22/04/15

FORECASTING
EXAM PAPER 2014

Q1 A. Trends - upward/downward TS goes up or down "Deterministic pattern".
Seasonality - TS Repeats itself every fixed period of time
Noise/error - Random fluctuation

Fig 1. - upward trend
Seasonality every 12 months
noise is random fluctuation

B. TS is not stationary in mean because it has a trend and also seasonality which indicates that the average sale of printing and writing paper

From figure 1 it is unclear if the TS is stationary in variance

C. - The timeplot indicates a fall in sales every August of every year (monthly data)
- A periodicity of 12 is suspected
- ACF - major autocorrelation coef at lag 12, 24, 36 indicates periodicity of 12
- PACF - Spike at lag 12 indicating

D i. Backshift operator is a mathematical operator to compute lags of time

$B y_t = y_{t-1}$

$B^2 y_t = \cancel{y} \; B B y_t = y_{t-2}$

$B^n y_t = y_{t-n}$

ii. $ARIMA(0,1,0)(0,1,0)_{12}$

$d=1 \quad D=1$

$(1-B^{12})^1 (1-B)^1 y_t = c + \varepsilon_t$

iii. $(1-B^{12})(1-B) = (1-B^{12} - B + B^{13}) y_t$

$y_t - y_{t-12} - y_{t-1} + y_{t-13} = c + \varepsilon_t$

E ARIMA $(0,1,1)(0,1,1)_{12}$

$S=12$ Seasonality $=12$, T.S is repeating itself every 12 months as seen From PACF and ACF.

$d=1$ mean is not Stationary

$D=1$ mean is not Stationary

$M(1)$ Moving average as major Spike on PACF at 1 and spikes on

$M(1)_s$ ACF at $S, 2S, 3S \ldots$

C Q2A SES Algorithm

$F_1 = y_1 \qquad 0 < \alpha < 1$

$F_{t+1} = F_t + \alpha(y_t - F_t)$

$F_{n+k} = F_{n+1} \qquad k \geq 1$

B Root mean Square Error $= \sqrt{\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2}$

$\varepsilon_t = y_t - F_t$

Mean absolute Percent error $= 100 \frac{1}{n} \sum_{t=1}^{n} \left| \frac{\varepsilon_t}{y_t} \right|$

Family of algorithms $SES(\alpha)$ that depends on $0 < \alpha < 1$

Choose α that minimises RMSE or MAPE.

$\hat{\alpha} = \text{argmin}_{\alpha}$ RMSE or $\hat{\alpha} = \text{argmin}_{\alpha}$ MAPE

Ci $y_i = \hat{\phi}_0 + \hat{\phi}_1 y_{i-1} + \varepsilon_i$

$y_{n+1} = \hat{\phi}_0 + \hat{\phi}_1 y_n + \varepsilon_{n+1}$ relevant to y_{n+1}.

We know $\varepsilon_{n+1} \sim N(0, \sigma^2)$ variance s^2.

$\mathbb{E}[y_{n+1}] = \mathbb{E}[\hat{\phi}_0 + \hat{\phi}_1 y_n + \varepsilon_{n+1}]$

$= \mathbb{E}[\hat{\phi}_0] + \hat{\phi}_1 \mathbb{E}[y_n] + \mathbb{E}[\varepsilon_{n+1}]$

$= \hat{\phi}_0 + \hat{\phi}_1 y_n$

FORECASTING
EXAM PAPER 2014 Q2

$Var[y_{n+1}] = \mathbb{E}[(y_{n+1} - \mathbb{E}[y_{n+1}])^2]$
$= \mathbb{E}[\varepsilon_{n+1}^2] = s^2$

95% PI $= [\hat{y}_{n+1} - 2s, \hat{y}_{n+1} + 2s]$

ii $y_{n+2} = \hat{\phi}_0 + \hat{\phi}_1 y_{n+1} + \varepsilon_{n+2}$
$= \hat{\phi}_0 + \hat{\phi}_1 (y_n + \varepsilon_{n+1}) + \varepsilon_{n+2}$
$= \hat{\phi}_0 + \hat{\phi}_1 y_n + \hat{\phi}_1 \varepsilon_{n+1} + \varepsilon_{n+2}$
$= \cancel{\hat{\phi}_0 + \hat{\phi}_1 y_n}$
$\hat{\phi}_0 + \hat{\phi}_1 \hat{\phi}_0 + \hat{\phi}_1^2 y_n$

$Var[y_{n+2}] = \mathbb{E}[\text{error term}^2]$
$= \mathbb{E}[(\hat{\phi}_1 \varepsilon_{n+1} + \varepsilon_{n+2})^2]$
$= \mathbb{E}[\hat{\phi}_1^2 \varepsilon_{n+1}^2 + 2\hat{\phi}_1 (\varepsilon_{n+1}\varepsilon_{n+2}) + \mathbb{E}[\varepsilon_{n+2}^2]]$
$= \hat{\phi}_1 s^2 + 2\hat{\phi}_1 + s^2$
independence

$= (1 + \hat{\phi}_1^2) s^2$

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

Di. $AIC = -2Log(L) + 2m$
AIC is used to choose the best ARIMA model
The model selected minimises the AIC.

- The term $-2Log(L)$ is similar to sum of squared errors.
- Same functionality as RMSE and MAPE but in addition has term 2m
- 2m term penalises complexity of models
- AIC allows a trade off between a model that goes as close as possible to observation values and a model that doesn't have too many explaining variables

Q 20 i Assumption $\varepsilon \sim N(0, \sigma^2)$ s^2 estimate of σ^2

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

Likelihood $p(\varepsilon_1, \varepsilon_2 \ldots \varepsilon_n)$ time n() the last term in IS

$$\log(p(\varepsilon_1, \varepsilon_2 \ldots \varepsilon_n)) = \log\left[\prod_{i=1}^{n} p(\varepsilon_i)\right]$$

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

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

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

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

$$-2\log\alpha = \sum_{i=1}^{n}\left(\log(2\pi s^2) \cancel{\ldots} + \varepsilon_i^2/s^2\right)$$

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

$$= n\log[2\pi s^2] + 1/s^2 \sum_{i=1}^{n} \varepsilon_i^2$$

$$s^2 \approx \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 \quad \text{(estimate of variance)}$$

$$-2\log\alpha = n\log(2\pi s^2) + n$$

Eq(2) $AIC = n + n\log 2\pi + n\log s^2 + 2m$

Eq(1) $AIC = -2\log\alpha + 2m$

$$= n\log(2\pi s^2) + n + 2m$$

$$= n\log(2\pi) + n\log(s^2) + n + 2m$$

Sub in and split term into 2.

22/04/15.

FORECASTING

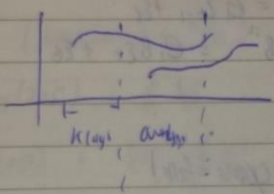
EXAM PAPER 2014

Q3

Q3 Ai. BIC - Bayesian Information Criterion

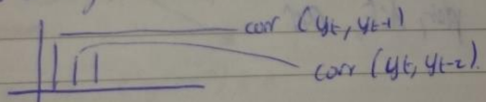
- Used to select best ARIMA model, model with lowest BIC
- $BIC = -2\log(L) + M \log N$
- n = ts length
- m = # parameters estimated
- L = Likelihood
- For ARIMA $m = p + d$
- For ARIMA $(p,d,q)(P,D,Q)_s$ $m = p + q + P + Q$
- → $-2\log(L)$ is proportional to $\sum \varepsilon_i^2$ (SSE)

Aii. ACF - Autocorrelation function - compares correlation between a time y_t and a lag k at y_{t+k}.



Correlation is a similarity measure, if y_t and y_{t+k} are similar, correl close to 1

ACF plot.



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

Aiii. PACF - Partial Autocorrelation function $(y_1 \dots y_n)$ TS

- Fit $Ar(k)$ $y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_k y_{t-k} + \varepsilon_t$
- $PACF[k] = \phi_k$.
- removes all info between y_{t+k} and the TS of y_t when the information is carried by previous explanatory variables y_{t-1} and $y_{t-(k+1)}$ is removed

6

B. $(1-B^2)y_t = \varepsilon_t$

$y_t - B^2 y_t = \varepsilon_t$

$y_t - y_{t-2} = \varepsilon_t$

ii. $(1+\phi_1 B - \phi_2 B^2)(1-B)y_t = (1-\theta_1 B)(1-B^{12})\varepsilon_t$

$1 - \phi_1 y_t - \phi_2 y_{t-2} - y_t + \phi_1 y_{t-2} + \phi_2 y_{t-3} = \varepsilon_t - \theta_1 \varepsilon_t - \varepsilon_{t-12} + \theta_1 \varepsilon_{t-13}$

C i. $y_t - y_{t-12} = \varepsilon_t$

$y_t - y_t B^{12} = \varepsilon_t$

ii. $y_t - \phi y_{t-1} - \phi_2 y_{t-12} = \theta \varepsilon_{t-1} + \varepsilon_t$

$y_t - \phi_1 B y_t - \phi y_t B^{12} = \theta_1 B \varepsilon_t + \varepsilon_t$

D i. arima || auto.arima

ii. Holtwinter (beta = false, gamma = false)

iii. predict || forecast

iv. tsdisplay || plot, ACF, PACF

Forecasting Exam Paper 2014

Q2 Di Assumptions $\varepsilon \sim N(0, \sigma^2)$

s^2 estimate of σ^2

Likelihood $p(\varepsilon_1, \varepsilon_2 \dots \varepsilon_n)$ - time n is the last datum in the time series

Residuals are independent

$\log(p(\varepsilon_1, \varepsilon_2 \dots \varepsilon_n)) = \log\left(\prod_{i=1}^{n} p(\varepsilon_i)\right)$

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

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

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

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

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

$= \sum_{i=1}^{n} \log(2\pi s^2) + \sum_{i=1}^{n} \frac{\varepsilon_i^2}{s^2}$

$= n\log(2\pi s^2) + \frac{1}{s^2} \sum_{i=1}^{n} \varepsilon_i^2$

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

Estimate of variance

$-2\log \alpha = n\log(2\pi s^2) + n$

eq(2) $AIC = n + n\log 2\pi + n\log s^2 + 2m$

eq(1) $AIC = -2\log \alpha + 2m$

$= n\log[2\pi s^2] + n + 2m$

$= n\log(2\pi) + n\log(s^2) + n + 2m$

2

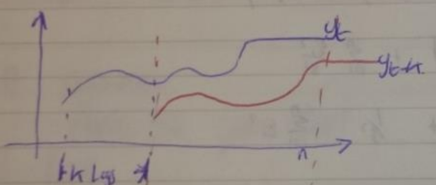
Q3

BIC - Bayesian Information Criterion

- BIC used to select the best ARIMA model, select model with the lowest BIC
- $BIC = -2 \log L + m \log n$
- n = time series length
- m = number of parameters estimated
- L = Likelihood
- For Arma $m = p+q$ ARIMA(p,d,q)
- for Seasonal ARIMA $m = p+q+P+Q$ ARIMA(p,d,q)(P,D,Q)s
- Try to minimise SSE without increasing too much the complexity of the model quantified by m.
  - $-2 \log L \propto \Sigma \varepsilon_i^2$ (SSE)
    proportional to

ACF - Auto correlation function

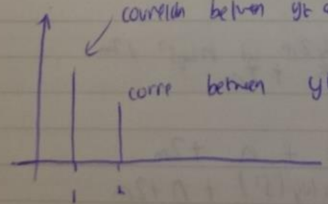
acf(k) - computes the correlation between a time y_t and a Lag k or y_{t+k}.



correlation is a similarity measure, where if y_t and y_{t-k} are similar, correlation is close to 1.

If y_t and y_{t-k} are dissimilar, correlation should be close to zero.

ACF plot



$$ACF(k) = \frac{\sum_{k=1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\left(\Sigma (y_t - \bar{y})^2 \; \Sigma (y_{t-k} - \bar{y})^2\right)^{1/2}}$$

12/12/14. Forecasting 2014 Exam Paper

3A iii PACF - Partial auto correlation function
$(y_1 \cdots y_n)$ time series
(i) fit $Ar(K)$:
$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots \phi_K y_{t-K} + \varepsilon_t$$

(2) $pacf(K) = \phi_K$
remove all info up until K values contribution
Capture the information between y_{t-K} and the time and y_t when the information carried by previous explanatory variables y_{t-1} and $y_{t-(K+1)}$ is removed

B i. $(1-B^2)y_t = \varepsilon_t$
$y_t - B^2 y_t = \varepsilon_t$

$y_t - y_{t-2} = \varepsilon_t$

ii $(1-\phi_1 B - \phi_2 B^2)(1-B)y_t = (1-\theta_1 B)(1-B^{12})\varepsilon_t$

$(1-\phi_1 B - \phi_2 B^2 - B + \phi_1 B^2 + \phi_2 B^3)y_t = (1-\theta_1 B - B^{12} + \theta_1 B^{13})\varepsilon_t$

$1-\phi_1 y_t - \phi_2 y_{t-2} - y_t + \phi_1 y_{t-2} + \phi_2 y_{t-3} = \varepsilon_t$ ~~$-\varepsilon_{t-1}$~~
$\varepsilon_t - \theta_1 \varepsilon_t - \varepsilon_{t-12} + \theta_1 \varepsilon_{t-13}$

C i $y_t - y_{t-12} = \varepsilon_t$

$y_t - y_t B^{12} = \varepsilon_t$

ii. $y_t - \phi_1 y_{t-1} - \phi_2 y_{t-12} = \theta_1 \varepsilon_{t-1} + \varepsilon_t$

$y_t - \phi B y_t - \phi_1 y_t B^{12} = \theta_1 B \varepsilon_t + \varepsilon_t$.

D i. arima | auto.arima
ii. Holtwinters (beta=false, gamma=false)
iii predict | Forecast
iv. tsdisplay | plot, ACF, pacf