Model Specification and Diagnostics

MAS 640 - Time Series Analysis and Forecasting

1/30/2018

Outline

- Chapter 6 - Model Specification
  - ACF/PACF/EACF for determining p, d, q
- Chapter 7 - Parameter Estimation
  - We will skip this
- Chapter 8 - Model Diagnostics
  - Does the model fit well?
  - Which model fits better?

Strategy

1. Plot the time series
2. Determine appropriate transformation for nonconstant variance
3. Difference or detrend as needed
4. Plot ACF, PACF, EACF to help identify appropriate order of p and q

Strategy

1. Plot the time series. Looking for
   - trends
   - seasonality
   - nonconstant variance
   - outliers
   - abrupt changes

Strategy

2. Determine appropriate transformation if nonconstant variance exists

   - Box-Cox procedure
   - Try several and stick with best

Strategy

3. Difference data as needed to remove trends
    - Will likely notice trend in time series plot
    - Can plot ACF on original series, ACF will decay *very* slowly
    - First difference likely sufficient, possibly 2nd order difference
    - *Rarely*, if ever, will you need to difference more than twice

Strategy

4. Plot sample ACF, PACF, and EACF on the stationary series to determine p and q

   - Can be original, transformed, differenced, or transformed/differenced series (whichever one is stationary)
   - p and q are generally never too high, say ≤ 4 (excluding seasonal models which we haven't covered yet)
     - Principle of Parsimony
   - Use knowledge of theoretical patterns in ACF, PACF, and EACF to guide selection of p and q
     - ACF cuts off at q for MA(q)
     - PACF cuts off at p for AR(p)
     - Wedge or upper left 0 in EACF table identifies p and q for ARMA(p, q) models

Identifying Model Order

- "The art of model selection is very much like the method of an FBI agent's criminal search. Most criminals disguise themselves to avoid being recognized."

- Similar for ACF, PACF, and EACF. Sampling variation can disguise theoretical patterns making it difficult to clearly ascertain p and q
- At the end of these steps...

  - probably have multiple candidate models
  - rare to have one definite model in practice

Next Steps (Chapter 8)

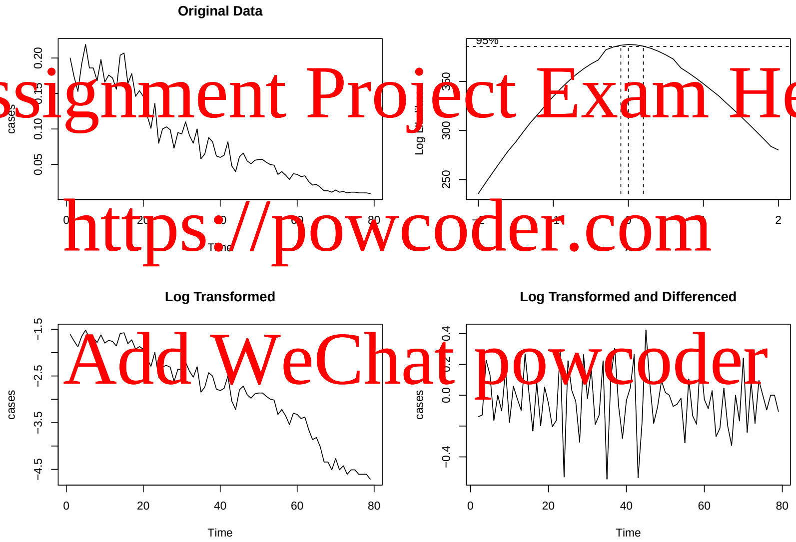
- Fit candidate models
- Compare them
- Diagnose assumptions

- Determine which seems most appropriate
- After selecting the "best" model, forecasting becomes the central focus (Chapter 9)

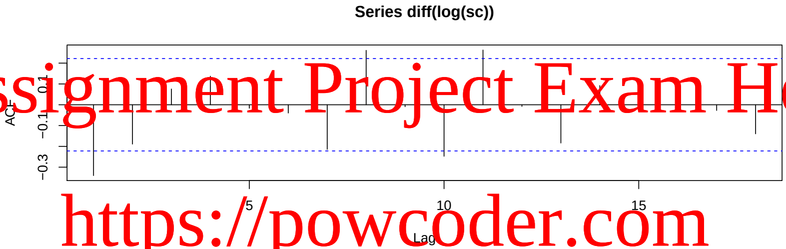
Supreme Court Data

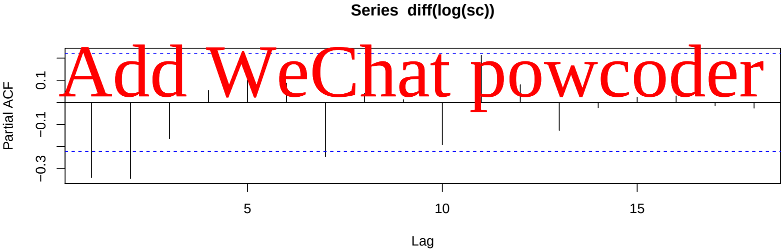
- The supreme court dataset represents the acceptance rate of cases appealed to the Supreme Court during 1926-2004.
- Convert to stationary, then identify p and q

Supreme Court Data



Supreme Court Data





Supreme Court Data

AR/MA

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	o	o	o	o	o	o	x	o	x	x	o	o	o
1	x	x	o	o	o	o	o	x	o	o	x	o	o	o
2	x	x	o	o	o	o	o	o	o	o	o	o	o	o
3	x	x	x	o	o	o	o	o	o	o	o	o	o	o
4	x	o	o	o	o	o	o	o	o	o	o	o	o	o
5	x	o	x	o	o	o	o	o	o	o	o	o	o	o
6	x	x	x	o	x	x	o	o	o	o	o	o	o	o
7	o	o	x	x	o	o	x	o	o	o	o	o	o	o

Supreme Court Data

- Based on everything, probably start with an MA(1) model
  - Why?


```
arima(log(sc), order=c(0, 1, 1))$coef
```

```
       ma1
-0.3556319
```

Supreme Court Data - MA(1) Model

After fitting an MA(1) model, we find the following regression equation

$$\nabla log(Y_t) = e_t - 0.356 e_{t-1}$$

- As with linear regression, we study residuals to determine appropriateness of model
- In time series, we want the residuals to look like white noise
  - Normally distributed
  - Independent
  - Constant variance (though this should have been taken care of in previous steps. . .)

Diagnostic Tools

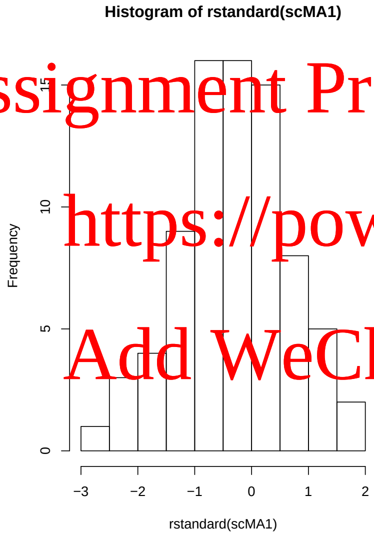
- **Normality**
  - Histogram of residuals
  - Normal QQ plot of residuals
  - Shapiro-Wilk Test

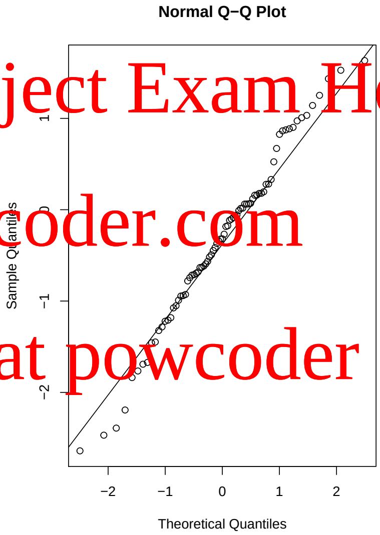
- **Independence**
  - Runs test
  - Plot of residual ACF
  - Ljung-Box Test

Normality


```
scMA1 <- arima(log(sc), order=c(0, 1, 1) )
hist(rstandard(scMA1)) # histogram
qqnorm(rstandard(scMA1)) # qqplot
qqline(rstandard(scMA1))
shapiro.test(rstandard(scMA1) ) # shapiro test
```

Histogram and QQ Plot





Shapiro-Wilk Test

- H_0: The standardized residuals are normally distributed
- H_a: The standardized residuals are not norally distributed


```
shapiro.test(rstandard(scMA1))
```



```
    Shapiro-Wilk normality test

data:  rstandard(scMA1)
W = 0.98632, p-value = 0.5639
```

Independence


```
runs(rstandard(scMA1)) # runs test
acf(rstandard(scMA1)) # residual ACF
Box.test(residuals(scMA1), lag=k, type='Ljung-Box', fitdf=1
tsdiag(scMA1) # produces relevant plots
```

Runs test

- The **runs test**
  - H_o: The standardized residuals are independent
  - H_a: The standardized residuals are independent


```r
runs(rstandard(scMA1))
```

```
$pvalue
[1] 0.864
```



```
$observed.runs
[1] 37
```

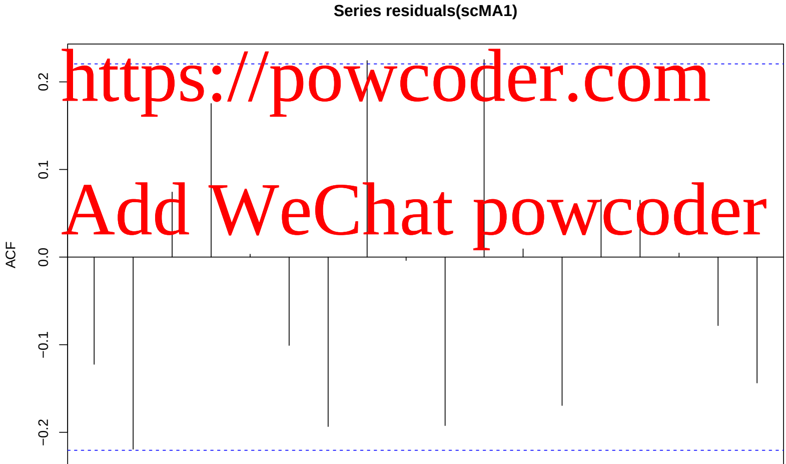


```
$expected.runs
[1] 38.21519


$n1
[1] 49
```

Residual ACF

- Under "white noise" assumption, the autocorrelations follow $N(0, \sigma^2 = 1/n)$
  - Plot them, if many are outside of $\pm 1.96/\sqrt(n)$ then independence appears to be violated



Ljung-Box Test

- In addition to examining lag k autocorrelation individually, we can assess their magnitude as a group
- For example, most of the ACFs may be moderate, but, taken as a group might look excessive
- Ljung-Box test developed for this scenario
  - H_0: ARMA(p, q) model is appropriate
  - H_a: ARMA(p, q) model is not appropriate

Ljung-Box Test


```
# tsdiag(scMA1) # runs test for multiple lags, want most p-
Box.test(residuals(scMA1), lag=10, type='Ljung-Box', fitdf=
```



```
	Box-Ljung test

data:  residuals(scMA1)
X-squared = 20.484, df = 9, p-value = 0.01515
```

Overfitting

- In addition to performing thorough residual analysis, overfitting can be a useful diagnostic technique to further assess the validity of an assumed model.

- Overfitting involves fitting a model more complicated than the one currenlty being considered and
  - examining the significance of the additional terms
  - examining the change in estimates from the assumed model

Overfitting

- When overfitting an ARMA(p, q) model, we consider the following two models

  1. ARMA($p+1$, q)
  2. ARMA(p, $q+1$)

Overfitting

Suppose my current model is an MA(1), or an ARMA(0, 1)

- Fit an ARMA(0, 2)
  - If additional MA term is insignificant → evidence that more complicated model is not needed
  - If additional MA term is significant → evidence that the MA(2) is worth considering
  - Residual analysis, see if things improved/worsened

- Fit an ARMA(1, 1)
  - Same as above . . .

- Can continue this process
  - add q, add p, diagnose until you reach a suitable model that fits well

Model Diagnostics with `sarima()`

- Diagnostic plots are provided in output
- Residual analysis as before
  1. Standardized residuals should look random (white noise)
  2. ACF of residuals should remain within boundary lines
  3. Normal Q-Q plot
  4. p values for Ljung-Box statistic
     - Tests for "randomness" of residuals
     - Most points should be above boundary line
     - If many are under it, start "overfitting" and hope for better results

Comparing Models

- Several models may look reasonable
  - Overfit supreme court data, you will find multiple ARIMA models that fit well

- Common to fit several and accept the model with the lowest AIC or BIC value
  - AIC - Akaike Information Criterion
  - BIC - Bayesian Information Criterion

Summary

- We now know how to
  - specify a model
    - ACF/PACF/EACF for p, d, q
  - diagnose how well it fits (are the residuals white noise)
    - Normality
    - Independence
  - Compare models
    - overfit and diagnose
    - compare AIC or BIC
- You are now well versed in modeling time series data in the ARIMA(p, d, q) family

Summary

- No model is perfect!
- Model specification and selection is not always clear cut
  - Lots of trial and error

- Not a "black box" exercise
- In the end, want a simple model that fits data well
- With this model, we focus on forecasting