+1-s}$

- Until no more observations available Subsequent forecast:

$F_{n+k} = (L_n + k \cdot b_n) S_{n+k-s}$

ARIMA MODELS

Linear Regression

- Collected data $(x_1,y_1)(x_2,y_2)$... x_i - predictor y_i - response
- Assume y is related to x by: $y_i = a + bx_i + \varepsilon_i$

Assumption that: ε_i is the error and $\varepsilon_i \sim N(0,\sigma^2)$

ε_i and ε_j are independent if $i \neq j$.

- Given a and b, y_i will be normally distributed with mean $a+bx_i$ and variance σ^2

Fitting the Model

- Best fitting values are the least squared estimates that minimise Residual Sum of Squares

$$RSS = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

- Also known as the Sum of Squared Errors (SSE)
- Estimates $\hat{a}, \hat{b}$ computed such that:

$$\frac{d(RSS)}{da} = 0 \qquad \frac{d(RSS)}{db} = 0$$

- Giving the least squared estimates: $\hat{b} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ $\hat{a} = \bar{y} - \hat{b}\bar{x}$

- $\hat{\sigma}^2$ estimated by $s^2 = \frac{\hat{RSS}}{n-2}$

$\hat{RSS}$ is RSS value computed with estimates $\hat{a}, \hat{b}$

Measuring Strength of the Linear Relationship

- Correlation Coefficient: $r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$

- Call $\tilde{x}$ all values of $(x_1-\bar{x}, \ldots, x_n-\bar{x})$ same for $\tilde{y}$

$r_{xy} = \frac{\langle\tilde{x},\tilde{y}\rangle}{\|\tilde{x}\|\,\|\tilde{y}\|}$

- where $\|\tilde{x}\|\,\|\tilde{y}\|$ is the norm of the vector of $\tilde{x}$ wrt $\tilde{y}$

$\langle\tilde{x},\tilde{y}\rangle$ is dot product of $\tilde{x}$ and $\tilde{y}$

We have: $\langle\tilde{x},\tilde{y}\rangle = \|\tilde{x}\| \cdot \|\tilde{y}\| \cos(\theta_{xy})$

20/04/15

REGRESSION

ACUTE NOTES

Coefficient of Determination $R^2 = r^2_{y\hat{y}} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$

For simple linear regression $R^2 = r^2_{xy}$

Evaluating Model Fit

- Look at residuals $\varepsilon_i = y_i - \hat{y}_i$
- Two possible problems
  - We have not fit the pattern of data. Any unmodelled relationships between x and y appear in the residuals. Scatter plot of of the residuals against the x_i should show up any unmodelled patterns
  - Normality assumption not correct. Residuals should be independent and normally distributed mean 0 and variance σ^2. Histogram of residuals can usually verify this

Outliers

- Observation with unusually large residual - observation that has been predicted badly by model
- Standardised residuals give a good indication as to whether an observation is an outlier
- Can cause problems with model fit.

Making Predictions

- Given a new value x, y should be normally distributed with mean $\alpha + \beta x$ and variance σ^2
- Replace α, β and σ^2 by their estimates and to forecast the value of y to be $\hat{y} = \hat{\alpha} + \hat{\beta} x$
- 95% prediction interval $\hat{\alpha} + \hat{\beta} x \pm 2s$

Statistical Tests In Regression

- F-test is used to determine if there is any significant linear relationship between x and y. Checks hyp that $\beta = 0$
- Test statistic is $F = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2)/(n-2)}$

which is compared to F-value having 1 and n-2 degrees of freedom

USING LINEAR REGRESSION TO MAKE FORECASTS

Time as an Explanatory Variable

- TS $(y_1 \dots y_n)$ use x_i values as the time i^{th} observation was taken

Indicator Variables: Modelling Seasonality

- Indicator variable - binary variable that takes value 0 or 1.
- For each month we can define an indicator variable for Jan Feb etc

$$Jan_i = \begin{cases} 1 & \text{if } i \text{ is corresponding to month January} \\ 0 & \text{otherwise} \end{cases}$$

- Fit by linear regression: $y_i = a + bi + \gamma_1 Jan_i + \gamma_2 Feb_i + \dots + \gamma_{12} Dec_i + \varepsilon_i$
- Only need 11 of 12 monthly effects as Error term will account for last month
  $\Rightarrow$ Get rid of Jan

Least Squares For Linear Regression

- Case when we have only one explanatory variable: $y = a + bx + \varepsilon$
- After n observations $\{(x_i, y_i)\}_{i=1 \dots n}$ we can write:

$$\begin{pmatrix} y_1 = a + bx_1 + \varepsilon_1 \\ \vdots \\ y_n = a + bx_n + \varepsilon_n \end{pmatrix}$$

- Can be rewrote as:

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{y} = \underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}}_{X} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\theta} \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{\varepsilon}$$

- Least Squares estimate of θ such that the derivative of the RSS is zero is:

$$\hat{\theta} = (X^T X)^{-1} X^T y \qquad (8.3)$$

Multiple Linear Regression

- Eq^n 8.3 remains the same except X and θ need to be expanded

$$X = \begin{bmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ \vdots & & \vdots \\ 1 & x_n & z_n \end{bmatrix} \qquad \theta = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

20/04/15

ARIMA MODELS

AR(p): Autoregressive Models

Definition: An auto regressive model of order p (denoted AR(p)) states that y_i is a linear function of the previous p values of the time series plus an error term:

$$y_i = \phi_0 + \phi_1 y_{i-1} + \phi_2 y_{i-2} + \dots + \phi_p y_{i-p} \varepsilon_i$$

where $\phi_1 \dots \phi_p$ are weights that we define/determine and ε_i are $\varepsilon \sim (0, \sigma^2)$

- Formula only holds for $i > p$
- Simplest Model $AR(1)$: $y_i = \phi_0 + \phi_1 y_{i-1} + \varepsilon_i$
- For fitting AR(1) model we have observations $y_1 \dots y_n$ that defines the linear system of n-1 equations:

$$\begin{bmatrix} y_2 = \phi_0 + \phi_1 y_1 + \varepsilon_2 \\ y_3 = \phi_0 + \phi_1 y_2 + \varepsilon_3 \\ \vdots \\ y_n = \phi_0 + \phi_1 y_{n-1} + \varepsilon_n \end{bmatrix}$$

1. Define $x_i = y_{i-1}$, called the lagged series. Not defined for $i=1$
2. AR(1) model is: $y_i = \phi_0 + \phi_1 x_i + \varepsilon_i$ - linear regression model can fit this model by doing linear regression of series against lagged series. Estimate parameters and carry out F-test
3. Regression fitted on n-1 points x_1 does not exist
4. Model is: $\hat{y}_i = \hat{\phi}_0 + \hat{\phi}_1 y_{i-1}$ for $i = 2 \dots n$
5. Estimate σ^2 by S^2:
$$S^2 = \frac{1}{n-1-2} \sum_{i=2}^{n} (y_i - \hat{\phi}_0 - \hat{\phi}_1 y_{i-1})^2$$

- have n-1 equations lose 2 dF from 2 parameters estimated
- 95% prediction interval for y_i when x_i is known: $\hat{\phi}_0 + \hat{\phi}_1 x_i \pm 2s$

12

Prediction Interval for AR(1) k Step ahead

1- Forecast for y_{n+1} and prediction interval?

⇒ AR model: $y_{n+1} = \hat{\phi}_0 + \hat{\phi}_1 y_n + \varepsilon_{n+1}$

⇒ We don't know value of $\varepsilon_{n+1} \sim N(0, \sigma^2)$ but we know $\hat{\phi}_0$, $\hat{\phi}_1$ and y_n so:

$$y_{n+1} = \underbrace{\hat{\phi}_0 + \hat{\phi}_1 y_n}_{\text{Forecast of } y_{n+1}} \pm 2s$$

$$s^2 = \frac{\sum_{i=2}^{n} \varepsilon_i^2}{n-3}$$

2 Forecast for y_{n+2}?

- AR(1) model: $y_{n+2} = \hat{\phi}_0 + \hat{\phi}_1 y_{n+1} + \varepsilon_{n+2}$

- Don't know y_{n+1} (we just know a prediction $\hat{y}_{n+1}$), so replacing y_{n+1} by its expression with y_n:

$$\begin{aligned} y_{n+2} &= \hat{\phi}_0 + \hat{\phi}_1 y_{n+1} + \varepsilon_{n+2} \\ &= \hat{\phi}_0 + \hat{\phi}_1(\hat{\phi}_0 + \hat{\phi}_1 y_n + \varepsilon_{n+1}) + \varepsilon_{n+2} \\ &= \underbrace{\hat{\phi}_0 + \hat{\phi}_1\hat{\phi}_0 + \hat{\phi}_1^2 y_n}_{\text{Forecast } \hat{y}_{n+2}} + \underbrace{\hat{\phi}_1 \varepsilon_{n+1} + \varepsilon_{n+2}}_{\text{error term}} \end{aligned}$$

- Note, forecast is only part we can compute (we know $\hat{\phi}_0$, $\hat{\phi}_1$, y_n) whereas we don't known value of the errors, only know how they behave statistically.

3 Prediction Interval for y_{n+2}?

- We know the forecast $\hat{y}_{n+2}$ and the error on this forecast.

- Need to estimate variance of error. First compute mean:

$$\mathbb{E}[\hat{\phi}_1 \varepsilon_{n+1} + \varepsilon_{n+2}] = \hat{\phi}_1 \underbrace{\mathbb{E}[\varepsilon_{n+1}]}_{=0} + \underbrace{\mathbb{E}[\varepsilon_{n+2}]}_{=0} \Rightarrow \underset{=0}{0}$$

- Variance: $\mathbb{E}[(\hat{\phi}_1 \varepsilon_{n+1} + \varepsilon_{n+2})^2] = \hat{\phi}_1^2 \underbrace{\mathbb{E}[\varepsilon_{n+1}^2]}_{=s^2} + 2\hat{\phi}_1 \underbrace{\mathbb{E}[\varepsilon_{n+1}\varepsilon_{n+2}]}_{=0} + \underbrace{\mathbb{E}[\varepsilon_{n+2}^2]}_{=s^2}$

$= (1 + \hat{\phi}_1^2) s^2$ $\quad \varepsilon_{n+1} \times \varepsilon_{n+2} = 0$ because of independence

- 95% CI is $y_{n+2} = \hat{y}_{n+2} \pm 2s\sqrt{(1+\hat{\phi}_1^2)}$

- CI getting larger as move further away from y_n

4 Forecasting k-steps ahead ($\hat{y}_{n+k}$) and 95% Prediction Interval?

- We know $y_{n+1} = \underbrace{\hat{\phi}_0 + \hat{\phi}_1 y_n}_{\text{Forecast } \hat{y}_{n+1}} \underbrace{\pm 2s}_{CI}$

11

20/04/15

FORECASTING

ARMA MODELS

$$y_{n+2} = \underbrace{\hat{\phi}_0 + \hat{\phi}_0 \times \hat{\phi}_0 + \hat{\phi}_1^2 y_n}_{\text{forecast}} \pm \underbrace{2S\sqrt{1+\hat{\phi}_1^2}}_{CI}$$

and
$$y_{n+3} = \hat{\phi}_0 + \hat{\phi}_1 y_{n+2} + \varepsilon_{n+3}$$
$$= \hat{\phi}_0 + \hat{\phi}_1 (\hat{\phi}_0 + \hat{\phi}_1 y_{n+1} + \varepsilon_{n+2}) + \varepsilon_{n+3}$$
$$= \hat{\phi}_0 + \hat{\phi}_1 (\hat{\phi}_0 + \hat{\phi}_1 (\hat{\phi}_0 + \hat{\phi}_1 y_n + \varepsilon_{n+1}) + \varepsilon_{n+2}) + \varepsilon_{n+3}$$
$$= \underbrace{\hat{\phi}_0 + \hat{\phi}_1 \hat{\phi}_0 + \hat{\phi}_1^2 \phi_0 + \hat{\phi}_1^3 y_n}_{\text{Forecast}} + \underbrace{\hat{\phi}_1^2 \varepsilon_{n+1} + \hat{\phi}_1 \varepsilon_{n+2} + \varepsilon_{n+3}}_{\text{error term}}$$

PI : $y_{n+3} = \hat{\phi}_0 + \hat{\phi}_1 \hat{\phi}_0 + \hat{\phi}_1^2 \hat{\phi}_0 + \hat{\phi}_1^3 y_n \pm 2S\sqrt{1+\hat{\phi}_1^2+\hat{\phi}_1^4}$

Propose the following formula:
$$y_{n+k} = \underbrace{\hat{\phi}_0 \left(\Sigma_{i=1}^{k} \hat{\phi}_1^{i-1}\right) + \hat{\phi}_1^{k} y_n}_{\text{Forecast}} + \underbrace{\Sigma_{i=1}^{k} \hat{\phi}_1^{i-1} \varepsilon_{n+k-i-1}}_{\text{error}} \quad 9.1$$

implying
$$y_{n+k} = \underbrace{\hat{\phi}_0 \left(\Sigma_{i=1}^{k} \hat{\phi}_1^{i-1}\right) + \hat{\phi}_1^{k} y_n}_{\text{Forecast}} \pm \underbrace{2S \sqrt{\sum_{i=1}^{k} \hat{\phi}_1^{2(i-1)}}}_{CI}$$

- By induction 9.1 eq^n is valid at step $k+1$
- Width of CI depends on term $\Sigma_{i=1}^{k} \hat{\phi}_1^{2(i-1)}$
- We recognise a geometric series and its limit:
$$\lim_{k\to\infty} \Sigma_{i=1}^{k} \hat{\phi}_1^{2(i-1)} = \frac{1}{1-\hat{\phi}_1^2}$$
- So for an AR(1) model, the CI is growing up to a finite limit
- it is bounded.

MA(q): Moving Average Process

Definition: A moving average model of order 1 is a time series model defined as:

$$y_t = \psi_0 - \psi_1 \varepsilon_{t-1} + \varepsilon_t$$

where ε_t are independent errors $\varepsilon_t \sim N(0, \sigma^2)$

A moving average model of order q, MA(q), defined as

$$y_t = \psi_0 + \psi_1 \varepsilon_{t-1} - \psi_2 \varepsilon_{t-2} - \dots - \psi_q \varepsilon_{t-q} + \varepsilon_t$$

- Errors are now used as explanatory variables in MA models.
- to $y_1 \dots y_n$, can write n eq^ns with convention $\varepsilon_0 = 0$:

$$\begin{cases} y_1 = \psi_1 \varepsilon_0 + \varepsilon_1 \\ y_2 = \psi_1 \varepsilon_1 + \varepsilon_2 \\ y_3 = \psi_1 \varepsilon_2 + \varepsilon_3 \\ y_n = \psi_1 \varepsilon_{n-1} + \varepsilon_n \end{cases} = \quad \begin{aligned} y_1 &= \varepsilon_1 \\ y_2 &= \psi_1 y_1 + \varepsilon_2 \\ y_3 &= \psi_1 (y_2 - \psi_1 y_1) + \varepsilon_3 \\ y_n &= \psi_1 y_{n-1} - \psi_1^2 y_{n-2} + \dots + (-\psi_1)^{n-1} y_1 + \varepsilon_n \end{aligned}$$

- estimate ψ_1 by minimising $S\varepsilon^2$, system is not linear, powers appear

FORECASTING

ARIMA MODELS

ARMA: AutoRegressive Moving Average Models

Definition: Combining AR and MA models, we can define ARMA(p,q) model as:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \psi_0 - \psi_1 \varepsilon_{t-1} - \ldots - \psi_q \varepsilon_{t-q} + \varepsilon_t$$

with p the order of the AR part, q the order of the MA part. ψ_0 and ϕ_0 can be put together to define a unique constant c

$$y_t = c + \phi_1 y_{t-1} + \ldots \phi_p y_{t-p} - \psi_1 \varepsilon_{t-1} \ldots - \psi_q \varepsilon_{t-q} + \varepsilon_t$$

- With ARMA models, hard to identify p, q using ACF and PACF function
- The parameters $\{\phi_i\}$ computed using minimizing sum of square errors

Stationary in mean and Variance

- A TS is called stationary in mean if it randomly fluctuates about a constant mean level
- Stationary in mean ⇒ no trend/cycle or seasonality

Stationary:

- The mean is constant (Stationary in mean)
- The variance is finite (Stationary in variance)
- The correlation between values in the time series depends only on the time distance between these values (Stationary in autocorrelation)

- Stationary in variance ⇒ TS is said to be stationary in variance if the variance in the TS does not change with time

- ARMA models can not handle TS that are not stationary in mean and variance.
- ARMA models should only be fitted on TS that are stationary in mean and stationary in variance (i.e. no trend or no seasonal pattern)

Using ACF and PACF to Select MA(q) and AR(p) Models

Model	ACF	PACF
AR(1)	Exponential decay: on + side if $\phi_1 > 0$ and alternating in sign, if $\phi_1 < 0$ starts with - side	Spike at lag 1, then 0, + side if $\phi_1 > 0$ and - if $\phi_1 < 0$
AR(p)	Exponential decay or damped sine wave. Exact pattern depends on signs and sizes of $\phi_1 \dots \phi_p$	Spikes at lag 1 to p, then zero
MA(1)	Spikes at lag 1, then 0, + spike if $\psi_1 < 0$ and - spike if $\psi_1 > 0$	Exponential decay, on + side if $\psi_1 < 0$ and alternating in sign, - side if $\psi_1 > 0$
MA(p)	Spikes at lag 1 to q then zero	Exponential decay or damped sine wave. Exact pattern depends on signs and sizes of $\psi_1 \dots \psi_p$

ACF and PACF

Definition ACF at lag k computed by

$$ACF(k) = \frac{E[(y_t - E[y_t])(y_{t-k} - E[y_{t-k}])]}{\sqrt{var[y_t]\, var[y_{t-k}]}}$$

In TS we want to measure the relationship between y_t and y_{t-k} when the effects of other time lags 1, 2.. k-1 have been removed. Autocorrelation doesn't measure this. PACF does

Definition: PACF of a TS at lag k as denoted α_k is found as follows.

1. Fit a linear regression of y_t to the first k lags (ie an AR(k) model)

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_k y_{t-k} + \varepsilon_t$$

2. Then $\alpha_k = \hat{\phi}_k$ the fitted value ϕ_k from regression (least squares)

20/04/15

FORECASTING

ARIMA MODELS

1 AR model with $\phi_0 = 0$, Show PACF coefficients are zero when $k > 1$

Model 1) $y_t = \phi_1 y_{t-1} + \varepsilon_t$

Computing PACF at order 2 for instance implies to fit an AR(2) model to our AR(1): $y_t = \phi_1 y_{t-1} + 0 y_{t-2} + \varepsilon_t$

- therefore pacf coefficient at lag 2 is 0. Same reasoning for $k > 1$
- At lag $k=1$, pacf coefficient is ϕ_1

2 Assume a MA(1) model with $\psi_0 = 0$

What is $\mathbb{E}[y_t]$?

$$\mathbb{E}[y_t] = \mathbb{E}[\psi_1 \varepsilon_{t-1} + \varepsilon_t]$$
$$= \psi_1 \mathbb{E}[\varepsilon_{t-1}] + \mathbb{E}[\varepsilon_t]$$
$$= \psi_1(0) + (0)$$
$$= 0$$

Variance of y_t?

$$Var[y_t] = \mathbb{E}[(y_t - \mathbb{E}[y_t])^2]$$
$$= \mathbb{E}[y_t^2] \quad (\mathbb{E}[y_t] = 0)$$
$$= \mathbb{E}[(\psi_1 \varepsilon_{t-1} + \varepsilon_t)^2]$$
$$= \mathbb{E}[\psi_1^2 \varepsilon_{t-1}^2 + \varepsilon_t^2 + 2\psi_1 \varepsilon_{t-1}\varepsilon_t]$$
$$= \psi_1^2 \mathbb{E}[\varepsilon_{t-1}^2] + \mathbb{E}[\varepsilon_t^2] + 2\psi_1 \mathbb{E}[\varepsilon_{t-1}\varepsilon_t]$$
$$= \psi_1^2 \sigma^2 + \sigma^2 + 2\psi_1(0)$$

$\varepsilon_t \sim N(0, \sigma^2)$ $\varepsilon_i \neq \varepsilon_j$ & $i \neq j$ independence

Covariance of y_t and y_{t-k}?

$$Cov[y_t, y_{t-k}] = \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-k} - \mathbb{E}[y_{t-k}])]$$
$$= \mathbb{E}[(y_t)(y_{t-k})] \quad \text{because } \mathbb{E}[y_t] = 0 \ \forall t$$
$$= \mathbb{E}[(\psi_1 \varepsilon_{t-1} + \varepsilon_t)(\psi_1 \varepsilon_{t-1-k} + \varepsilon_{t-k})]$$
$$= \psi_1^2 \mathbb{E}[\varepsilon_{t-1-k}\varepsilon_{t-1}] + \psi_1 \mathbb{E}[\varepsilon_{t-k}\varepsilon_{t-1}] + \psi_1 \mathbb{E}[\varepsilon_t \varepsilon_{t-1-k}] + \mathbb{E}[\varepsilon_t \varepsilon_{t-k}]$$

$= 0 \ \forall k \geq 1$ $\quad 0 \ \forall k > 1;\ \sigma^2$ for $k=1$ $\quad 0 \ \forall k$ $\quad 0 \ \forall k > 0$

So $cov[y_t, y_t]$ $(\psi_1^2+1)\sigma^2$,
$cov[y_t, y_{t-1}] = \psi_1\sigma^2$
$cov[y_t, y_{t-k}] = 0 \quad \forall k \geq 1$

Correlation of y_t and y_{t-k}
Corr = cov/variance

$$Corr = \frac{Cov[y_t, y_{t-k}]}{\sqrt{Var[y_t]Var[y_{t-k}]}} \Rightarrow \begin{cases} 1 & \text{if } k=0 \\ \psi_1/\psi_1^2+1 & \text{if } k=1 \\ 0 & \text{otherwise } t>1 \end{cases}$$

3 Form of the ACF function for a MA(1) model.
ACF plots the lags k on the x axis and the y axis reports the correlation $Corr[y_t, y_{t-k}]$

Least Squares algorithm for MA models?

- MA(1) with $\psi_0 = 0$: $y_t = \psi_1 \varepsilon_{t-1} + \varepsilon_t$
- rewrite it: $y_t = \psi_1 y_{t-1} - \psi_1^2 \varepsilon_{t-2} + \varepsilon_t$

$$= \psi_1 y_{t-1} - \psi_1^2 y_{t-2} + \psi_1^3 \varepsilon_{t-3} + \varepsilon_t$$

$\vdots$

$$y_t = \psi_1 y_{t-1} - \psi_1^2 y_{t-2} + \dots + (-1)^t \psi_1^{t-1} y_1 + \psi_1^t \varepsilon_0 + \varepsilon_t$$

Cannot use least squares $\Rightarrow$ not linear

Backshift Operator

Definition: Backshift operator used to denote a lagged series by using B:

$$By_t = y_{t-1}$$

For lags of length k, we apply B, k times:

$$y_{t-2} = By_{t-1} = B^2 y_t \qquad \text{In general: } B^k y_t = y_{t-k}$$

- Can use for differentiating: $y'_t = y_t - y_{t-1} = y_t - By_t = (1-B)y_t$
- Backshift operator is multiplicative ie add powers.

21/04/15

FORECASTING

ARIMA MODELS

AIC and BIC

$AIC = -2\log(L) + 2m$ Akaike

$BIC = -2\log(L) + m\log n$ Bayesian

- L is the likelihood of the data with a certain model
- N is number of observations
- M is number of parameters in the model = $m = p+q$ for an ARMA(p,q) model

- Penalises model complexity against accuracy

CIS: ARIMA(p,d,q)

- ARMA no good for TS with trend ⇒ not stationary in mean and variance

Differencing a time series

- TS y_t, first order differencing: $y'_t = y_t - y_{t-1}$
  use B $= y_t - By_t = (1-B)y_t$

Integrated Differencing

Definition: ARIMA(p,d,q) Trends in a TS can be removed by differencing the TS

$$\underbrace{(1-\phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)}_{AR(p)} \underbrace{(1-B)^d}_{I(d)} y_t = c + \underbrace{(1-\psi_1 B - \psi_2 B^2 - \dots - \psi_q B^q)\varepsilon_t}_{MA(q)}$$

Which ARIMA (p,d,q) to use?

- p, d, q < 3
- Many ~~similar~~ different ARIMA models give similar predictions

1. Plot the data.
2. Check if data is stationary. Look at ACF, PACF. Stationary is implied by the ACF or PACF dropping quickly to zero

3. If non stationary, difference the data. Practically at most, two differences need to be taken to reduce a series to stationary. Verify stationarity by plotting the differenced series and looking at the ACF and PACF

4. Once stationarity obtained, look at ACF and PACF to see if there is any remaining pattern. Check against the theoretical behaviour of the MA and AR models to see if they fit

5. Minimise AIC or BIC

~~SEAS~~

SEASONAL ARIMA$(p,d,q)(P,D,Q)_s$

- TS having a trend and/or seasonality and are not stationary in mean
- ARIMA models can not cope with seasonality, only trend

Definition: Seasonal Autoregressive Integrated Moving average: ARIMA$(p,d,q)(P,D,Q)_s$

$$\underbrace{(1-\phi_1 B-\phi_2 B^2-\dots-\phi_p B^p)}_{AR(p)}\underbrace{(1-\beta_1 B^s-\beta_2 B^{2s}-\dots-\beta_P B^{Ps})}_{AR_s(P)}\underbrace{(1-B)^d}_{I(d)}\underbrace{(1-B^s)^D}_{I_s(D)} y_t =$$

$$= c + \underbrace{(1-\psi_1 B-\psi_2 B^2-\dots-\psi_q B^q)}_{MA(q)}\underbrace{(1-\theta_1 B^s-\theta_2 B^{2s}-\dots-\theta_Q B^{Qs})}_{MA_s(Q)}\varepsilon_t$$

- AR(p) auto regressive part of order p
- MA(q) moving average part of order q
- I(d) integrated differencing order d
- $AR_s(P)$ Seasonal Autoregressive part of order P
- $MA_s(Q)$ Seasonal Moving Average part of order Q
- $I_s(D)$ Seasonal differencing of order D
- S is the seasonal pattern appearing i.e S=12 etc

21/04/15.

FORECASTING 19

ARIMA MODELS

- Idea behind seasonal ARIMA is to look at what the best explanatory variables to model a seasonal pattern
- Australian beer, seasonal pattern of 12 months
- Best predictors are y_{t-12}, y_{t-24}... or ε_{t-12} ε_{t-24} etc

Using ACF and PACF to identify seasonal ARIMA's

- Can use ACF and PACF to identify P or Q
- For ARIMA $(0,0,0)(P,0,0)_s$ should see major peaks on PACF at S, 2S ... PS. On the ACF, the coefficients at lags S, 2S, 3S ... PS should form an exponential decrease or damped sinewave

- ARIMA $(0,0,0)(0,0,Q)_s$ - major peaks on ACF at S, 2S ... QS. On PACF coef at lags S, 2S ... QS should form an exponential decrease or damped sine wave

- When identifying P, Q, only look at multiples of S

- Can't deal with TS that are not stationary in variance

Preparing Time Series for Analysis

- Seasonal ARIMA$(p,d,q)(P,D,Q)_s$ can only deal with TS with stationary variance
- Amplitude increasing over time in time series → typical of TS which is not ~~stat~~ stationary in variance
- Mathematical functions can be used to make stationary in variance
- 4 ways reduce variance by differing amounts. Which one to use depends on how much the variance is increasing with time

Square root	$\sqrt{y_i}$	↓
Cube root	$\sqrt[3]{y_i}$	Increasing
Log	$\log(y_i)$	Strength
Negative Reciprocal	$-1/y_i$	↓

- Sometimes TS need to be normalised or adjusted to fit a model
- Months 28-31 days
- Average month length = 365.25/12
- Average month length / no days in month i:

$w_i = y_i \times$ average month length / no days in month i

$= y_i \times 365.25 / (12 \times$ no. days in month i.)

- Trading days

$w_i = y_i \times \frac{\text{no. of trading days in an average month}}{\text{no. trading days in month i.}}$

SES = ARIMA (0,1,1)

Holt linear Meth DES = ARIMA (0,2,2).

Holt additive = ARIMA (0,1,S+1)(0,1,0)$_S$

Holt multiplicative no ARIMA equivalent.

$y_t = c + \alpha \varepsilon_{t-4} + \varepsilon_t$ $\qquad \varepsilon_{t-4} = y_{t-4} - c - \alpha\varepsilon_{t-8}$

$= c + \alpha(y_{t-4} - c - \alpha\varepsilon_{t-8}) + \varepsilon_t$

$= c + \alpha y_{t-4} - \alpha c - \alpha^2\varepsilon_{t-8} + \varepsilon_t$ $\qquad \varepsilon_{t-8} = y_{t-8} - c - \alpha\varepsilon_{t-12}$

$= c + \alpha y_{t-4} - \alpha c - \alpha(y_{t-8} - c - \alpha\varepsilon_{t-12})$

$= (1 - \alpha + \alpha^2 - \alpha^3)c + \alpha y_{t-4} - \alpha^2 y_{t-8} - \alpha^3 y_{t-12} - \alpha^4 \varepsilon_{t-12}$ etc

α lag 0 α^2 lag 4 ... etc

$|\alpha| < 0 \Rightarrow$ exponential decrease or damped sine wave

$E[y_t] = E[c + \alpha\varepsilon_{t-4} + \varepsilon_t]$

$= E[c] + E[\alpha\varepsilon_{t-4}] + E[\varepsilon_t]$

$= c + \alpha E[\varepsilon_{t-4}] + E[\varepsilon_t]$

$= c$

$Var[y_t] = E[(y_t - E[y_t])^2]$ $\qquad E[y_t] = c$

$= E[(c + \alpha\varepsilon_{t-4} + \varepsilon_t - c)^2]$

$= E[(\alpha\varepsilon_{t-4} + \varepsilon_t)^2]$

$= E[\alpha^2\varepsilon_{t-4}^2 + 2\alpha\varepsilon_{t-4}\varepsilon_t + \varepsilon_t^2]$

$= \alpha^2 E[\varepsilon_{t-4}^2] + 2\alpha E[\varepsilon_{t-4}\varepsilon_t] + E[\varepsilon_t^2]$

$= \alpha^2\sigma^2 + \sigma^2$

$= (1+\alpha^2)s^2 = Var[y_t]$

$Cov[y_t, y_{t+h}] = E[(y_t - E[y_t])(y_{t+h} - E[y_{t+h}])]$

$E[(y_t - E[y_t])(y_{t+h} - E[y_{t+h}])]$

$[(\cancel{c} + \alpha\varepsilon_{t-4} + \varepsilon_t - \cancel{c})(\cancel{c} + \alpha\varepsilon_{t-4+h} + \varepsilon_{t+h} - \cancel{c})]$

$\alpha^2\varepsilon_{t-4}\varepsilon_{t-4+h} + \varepsilon_{t-4}\varepsilon_{t+h} + \alpha\varepsilon_t\varepsilon_{t-4+h} + \varepsilon_t\varepsilon_{t+h}$

$= E$

if $h=0$ $\quad \alpha^2 s^2 + s^2$

if $h=4$ $\quad \alpha\, \delta^2$ $\qquad \alpha s^2$

0 otherwise

$Corr = \dfrac{Cov[y_t, y_{t+h}]}{\sqrt{Var[y_t]\, Var(y_{t+h})}} = (1+\alpha^2)s^2$ $\qquad h=0 \to 0$

$h = 4 \quad = \dfrac{\alpha}{1+\alpha^2}$

otherwise $= 0$

Assumption $\varepsilon_i \sim N(0, \sigma^2)$ ε_i ε_T independent if $i \neq T$.

Likelihood $p(\varepsilon_1, \varepsilon_2 \ldots \varepsilon_n)$ $\prod_{i=1}^{n} (p(\varepsilon_i))$

$\log p = \log \prod_{i=1}^{n} (p(\varepsilon_i))$

$\log p = \log \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\varepsilon_i^2}{2\sigma^2}\right]$

$\log p = \sum_{i=1}^{n} \left(\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{\varepsilon_i^2}{2\sigma^2}\right)$

$\log p = \sum_{i=1}^{n} \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{\varepsilon_i^2}{2\sigma^2}\right)$

$-2\log p = \sum_{i=1}^{n} \log(2\pi\sigma^2) \; \Sigma \frac{\varepsilon_i^2}{2\sigma^2}$ $\qquad \cancel{\Sigma} = S^2 \approx \frac{\Sigma \varepsilon_i}{2}$

$n \log 2\pi\sigma^2$

$= n\log 2\pi\sigma^2 \quad + n$

$-2\log(L) = n\log 2\pi\sigma^2 + n$

$AIC = -2\log L + 2m$

$= n\log 2\pi\sigma^2 + n \quad + 2m$

$= n\log 2\pi + n\log\sigma^2 + n + 2m$

AR(1) $y_i = \phi_0 + \phi_1 y_{i-1} + \varepsilon_i$

$y_{n+1} = \phi_0 + \phi_1 y_n + \varepsilon_{n+1}$

we know ϕ_0 ϕ_1 and y_n $\Rightarrow$ $y_{n+1} = \phi_0 + \phi_1 y_n \pm 2s$ $\quad s^2 = \frac{\sum \varepsilon_i^2}{n-3}$

$y_{n+2} \quad \phi_0 + \phi_1 y_{n+1} + \varepsilon_{n+2}$ $\qquad$ Sub in $y_{n+1} = \phi_0 + \phi_1 y_n + \varepsilon_{n+1}$

$= \phi_0 + \phi_1(\phi_0 + \phi_1 y_n + \varepsilon_{n+1}) + \varepsilon_{n+2}$

$= \phi_0 + \phi_1 \phi_0 + \phi_1^2 y_n + \phi_1 \varepsilon_{n+1} + \varepsilon_{n+2}$

$E[\phi_1 \varepsilon_{n+1} + \varepsilon_{n+2}] = 0$

$Var[(\phi_1 \varepsilon_{n+1} + \varepsilon_{n+2})^2] = Var = E[\phi_1^2 \varepsilon_{n+1}^2 + 2\phi_1 \varepsilon_{n+1}\varepsilon_{n+2} + \varepsilon_{n+2}^2]$

$= \phi_1^2 s^2 + s^2$

$y_{n+2} = \hat{y}_{n+2} \pm 2\sqrt{(1+\phi_1^2)}$

$y_{n+k} \quad \phi_0 \sum_{i=1}^{k} \phi_1^{i-1} + \phi_1^k y_n + \sum_{i=1}^{k} \varepsilon_{n+k-i+1} \phi_1^{(i-1)}$

y_{n+k} 95% $\phi_0 \sum_{i=1}^{k} \phi_1^{i-1} + \phi_1^k y_n \pm 2s\sqrt{\sum_{i=1}^{k} \phi_1^{2(i-1)}}$

MA(1) $E[y_t] = E[\psi_1 \varepsilon_{t-1} + \varepsilon_t] = 0$

$Var = E[(y_t - E[y_t])^2] = E[(\psi_1 \varepsilon_{t-1} + \varepsilon_t)^2]$

$E(\psi_1^2 \varepsilon_{t-1}^2 + \psi_1 \varepsilon_{t-1}\varepsilon_t + \varepsilon_t^2)$

$= \psi_1^2 s^2 + s^2$

$cov[y_t, y_{t-k}] = E[(y_t - E[y_t])(y_{t-k} - E[y_{t-k}])]$

$E[(\psi_1 \varepsilon_{t-1} + \varepsilon_t)(\psi_1 \varepsilon_{t-1-k} + \varepsilon_{t-k})]$

$\psi_1^2 \varepsilon_{t-1}\varepsilon_{t-1-k} + \psi_1 \varepsilon_{t-1}\varepsilon_{t-k} + \psi_1 \varepsilon_t \varepsilon_{t-1-k} + \varepsilon_t \varepsilon_{t-k}$

$k=0 \quad \psi_1^2 s^2 + s^2$ $\qquad$ corr $= \frac{cov(y_t\, y_{t+k})}{Var(y_t)\, var(y_{t+k})}$ $\overline{\psi_1^2 + 1}$ $\qquad = 1$

$k=1 \quad \psi_1 s^2$ $\qquad\qquad\qquad\qquad\qquad\qquad \frac{\psi_1}{\psi_1^2+1}$

other $= 0$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ 0 other

FORECASTING

03/05/15

Time Series Components

Trend: A long term increase or decrease occurs

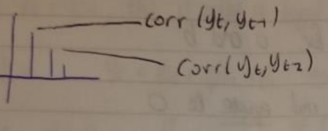
Seasonal: Series influenced by seasonal factors. Thus, series exhibits a behaviour that more or less repeats over a fixed period of time, such as a year.

Error: Corresponds to random fluctuations that cannot be explained by a deterministic pattern

ACF

- Similarity between observations as a function of the time lag between them
- Measure of how much the current value is influenced by previous values

For TS $y_1 \dots y_n$, autocorrelation at lag k is: $\frac{\sum_{t=k+1}^{n} (y_t-\bar{y})(y_{t-k}-\bar{y})}{\sqrt{\sum_{t=1}^{n} (y_t-\bar{y})^2 \sum (y_{t-k}-\bar{y})^2}}$



PACF

- Extension of autocorrelation, where the dependence on the intermediate elements (those within lag) are removed
- ~~Found by~~ Measures relationship between y_t and y_{t-k} when the effects of other time lags $1, 2, \dots k-1$ have been removed
- Fit a linear regression of y_t to the first k lags (ie. fit an AR(k) model to the ts).
  $y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} \cdots + \phi_k y_{t-k} + \varepsilon_t$
- Then as ϕ_k fitted value of y_{t-k} is our pacf value

SES - No Trend No Seasonal Component

Initialise $F_1 = y_1$ Choose $0 < \alpha < 1$ HoltWinters(ts, beta=FALSE, GAMMA=FALSE)

Forecast: $F_{t+1} = F_t + \alpha(y_t - F_t)$ $\alpha \approx 0$ ⇒ all forecasts first y

Until no more observations then: $F_{n+k} = F_{n+1} \quad \forall k \geq 1$ $\alpha = 1$ ⇒ all forecast previous value

SES forecasts are a weighted sum of ALL previous observations

DES ~ Holt's Linear Model Trend No Seasonality

Initialise $L_1 = y_1$ $b_1 = y_2 - y_1$ $F_1 = y_1$ Choose $0 \leq \alpha \leq 1$ $0 \leq \beta \leq 1$

Compute and forecast: $L_t = \alpha y_t + (1-\alpha)(L_{t-1} + b_{t-1})$

$b_t = \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}$

$F_{t+1} = L_t + b_t$

Until no more observations: $F_{n+k} = L_n + k b_n \quad \forall k \geq 1$ HoltWinters(ts, ~~beta=FALSE~~, GAMMA=FALSE)

HoltWinters(ts, gamma=FALSE) or alpha=0.2 " "

PERFORMANCE

SSE: $\sum_{t=1} (y_t - F_t)^2$

RMSE (Root mean Square Error): $\sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - F_t)^2} = \sqrt{\frac{SSE}{N}}$

MAPE (Mean Absolute Percent Error) $100 \frac{1}{n} \left| \frac{y_t - F_t}{y_t} \right|$

Holt Winters Exponential with Seasonality (and trend) HoltWinters(Beer)
- (α, β, γ)
- Additive or Multiplicative method HoltWinters(Beer, seasonal = "multiplicative")
- Select lowest SSE, RMSE or MAPE

Linear Regression
- $y_i = a + bx_i + \varepsilon_i$.
- ε_i is the error and is normally distributed with mean 0 and unknown variance σ^2
- ε_i and ε_j are independent when $i \neq j$

Fit model using least squares algorithm for a and b

$RSS = \sum_{i=1}^{n} (y_i - a - bx_i)^2$ Differentiate and equate to 0.

Correlation Coefficient: $r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 (\sum_{i=1}^{n} (y_i - \bar{y})^2)}}$

Coefficient of Determination $R^2 = r^2_{y\hat{y}} = \frac{\sum_{i=1}^{n} (\hat{y} - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$ SLR $\Rightarrow R^2 = r_{xy}^2$

Evaluating Model fit: look at residuals $\varepsilon_i = y_i - \hat{y}_i$
- Scatter plot of residuals against x_i show up any unmodelled pattern (have not fit the pattern of data)
- Normally distributed error assumption not correct. Histogram of residuals can usually verify this

Outliers: Unusually large residual - outlier. Standardised residuals give good indication

Can use binary variables to model seasonality

Least Squares Algorithm.

N observations $(x_i, y_i)_{i=1 \ldots n}$, write the following linear system.

$$\begin{cases} y_1 = a + bx_1 + \varepsilon_1 \\ y_2 = a + bx_2 + \varepsilon_2 \\ \vdots \\ y_n = a + bx_n + \varepsilon_n \end{cases}$$

$$\underset{y}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}} = \underset{X}{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}} \underset{\theta}{\begin{bmatrix} a \\ b \end{bmatrix}} + \underset{\varepsilon}{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}$$

$\hat{\theta} = (X^T X)^{-1} X^T Y$ (Least Squared Estimate)

In multiple $X = \begin{bmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ & \vdots & \\ 1 & x_n & z_n \end{bmatrix}$ $\theta = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$

FORECASTING 3

03/05/15

Autoregressive Models AR(P)

Definition: Autoregressive model of order p, states that y_i is the linear function of the previous p values of the series plus an error term:

$$y_i = \phi_0 + \phi_1 y_{i-1} + \phi_2 y_{i-2} + \dots + \phi_p y_{i-p} + \varepsilon_i$$

where $\phi_1 \dots \phi_p$ are weights we have to determine and $\varepsilon_i \sim N(0, \sigma^2)$

Only holds for $i > p$. Have to define $y_1, y_2 \dots y_p$ beforehand

AR(1) $\phi_0 + \phi_1 y_{i-1} + \varepsilon_i$

Have $n-1$ eqns

$$y_2 = \phi_0 + \phi_1 y_1 + \varepsilon_2$$
$$y_3 = \phi_0 + \phi_1 y_2 + \varepsilon_3$$
$$y_n = \phi_0 + \phi y_{n-1} + \varepsilon_n$$

1. Define $x_i = y_{i-1}$, this is called the lagged series. x_i only defined for $i = 2..n$

2. AR(1) is: $y_i = \phi_0 + \phi x_i + \varepsilon_i \Rightarrow$ linear regression mode use least squares estimate.

Estimate σ^2 by s^2: $s^2 = \frac{1}{n-1-2} \sum (y_i - \hat{\phi}_0 - \hat{\phi}_1 y_{i-1})^2$

Only $n-1$ equations in linear system to estimate $(\hat{\phi}_0, \hat{\phi}_1)$ and there are 2 parameters in our model

95% PI is $\hat{\phi}_0 + \hat{\phi}_1 x_i \pm 2s$

Prediction Interval for AR(1) k steps ahead

Forecast for y_{n+1}? 95% PI? $y_{n+1} = \hat{\phi}_0 + \hat{\phi}_1 y_n + \varepsilon_{n+1}$

We don't know value of $\varepsilon_{n+1} \sim N(0, s^2)$ but we know $\hat{\phi}_0 \hat{\phi}_1$ and y_n:

$$y_{n+1} = \hat{\phi}_0 + \hat{\phi}_1 \pm 2s \qquad s^2 = \sum \varepsilon_i^2 / n-3$$

Forecast for y_{n+2}? $y_{n+2} = \phi_0 + \phi_1 y_{n+1} + \varepsilon_{n+2}$

We don't know y_{n+1}, so we replace y_{n+1} by its expression wrt y_n:

$$\begin{aligned} y_{n+2} &= \phi_0 + \phi_1 y_{n+1} + \varepsilon_{n+2} \\ &= \phi_0 + \hat{\phi}_1 (\hat{\phi}_0 + \hat{\phi}_1 y_n + \varepsilon_{n+1}) + \varepsilon_{n+2} \\ &= \underbrace{\hat{\phi}_0 + \hat{\phi}_1 \hat{\phi}_0 + \hat{\phi}_1^2 y_n}_{\text{Forecast } \hat{y}_{n+2}} + \underbrace{\hat{\phi}_1 \varepsilon_{n+1} + \varepsilon_{n+2}}_{\text{error term}} \end{aligned}$$

Prediction Interval for y_{n+2}?

We know forecast $\hat{y}_{n+2}$ and error of this forecast. Need to estimate variance of error.

Compute mean first: $\mathbb{E}[\hat{\phi}_1 \varepsilon_{n+1} + \varepsilon_{n+2}] = \hat{\phi}_1 \mathbb{E}[\varepsilon_{n+1}] + \mathbb{E}[\varepsilon_{n+2}] = 0$

Compute variance of error term:

$$\mathbb{E}[(\hat{\phi}_1 \varepsilon_{n+1} + \varepsilon_{n+2})^2] = \hat{\phi}_1^2 \underbrace{\mathbb{E}[\varepsilon_{n+1}^2]}_{s^2} + 2\hat{\phi}_1 \underbrace{\mathbb{E}[\varepsilon_{n+1}\varepsilon_{n+2}]}_{0} + \underbrace{\mathbb{E}[\varepsilon_{n+2}^2]}_{s^2}$$

$$= \hat{\phi}_1^2 s^2 + s^2$$

95% PI = $y_{n+2} = \hat{y}_{n+2} \pm 2s\sqrt{(\hat{\phi}_1^2 + 1)}$

CI getting larger as we move any further from last observation y_n

Forecasting k steps ahead y_{n+k}?

We know $y_{n+1} = \hat\phi_0 + \hat\phi_1 y_n \pm 2s$

and $y_{n+2} = \hat\phi_0 + \hat\phi_1^2 y_n + \hat\phi_0\hat\phi_1 \pm 2s\sqrt{(\hat\phi_1^2+1)}$

and
$$y_{n+3} = \phi_0 + \phi_1 y_{n+2} + \varepsilon_{n+3}$$
$$= \hat\phi_0 + \hat\phi_1(\hat\phi_0 + \hat\phi_1 y_{n+1} + \varepsilon_{n+2}) + \varepsilon_{n+3}$$
$$= \hat\phi_0 + \hat\phi_1(\phi_0 + \hat\phi_1(\phi_0 + \hat\phi_1 y_n + \varepsilon_{n+1}) + \varepsilon_{n+2}) + \varepsilon_{n+3}$$
$$= \underbrace{\hat\phi_0 + \hat\phi_1\hat\phi_0 + \hat\phi_1^2\hat\phi_0 + \hat\phi_1^3 y_n}_{\text{forecast } \hat y_{n+3}} + \underbrace{\hat\phi_1^2\varepsilon_{n+1} + \hat\phi_1\varepsilon_{n+2} + \varepsilon_{n+3}}_{\text{error term}}$$

95% PI: $y_{n+3} = \hat\phi_0 + \hat\phi_1\hat\phi_0 + \hat\phi_1^2\hat\phi_0 + \hat\phi_1^3 y_n \pm 2s\sqrt{1+\hat\phi_1^2+\hat\phi_1^4}$

Propose following formula: $y_{n+k} = \hat\phi_0\left(\sum_{i=1}^{k} \hat\phi_1^{i-1}\right) + \hat\phi_1^k y_n + \sum_{i=1}^{k} \hat\phi_1^{i-1}\varepsilon_{n+k-i-1}$ 9.1

Implying: $y_{n+k} = \hat\phi_0\left(\sum_{i=1}^{k} \hat\phi_1^{i-1}\right) + \hat\phi_1^k y_n \pm 2s\sqrt{\sum_{i=1}^{k} \hat\phi_1^{2(i-1)}}$ 9.2

By induction we can show that eq^n 4.1 is valid at step k+1:

$$y_{n+k+1} = \phi_0 + \hat\phi_1 y_{n+k} + \varepsilon_{n+k+1}$$
$$= \hat\phi_0 + \hat\phi_1\left(\phi_0\left(\sum_{i=1}^{k} \hat\phi_1^{i-1}\right) + \hat\phi_1^k y_n + \sum_{i=1}^{k} \hat\phi_1^{i-1}\varepsilon_{n+k-i-1}\right) + \varepsilon_{n+k+1}$$

SEE NOTES

Note that the width of CI depends on term: $\sum_{i=1}^{k} \hat\phi_1^{2(i-1)}$

Recognise a geometric series and its limit: $\lim_{k\to\infty} \sum_{i=1}^{k} \hat\phi_1^{2(i-1)} = \frac{1}{1-\hat\phi_1^2}$

So for AR(1), the CI is growing up to a finite limit (it is bounded)

MA(q): Moving Average Process

Definition: MA(1) is defined as: $y_t = \psi_0 - \psi_1\varepsilon_{t-1} + \varepsilon_t$

where ε_t are independent errors, $\sim N(0,\sigma^2)$

A moving average model of order q, MA(q):
$$y_t = \psi_0 + \psi_1\varepsilon_{t-1} - \psi_2\varepsilon_{t-2} - \dots - \psi_q\varepsilon_{t-q} + \varepsilon_t$$

- Errors are now used as explanatory variables in MA model.
- Write it with convention $\varepsilon_0 = 0$

$$\begin{cases} y_1 = \psi_1\varepsilon_0 + \varepsilon_1 \\ y_2 = \psi_1\varepsilon_1 + \varepsilon_2 \\ y_3 = \psi_1\varepsilon_2 + \varepsilon_3 \\ \vdots \\ y_n = \psi_1\varepsilon_{n-1} + \varepsilon_n \end{cases} \quad = \quad \begin{aligned} & y_1 = \varepsilon_1 \\ & y_2 = \psi_1 y_1 + \varepsilon_2 \\ & y_3 = \psi_1(y_2 - \psi_1 y_1) + \varepsilon_3 \\ & y_n = \psi_1 y_{n-1} - \psi_1^2 y_{n-2} + \dots (-\psi_1)^{n-1} y_1 + \varepsilon_n \end{aligned}$$

- Estimate ψ_1 by minimising SSE, However System of equations is non linear wrt the parameter ψ_1 (powers of ψ_1 appear)
- Simple least squares algorithm used for linear regression and AR model cannot be used when there is an MA component in model

ARMA(p,q) Autoregressive Moving Average Models — No trend, No seasonality

Definition: Combining AR and MA models, ARMA(p,q).

$$y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \psi_0 - \psi_1 \varepsilon_{t-1} - \dots - \psi_q \varepsilon_{t-q} + \varepsilon_t$$

with p the order of the AR part and q order of MA. ψ_0 and ϕ_0 can be put together to define a unique constant c.

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \psi_1 \varepsilon_{t-1} - \dots - \psi_q \varepsilon_{t-q} + \varepsilon_t$$

Parameters ϕ_i are computed by minimising the SSE (algorithm - not do)

Expected value of MA(1) $y_t = \psi_0 - \psi_1 \varepsilon_{t-1} + \varepsilon_t$

$$E[\psi_0 - \psi_1 \varepsilon_{t-1} + \varepsilon_t] = \psi_0 \quad \text{or } 0 \text{ if } \psi_0 = 0$$

Same for MA(2)

No trend appearing in simulated ARMA model

Stationary in mean and variance: TS is stationary in mean if it randomly fluctuates about a constant mean level

- The mean is constant (Stationary in mean)
- The variance is finite (Stationary in variance) - it does not change with time
- The correlation between values in the TS depends only on the time distance between these values (Stationary in autocorrelation)

- ARMA models cannot handle TS that are not stationary in mean and variance
- ARMA should only be used on a TS that is stationary in mean and variance (i.e. no trend or seasonality)

Using ACF and PACF to select MA(q) and AR(p) models

Model	ACF	PACF
AR(1)	Exponential Decay: on + side if $\phi_1 > 0$ [and alternating signs, starting on - side if $\phi_1 < 0$] alt for negative	Spike at lag 1 then 0. + side if $\phi_1 > 0$ - side if $\phi_1 < 0$
AR(p)	Exponential decay or damped sine wave. Exact pattern depends on signs and sizes of $\phi_1 \dots \phi_p$	Spikes at lags 1 to p and then 0

6

$MA(q)$ Spike at lag 1 lag then zero

Exponential decay or damped sine wave. Exact pattern depends on the signs and sizes of ψ_1, ψ_2

Model	ACF	PACF
$MA(1)$	Spike at lag 1, then 0; + side if $\psi_1 < 0$ and − spike if $\psi_1 > 0$	Exponential decay on + side if $\psi_1 < 0$ and [alternating sign starting on + side if $\phi_1 < 0$]

Assuming $AR(1)$ model with $\phi_0 = 0$, show that the PACF coef are zero when $k > 1$

- By definition $y_t = \phi_1 y_{t-1} + \varepsilon_t$

- Computing the PACF at order 2 for instance, implies to fit a $AR(2)$ to an $AR(1)$

$$y_t = \phi_1 y_{t-1} + 0\, y_{t-2} + \varepsilon_t$$

- Therefore PACF coef at lag 2 is 0. Same reasoning for value of $k > 1$.

Assume a $MA(1)$ model with $\psi_0 = 0$

What is $\mathbb{E}[y_t]$?

$$\mathbb{E}[y_t] = \mathbb{E}[\psi_1 \varepsilon_{t-1} + \varepsilon_t]$$
$$= \psi_1 \mathbb{E}[\varepsilon_{t-1}] + \mathbb{E}[\varepsilon_t] = 0$$

Variance of y_t?

$$Var[y_t] = \mathbb{E}[(y_t - \mathbb{E}[y_t])^2]$$
$$= \mathbb{E}[(y_t)^2] \quad \text{Since } \mathbb{E}[y_t] = 0$$
$$= \mathbb{E}[(\psi_1 \varepsilon_{t-1} + \varepsilon_t)^2]$$
$$= \mathbb{E}[\psi_1^2 \varepsilon_{t-1}^2 + \varepsilon_t^2 + 2\psi_1 \varepsilon_{t-1}\varepsilon_t]$$
$$= \psi_1^2 \mathbb{E}[\varepsilon_{t-1}^2] + \mathbb{E}[\varepsilon_t^2] + 2\psi_1 \mathbb{E}[\varepsilon_{t-1}\varepsilon_t]$$
$$= \psi_1^2 \sigma^2 + \sigma^2$$

Covariance of y_t and y_{t-k}

$$Cov[y_t, y_{t-k}] = \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-k} - \mathbb{E}[y_{t-k}])] \quad \text{Def of covariance}$$
$$= \mathbb{E}[(y_t)(y_{t-k})] \quad \text{Because } \mathbb{E}[y_t] = 0 \ \forall t.$$
$$= \mathbb{E}[(\psi_1 \varepsilon_{t-1} + \varepsilon_t)(\psi_1 \varepsilon_{t-1-k} + \varepsilon_{t-k})]$$
$$= \mathbb{E}[\psi_1^2 \varepsilon_{t-1-k}\varepsilon_{t-1} + \psi_1 \varepsilon_{t-1-k}\varepsilon_t + \psi_1 \varepsilon_{t-k}\varepsilon_{t-1} + \varepsilon_t \varepsilon_{t-k}]$$

$$= \underbrace{\psi_1^2 \mathbb{E}[\varepsilon_{t-1-k}\varepsilon_{t-1}]}_{=0\ \forall k \geq 1} + \underbrace{\psi_1 \mathbb{E}[\varepsilon_{t-k}\varepsilon_{t-1}]}_{0\ \forall k > 1;\ \sigma^2 \text{ for } k=1} + \underbrace{\psi_1 \mathbb{E}[\varepsilon_t \varepsilon_{t-1-k}]}_{0\ \forall k} + \underbrace{\mathbb{E}[\varepsilon_t \varepsilon_{t-k}]}_{0\ \forall k > 0}$$

So $Cov[y_t, y_t] = (\psi_1^2 + 1)\sigma^2$

$Cov[y_t, y_{t-1}] = \psi_1 \sigma^2$

$Cov[y_t, y_{t-k}] = 0 \quad \forall k > 1$

Correlation of y_t, y_{t-k}

$$Corr[y_t, y_{t-k}] = \frac{Cov[y_t, y_{t-k}]}{\sqrt{Var[y_t]\, Var[y_{t-k}]}} = \begin{cases} 1 & \text{if } k = 0 \\ \dfrac{\psi_1}{\psi_1^2 + 1} & \text{if } k = 1 \\ 0 & \text{otherwise } k > 1 \end{cases}$$

ACF?

ACF plots the lags k on the x axis and y axis report the correlation $[y_t, y_{t-k}]$

FORECASTING

$AIC = -2\log(L) + 2m$
$BIC = -2\log(L) + m\log N$

AIC approximated by $\approx n(1+\log(2\pi)) + n\log(s^2) + 2m$

Assume $\varepsilon_i \sim N(0,\sigma^2)$ and ε_i and ε_j independent if $i \neq j$.

Likelihood $p(\varepsilon_1, \varepsilon_2 \dots \varepsilon_n)$.

$\log(p(\varepsilon_1, \varepsilon_2 \dots \varepsilon_n)) = \log\left[\prod_{i=1}^{n} p(\varepsilon_i)\right]$
$= \log\left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{\varepsilon_i^2}{2\sigma^2}\right]\right]$

$= \sum_{i=1}^{n}\left[\log\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{\varepsilon_i^2}{2\sigma^2}\right]$

$\log \alpha = \sum_{i=1}^{n}\left(-\frac{1}{2}\log(2\pi s^2) - \frac{1}{2}\frac{\varepsilon_i^2}{2\sigma^2}\right)$

$-2\log \alpha : \sum_{i=1}^{n}\log(2\pi s^2) + \varepsilon_i^2/s^2$
$= \sum_{i=1}^{n}\log(2\pi s^2) + \sum \varepsilon_i^2/s^2$
$= n\log(2\pi s^2) + 1/s^2 \sum \varepsilon_i^2$

$s^2 \approx \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2$ estimate of var

$\Rightarrow -2\log \alpha = n\log 2\pi s^2 + n$
$\Rightarrow$ sub this.

ARIMA(p,d,q)

can't really cope with trends

Differencing $y'_t = y_t - y_{t-1}$

Can use B to express differences $y'_t = y_t - y_{t-1} = y_t - By_t = (1-B)y_t$

ARIMA(p,d,q)

$$\underbrace{(1-\phi_1 B - \phi_2 B^2 \dots \phi_p B^p)}_{AR(p)}\underbrace{(1-B)^d}_{I(d)} y_t = c + \underbrace{(1-\psi_1 B - \psi_2 B^2 \dots - \psi_q B^q)}_{MA(q)}\varepsilon_t$$

Time $\Rightarrow$	Data	Differenced data	Fitted value	Error	Fitted value
t	y_t	$x_t = y_t - y_{t-1}$	$\hat{x}_t = \dots$	$\varepsilon_t = x_t - \hat{x}_t$	$\hat{y}_t = \hat{x}_t + y_{t-1}$

ARIMA(p,d,q)(P,D,Q)s

$$\underbrace{(1-\phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)}_{AR(p)}\underbrace{(1-\beta_1 B^s - \beta_2 B^{2s} - \dots - \beta_P B^{Ps})}_{AR_s(P)}\underbrace{(1-B)^d}_{I(d)}\underbrace{(1-B^s)^D}_{I_s(D)} y_t =$$

$$= c + \underbrace{(1-\psi_1 B - \psi_2 B^2 - \dots - \psi_q B^q)}_{MA(q)}\underbrace{(1-\theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs})}_{MA_s(Q)}\varepsilon_t$$

For ARIMA(0,0,0)(P,0,0)s major spikes on PACF at s, 2s Ps. On ACF coefficients at lags s, 2s … Ps should form an exponential decrease or damped sine wave

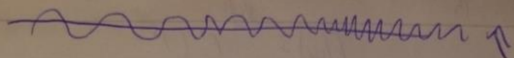
For ARIMA(0,0,0)(0,0,Q)s major spike on ACF at s, 2s … Qs. On PACF coef at lags s, 2s … Q should form exponential decrease or damped sine wave

3 1/12/14

Forecasting

ARMA (p,q)

constant variance



constant mean

$\varepsilon_t \sim N(0,\sigma^2)$ U also Stationary mean/variance from hypothesis
with ARIMA(p,d,q) (P,D,Q)

Removing trend using difference (D) indicates t.s. not stationary in mean

d: differencing used to remove trend. Trend indicates T.S. is not stationary in mean

P,D,Q introduced to deal with seasonality.
Seasonality in a time series indicates that the time series is not stationary in mean

ARIMA doesn't deal with time series with non constant variance

If not constant variance, transform data and create new t.s

if y_t not suitable for ARIMA (variance fluctuates over time) $X_t = f(y_t)$
Will need inverse of this function to transform prediction back

$y_t \xrightarrow{f_1} x_t \Rightarrow$ fit arima(x_t) — AIC/BIC in these models can not be compared

$\xrightarrow{f_2} z_t \Rightarrow$ fit ARIMA(z_t) —

$\downarrow$

SSE $y_t - y_{f_1}$
$y_t - y_{f_2}$

Pick the model that makes residuals have constant mean and variance
no other fuss

2
Suggested function
$f: \mathbb{R}$
$\mathbb{R}'$

log f^{-1}exp
$\sqrt{y_t}$
$f(y_t)^{1/3}$, $f^{-1}(\ )^3$

← most common

Prediction of y_{n+1}
95% prediction interval
We have $\hat{x}_{n+1}$ and 95% $[l_{n+1}, h_{n+1}]$

l_{n+1} $\hat{x}_{n+1}$ h_{n+1}

Prediction $y_{n+1} = f^{-1}(\hat{x}_{n+1})$ $[f^{-1}(l_{n+1}), f^{-1}(h_{n+1})]$
Will not have symmetry in new interval

Monthly data - adjust for number of days in month February 28 days Jan: 31
Adjust for number of trading days in data