Statistics 5350/7110
Forecasting

Lecture 18
Building ARIMA Models

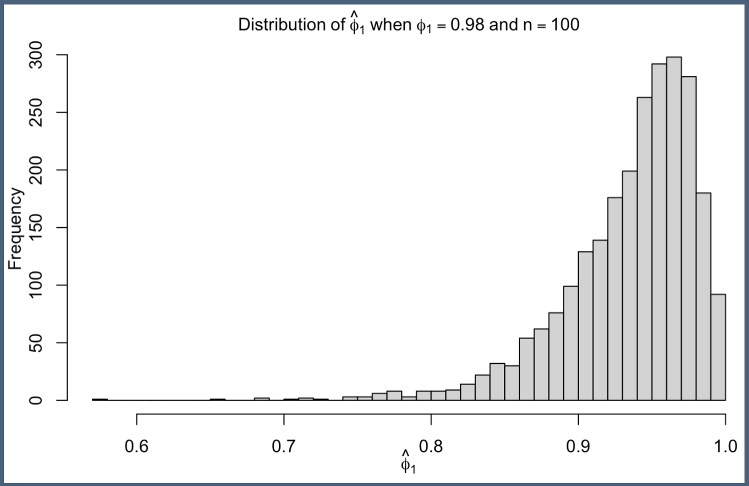
Professor Robert Stine

Preliminaries

- Questions?

- Assignments
  - Assignment questions
  - Dataset issues
  - Due Thursday

- Quick review
  - Forecasting ARIMA models
  - Forecasts diverge, uncertainty grows
  - RW vs AR(1) when ϕ approaches 1



Today's Topics

Text, §5.2

- Exponential smoothing
  - Popular "model-free" smoothing method for a time series
  - One-sided moving average with geometric weights
  - Equivalent to IMA(1,1)

- Building ARIMA models
  - Review diagnostic process
  - Interpreting residual ACF
  - Portmanteau test of residuals, Box-Pierce-Ljung test
  - Parameter drift, process changes

- Examples
  - US GDP time series

Exponential Smoothing

Exponential Smoothing

Example 5.5

- Geometric average
  - One-sided smoothing of data, using only values from past
  - Weights fall off geometrically (sound familiar?)
  - Divide by sum of weights so that we have a weighted average of the past values

as if we had the infinite history of the process

$$S_t = \frac{X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \cdots}{1 + \lambda + \lambda^2 + \cdots} = \frac{\sum_{h=0}^{\infty} \lambda^h X_{t-h}}{\sum_{h=0}^{\infty} \lambda^h} = (1-\lambda) \sum_{h=0}^{\infty} \lambda^h X_{t-h} \quad \text{for } 0 < \lambda < 1$$

  - Larger the value of λ, the smoother the sequence S_t
- Re-express as a recursion
  - Write down expression for adjacent values

$$\begin{aligned} S_t &= (1-\lambda)\left(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \cdots \right) \\ S_{t-1} &= (1-\lambda)\left(X_{t-1} + \lambda X_{t-2} + \lambda^2 X_{t-3} + \cdots \right) \end{aligned}$$

resembles "Gauss" derivation of the sum of a geometric series

  - Recognize that

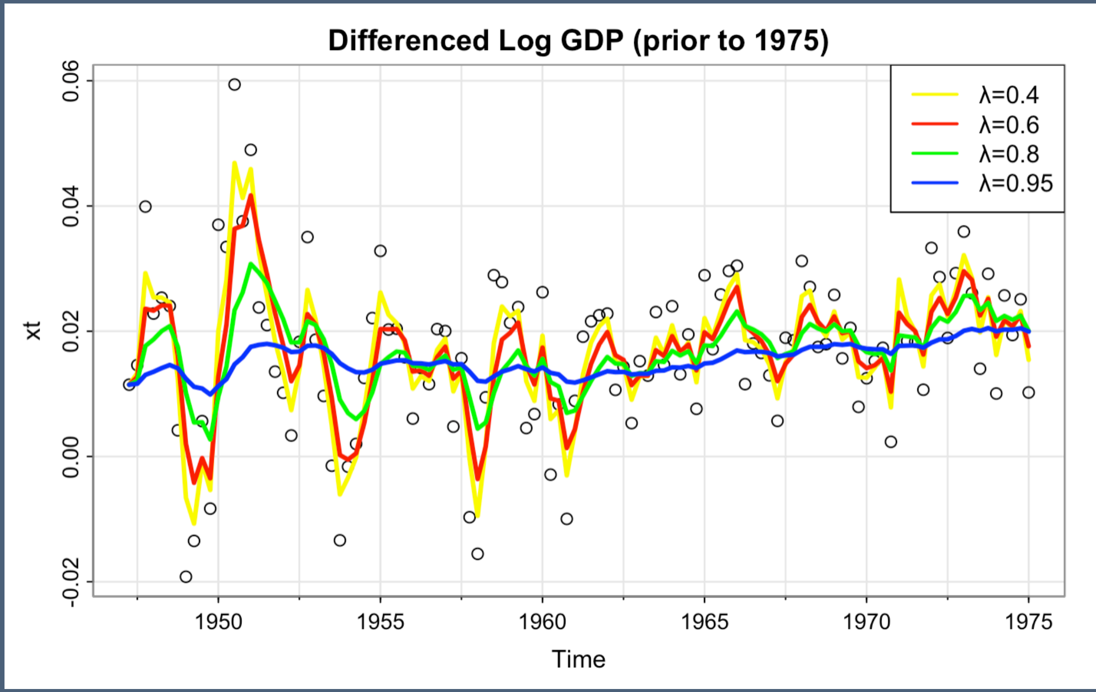
$$S_t - \lambda S_{t-1} = (1-\lambda) X_t \quad \Rightarrow \quad S_t = \lambda S_{t-1} + (1-\lambda) X_t$$

Recursive spreadsheet formula, provided you start with a reasonable choice for S_0.

Examples of Exponential Smoothing

- Exponential smooths of differenced quarterly log GDP
  - Several choices of the smoothing parameter

$$S_t = \lambda\, S_{t-1} + (1 - \lambda)\, X_t$$



Which is the best choice for the smoothing parameter?

Exponential Smoothing as ARIMA

Example 5.5

- Forecasts of an ARIMA(0,1,1) process
  - Write process as (change sign on MA term, $\lambda = -\theta$)

$$\nabla X_t = X_t - X_{t-1} = w_t - \lambda w_{t-1}$$

  and note that

$$w_t = \nabla X_t + \lambda w_{t-1}$$

  - Re-express the forecasts of the process

$$\begin{aligned} \widehat{X}_t &= X_{t-1} - \lambda w_{t-1} \\ &= X_{t-1} - \lambda (\nabla X_{t-1} + \lambda w_{t-2}) \\ &= X_{t-1} - \lambda X_{t-1} + \lambda \left(X_{t-2} - \lambda w_{t-2}\right) \\ &= (1-\lambda) X_{t-1} + \lambda \widehat{X}_{t-1} \end{aligned}$$

  which is an exponential smooth if we identify $\widehat{X}_t = S_t$ with a <u>time shift</u> (forecast vs smooth)

- Implications
  - Reminder that forecasts from ARIMA(0,1,1) model are geometrically weighted average
  - Suggests method for estimating the smoothing parameter

Comparing Smooth vs Forecast

- Example
  - Slice of nominal GDP data
  - Smoothing uses X_t to compute S_t

$$S_t = \lambda\, S_{t-1} + (1-\lambda)\, X_t$$

  - IMA interpretation treats S_t as a forecast of the next value

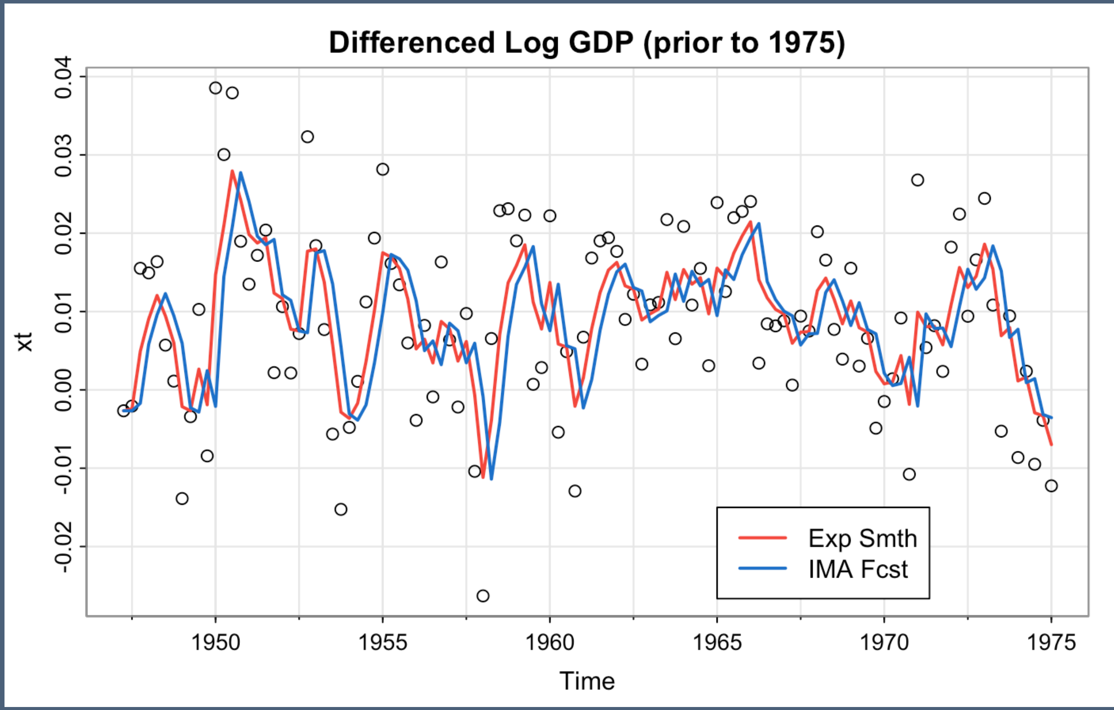
$$\hat{X}_t = \lambda\, \hat{X}_{t-1} + (1-\lambda)\, X_{t-1}$$

  - Estimate

$$\hat{\lambda} = -\,\hat{\theta} \approx 0.6$$

Coefficients:

	Estimate	SE	t.value	p.value
ma1	-0.5895	0.1777	-3.3173	0.0012
constant	-0.0001	0.0005	-0.1902	0.8495



Model Diagnostics

Illustrated with models for GDP

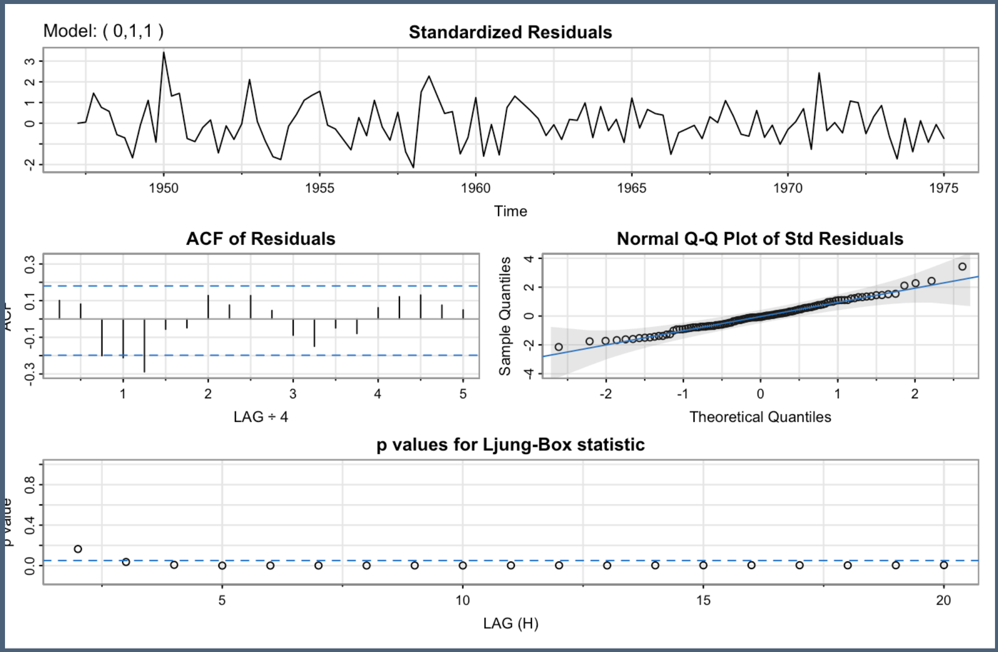
Diagnostics for the GDP Model

- Familiar time series diagnostics
  - Sequence plot of model residuals

    Standardizing mainly affects initial values (textbook p 107, eq 5.12)
  - ACF of residuals

    Caution: For most models, standard error is less than 1/√n
  - QQ plot of residuals

    Residuals tend to look more Gaussian than actual errors



- Portmanteau test
  - Accumulates squared residual autocorrelations
  - Plot shows p-values of tests with increasing amount of accumulation

Portmanteau Test

- Residual autocorrelation
  - Ideal model errors are white noise, uncorrelated with vaguely normal distribution
  - Estimated residual autocorrelations summarize dependence, but encounter multiplicity issues
  - Avoid looking individually by summing them…
  - Box-Pierce-Ljung statistic (equation 5.13)

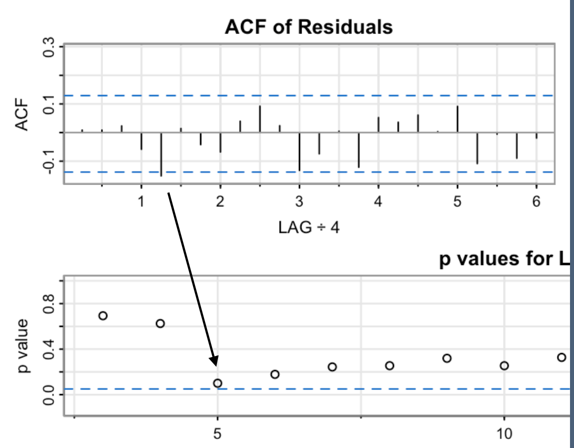
$$Q_H = n(n+2)\sum_{h=1}^{H}\frac{\hat{\rho}_e^2(h)}{n-h} \quad \Rightarrow \quad H << n \quad \Rightarrow \quad Q_H \approx n\sum_h \hat{\rho}_e^2(h)$$

denote residuals as $\{e_t\}$

  - Approximately chi-squared under null hypothesis of white noise

$$Q_H \sim \chi^2_{H-p-q}$$

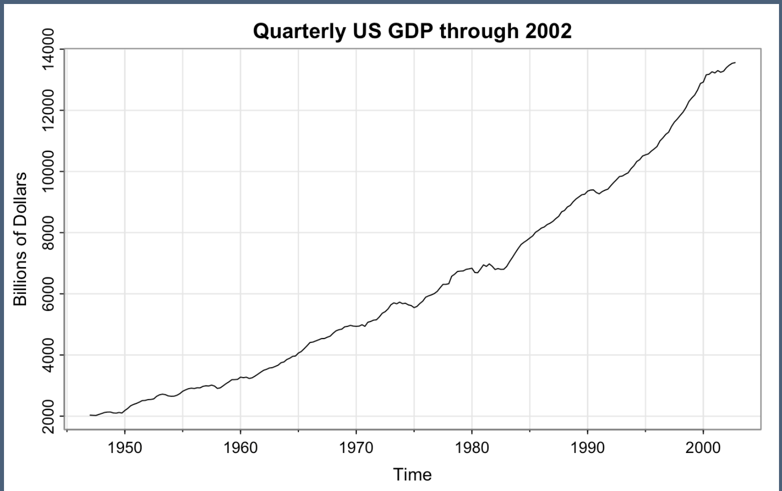
- Discussion
  - What to use for H?
    Textbook advice: "… not too large." (p 107)
  - Power of the test?



GDP Example: Differences

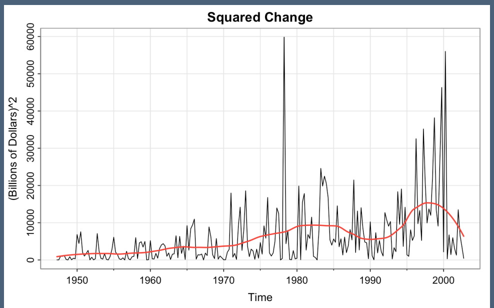
Textbook §5.2

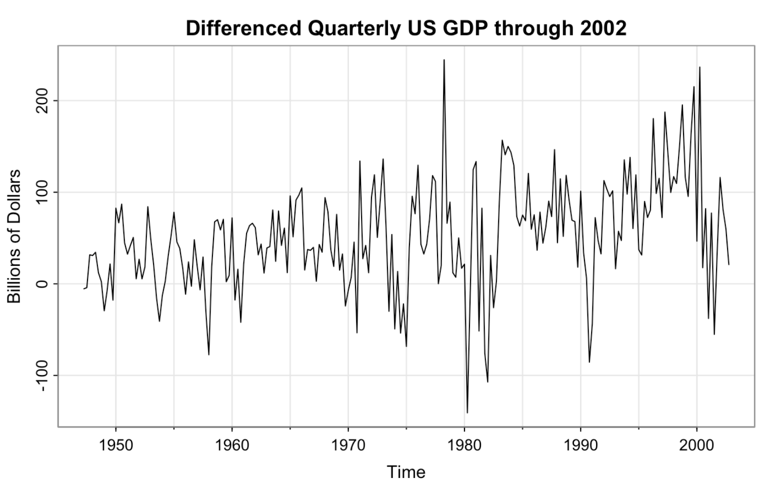
- Time series clearly non stationary
  - Use time period as in textbook, 1947-2002
  - Inflation adjusted, indexed to 1996 dollar



- Differenced data more stationary
  - Differencing resembles differentiation in calculus
  - Does the trend seem linear?
  - Are the variances stable?
  - Why do you care?

loess smooth

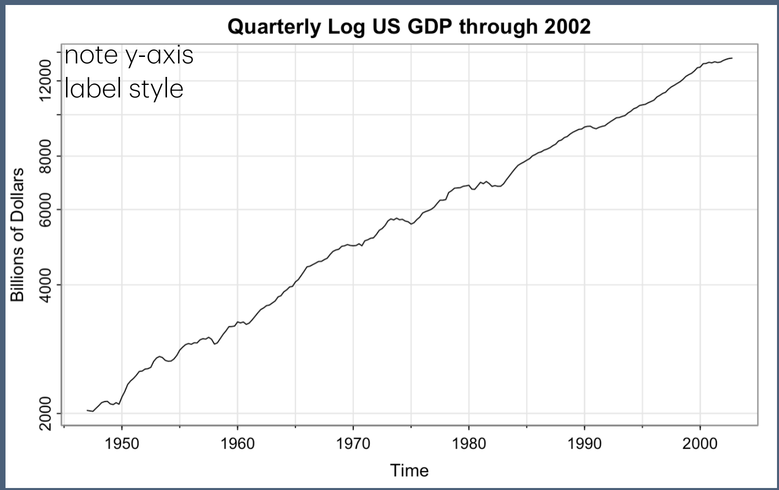




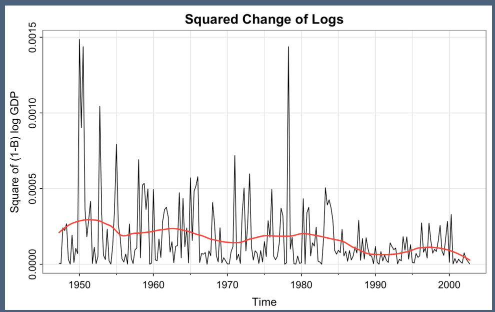
GDP Example: Log Differences

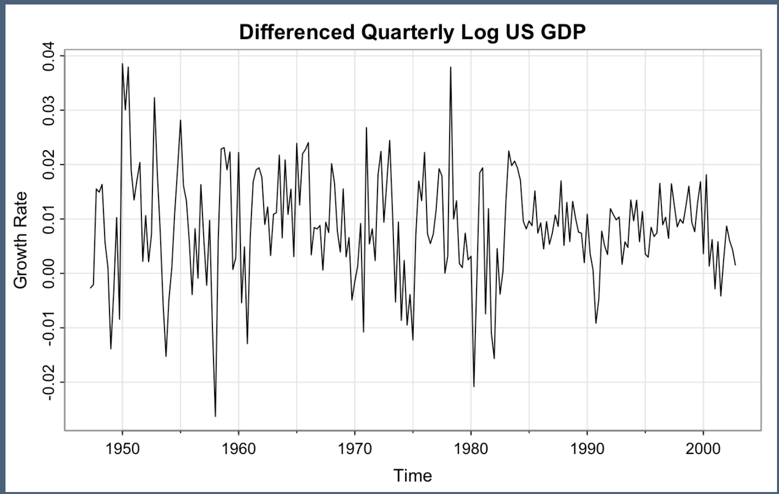
Textbook §5.2

- Switch to log scale
  - Trend is more linear in appearance



- Differenced log data
  - Interpreted as growth rate
  - Stationary? "Reverses" the pattern in differences.
  - More volatile in early period than later!

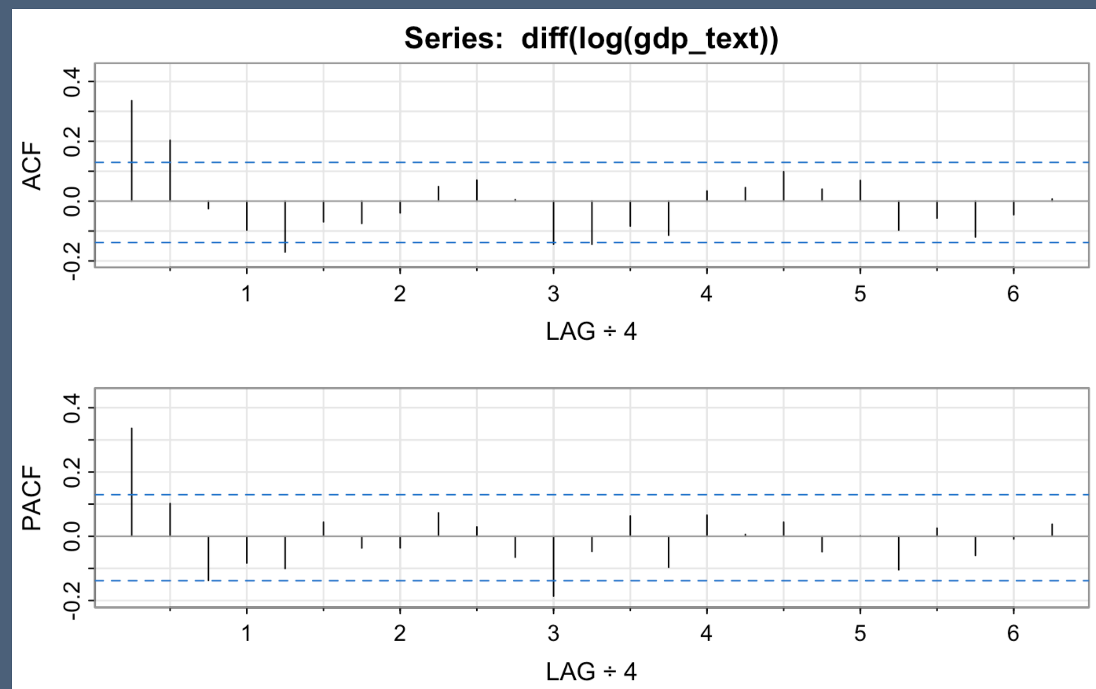




GDP Example: Model Selection

- Use differences of logs
  - Common because of the interpretation as a growth rate
  - I'll use here to match the textbook presentation
  - Heuristic
- What model to use?
  - ACF and PACF of differences of logs
- Model selection criteria
  - Textbook emphasizes ARIMA(0,1,2)
  - AIC likes an ARIMA(3,1,2)
  - BIC prefers ARIMA(1,1,0) [also in text]
  - None use the ARIMA(0,1,1) specification

BIC	q=0	q=1	q=2
p=0	21.39	6.97	0.48
p=1	0.00	4.17	5.16
p=2	3.00	5.95	8.61



GDP Example: ARIMA(0,1,2)

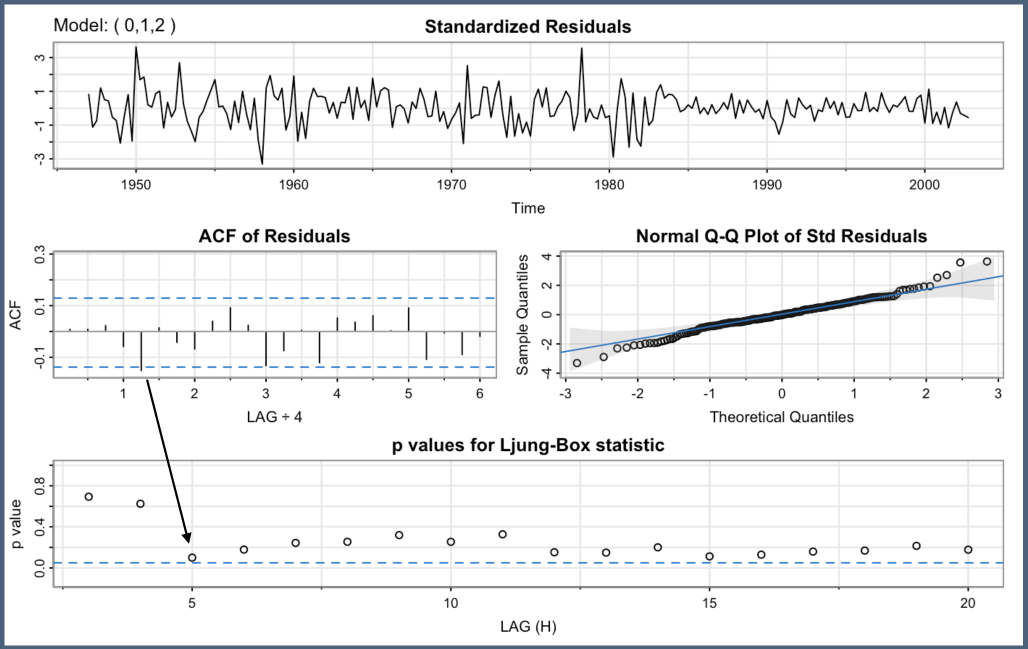
- Estimated model
  - Differences are MA(2)
  - Statistically significant parameter estimates
  - Non-zero drift term

```
Coefficients:
         Estimate     SE t.value p.value
ma1        0.2958 0.0653  4.5267   0e+00
ma2        0.2252 0.0625  3.6025   4e-04
constant   0.0085 0.0009  8.9360   0e+00

sigma^2 estimated as 8.635069e-05 on 220
```

Slightly different from results in text

- Diagnostics
  - Standardized residuals
  - Sequence plot shows decreasing variation
  - QQ plot shows slightly fat-tailed
  - Initial ACF very small
  - Portmanteau test of ACF shows p-values of test based on cumulative sum of squared residual ACF



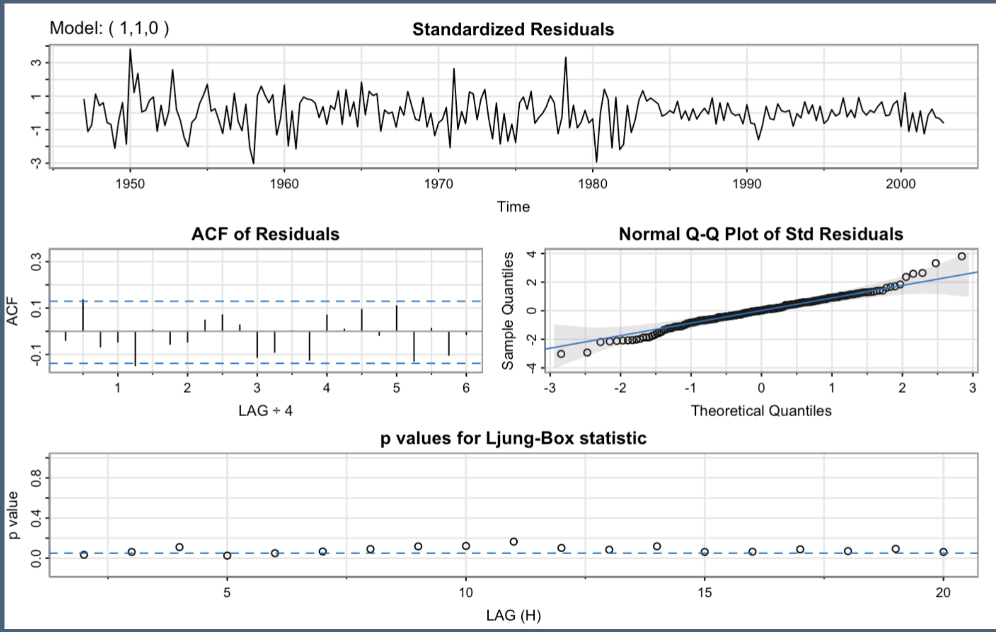
GDP Example: ARIMA(1,1,0)

```
Coefficients:
         Estimate     SE t.value p.value
ar1        0.3370 0.0631  5.3416       0
constant   0.0085 0.0009  8.9192       0

sigma^2 estimated as 8.829692e-05 on 221
```

```
sigma^2 estimated as 8.635069e-05
```

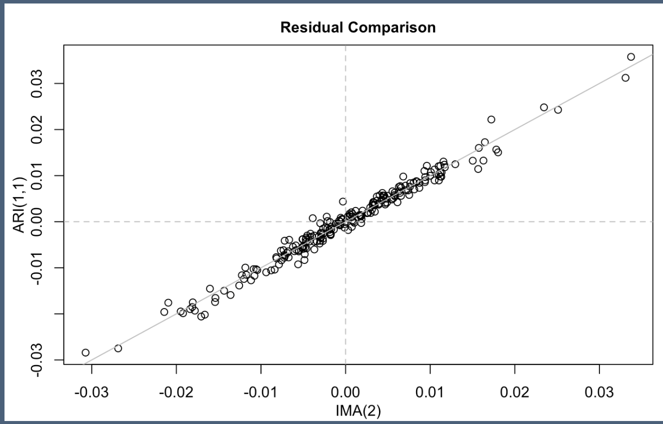
- Estimated model
  - Differences are AR(1)
  - Statistically significant estimates
  - Non-zero drift term
  - Larger estimated error variance



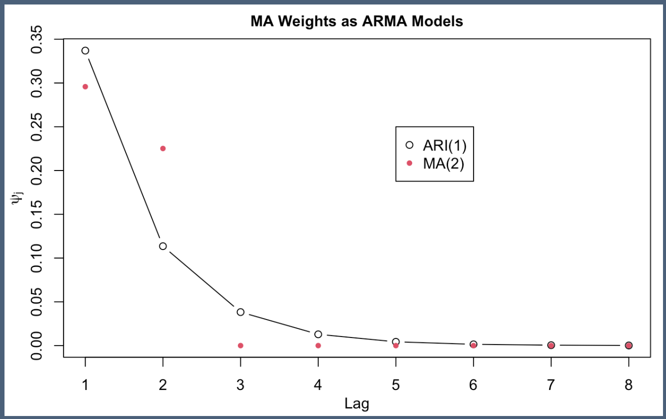
- Diagnostics
  - Sequence plot again shows decreasing variation
  - QQ plot close to diagonal (close to normal)
  - Initial ACF shows more autocorrelation, hence portmanteau test of ACF shows smaller p-values

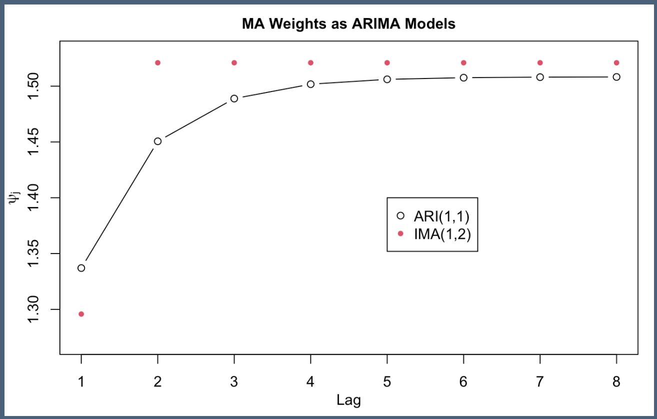
GDP Example: Model Comparison

- How similar are the fits?
  - Scatterplot of residuals from the two models
  - Hover on the X=Y diagonal
  - Thus, similar fits though not identical



- Comparison of MA weights

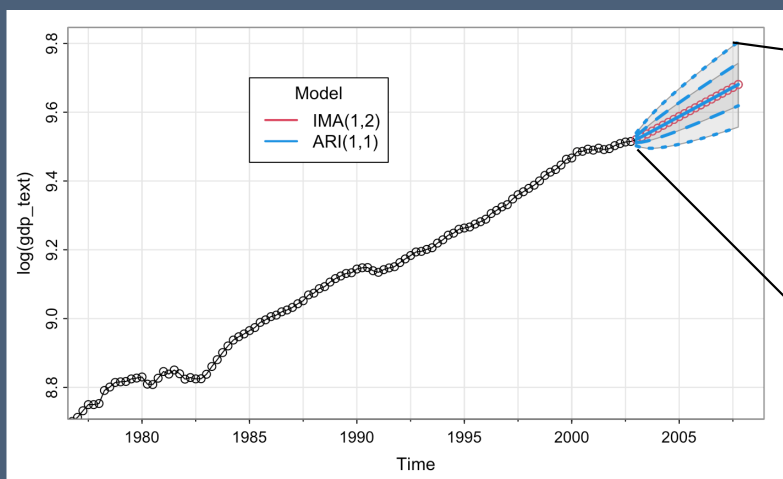


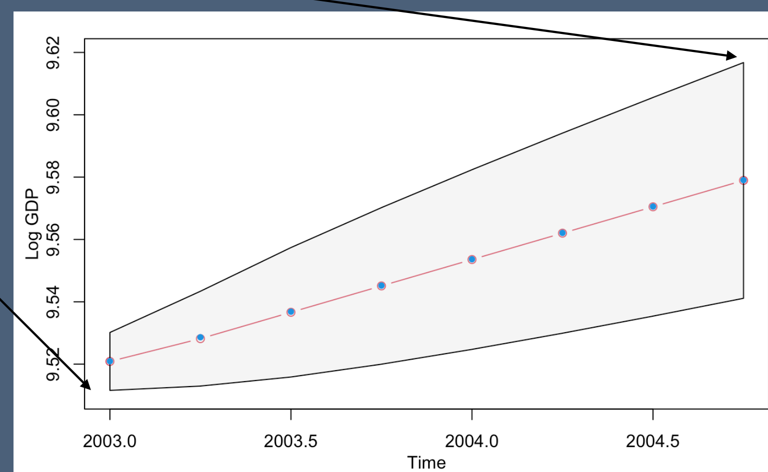


GDP Example: Forecast Comparison

- Forecasts of next 2 years
  - Remarkably similar forecasts and intervals

	2003	2003.25	2003.5	2003.75	2004	2004.25	2004.5	2004.75
IMA(1,2)	9.521	9.528	9.537	9.545	9.554	9.562	9.570	9.579
ARI(1,1)	9.521	9.529	9.537	9.545	9.554	9.562	9.571	9.579





Modeling Process

ARIMA Modeling Process

- Understand your data
  - Am I modeling the right time series?
  - Are there changes in measurement style, definitions?
  - Has the world changed during this time period? Has the population drifted?

Fit the same model in different time segments

- Plot the time series
  - Are there outliers? Are these evident coding errors or evidence of abrupt substantive changes?
  - Does the data suggest non-linear variation? (e.g. smooth trough and sharp peak)
  - Is there missing data? Why is it missing? Inactivity or a system failure? [not in our data!]
- Transform or difference the data as needed
  - Is the process stationary?
  - Does a log transformation make sense? (percentages)
- Identify the dependence in the time series
  - What are the patterns in the ACF/PACF? Slow or fast decay? Cut-off abruptly?
  - Use model selection criteria as suggestions

ARIMA Modeling Process Continued

- Specify and fit a model (pick p, d, and q)
  - Is fitted model stationary/invertible?
  - Are the reported estimated parameters statistically significant?
    Does the model contain redundant terms, factors that cancel in $\Phi(B)$ and $\Theta(B)$ (collinearity)
- Check the residuals $\{e_t\}$
  - Is the Q_H statistic significant? At what choice of of H?
  - Are the residuals roughly normally distributed in the QQ-plot?
  - Does the sequence plot of residuals look like white noise?
    Do you see changes in general properties of the residuals (e.g. change in mean or variance)
  - Do regression diagnostics reveal problems? (leverage, partial regression, …)
- Compare forecasts of multiple models as needed
  - How different are the forecasts from "comparable" models?
  - Do you have a baseline for comparison?
- Pick a model, or live with model averaging…

What's next?

- Modeling examples
  - Practice with simulated data (where we know the right answer)
  - Applying these ideas to real data

- Model testing
  - How do we decide when a model is good enough?