Overview of time series analysis steps

- We discussed how to:
  - Analyze the properties of a time series.
  - Identify candidate (ARIMA) model(s).
- Next:
  - Fit the model through (ML) estimation.
  - Check whether the model "fits well'.
  - Forecast future outcomes.

Estimating parameters of an ARMA model

- An ARMA(p, q), defined as:

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + e_t \\ &\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}, \end{aligned}$$

  with $e_t \sim WN(0, \sigma_e^2)$, has unknown parameters $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ and σ_e.
- We discussed two methods for estimating the parameters:
  - Method of moments (MoM)
  - Maximum likelihood estimation (MLE)
- By default, we'll use MLE (because MoM doesn't work well for fitting MA models).
- To use ML estimation, we do need to specify probability distributions.
- We will assume that white noise and the Y_t's are normally distributed unless otherwise specified.

ML estimation for time series

- The idea is simple: given $Y_1, \ldots, Y_t$, we want to find those parameters of the ARMA(p, q) model that maximize the likelihood function $f(y_1, \ldots, y_n)$.
- However, dealing with the Y_t's directly is complicated because of the autocorrelation: $f(y_1, \ldots, y_n) \neq \prod_{i=1}^{n} f(y_i)$.
- We discussed the main ideas for the AR(1) model with mean $E(Y_t) = \mu$.

ML estimation for the AR(1)-model

- We first derived that:

$$f(y_t, \ldots, y_1) = f(y_t|y_{t-1}, \ldots, y_1) \cdots f(y_2|y_1) f(y_1),$$

$$f(y_{t+1}|y_t, y_{t-1}, \ldots, y_1) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2\sigma_e^2}(y_{t+1} - \mu - \phi(y_t - \mu))^2\right)$$

$$f(y_1) = \frac{1}{\sqrt{2\pi\sigma_e^2/(1-\phi^2)}} \exp\left(\frac{-1}{2\sigma_e^2/(1-\phi^2)}(y_1 - \mu)^2\right),$$

to then find

$$\begin{aligned} L(\phi, \mu, \sigma_e^2) &= \prod_{t=2}^{n} f(y_t|y_{t-1}, y_{t-2}, \ldots, y_1) f(y_1), \\ &= (2\pi\sigma_e^2)^{-n/2}(1-\phi^2)^{1/2} \exp\left(-\frac{1}{2\sigma_e^2} S(\phi, \mu)\right), \text{ with} \\ S(\phi, \mu) &= (1-\phi^2)(Y_1 - \mu)^2 + \sum_{t=2}^{n}((Y_t - \mu) - \phi(Y_t - \mu))^2. \end{aligned}$$

Large-sample properties of the estimators

- Important question: if we fit an ARMA(p,q) model to data from an ARMA(p,q) process, how close do we expect the estimates to be to the true parameters?
- For ML estimators, standard theory gives the large sample distribution for the estimators.
- For large n, the estimators are approximately unbiased, normally distributed and we can derive the expression for the sampling variance.
  - E.g. for the ML estimate $\hat{\phi}$ for ϕ in an AR(1) model:
$$\hat{\phi} \sim N\left(\phi, \frac{1-\phi^2}{n}\right), \text{ approximately.}$$
  - How do we use this information in practice? An approximate 95% confidence interval for ϕ is given by $\hat{\phi} \pm 1.96 \cdot s\{\hat{\phi}\}$, where the standard error (SE) for $\hat{\phi}$, $s\{\hat{\phi}\}$ is given by $\sqrt{\frac{1-\hat{\phi}^2}{n}}$.
- MLE estimates for ARIMA(p,d,q) model parameters based on data "data" can be obtained in R as follows:

```
arima(data, order = c(p,d,q), method = "ML")
```

Examples for data series 2

```
> arima(data2, order=c(2,1,0),method='ML')

Call:
arima(x = data2, order = c(2, 1, 0), method = "ML")

Coefficients:
          ar1     ar2
      -0.4287  0.3544
s.e.   0.0538  0.0540

sigma^2 estimated as 0.9545:  log likelihood = -419.12,  aic
```

- Conclusion: when fitting an ARIMA(2,1,0) model to the Y_t's with ML estimation, we find that $\hat{\phi}_1 = -0.43$ with approximate SE $s\{\hat{\phi}_1\} = 0.0538$, and $\hat{\phi}_2 = 0.35$ with approximate SE $s\{\hat{\phi}_2\} = 0.0540$.
- An approximate 95% CI for ϕ_2 is given by $\hat{\phi}_2 \pm 1.95 \cdot s\{\hat{\phi}_2\}$.

Example for data series 3

- Note: as before, the AR and MA coefficients are given below, with their SEs.
- IMPORTANT (again!): In R, the θ's are always reported with opposite sign! E.g. in R, think about an MA process as

$$Y_t = e_t + \theta_1^* e_{t-1} + \theta_2^* e_{t-2} + \ldots + \theta_q^* e_{t-q},$$

  where $\theta_k^* = -\theta_k$ in the MA notation we use in the class/the book.

```
> arima(data3, order=c(3,0,5),method='ML')

Call:
arima(x = data3, order = c(3, 0, 5), method = "ML")

Coefficients:
          ar1      ar2     ar3      ma1     ma2     ma3
      -0.8195  -0.1226  0.5805  -0.7849  0.6084  0.6043  -0.2
s.e.   0.0149   0.0212  0.0148   0.0155  0.0123  0.0123   0.0
```

Model diagnostics

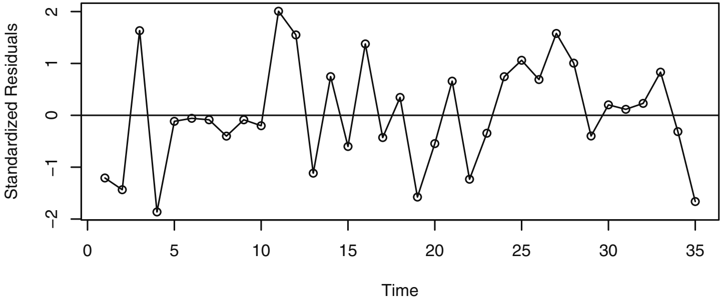
- How to check if the model "fits the data well"?
- Use residual analysis!
- Residual $\hat{e}_t$ = actual Y_t - predicted Y_t by the model, e.g. for an AR(p) model:

  $$\hat{e}_t = Y_t - \hat{\theta}_0 - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} - \ldots - \hat{\phi}_p Y_{t-p}$$

  (with some more details to obtain the first residuals).
- Often, standardized residuals are used with common variance 1: $\hat{s}_t = \hat{e}_t / \sqrt{\widehat{Var}(\hat{e}_t)}$.
- If the model was correctly specified, and the parameter estimates are reasonably close to the true values, then the residuals $\hat{e}_t$ should have *nearly* the properties of normally distributed white noise e_t.
- Things to check: Zero mean; Constant variance; Normality; Outliers; Autocorrelation.
- How?

Example of residual plot: color time series (tut)

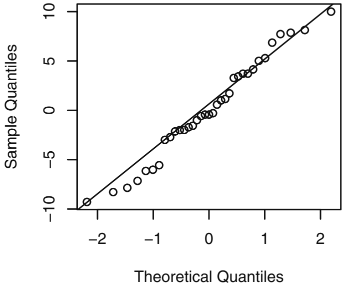
Exhibit 8.1 Standardized Residuals from AR(1) Model of Color



Time series plot of residual can be used to visually check lack of trends, constant variance and outliers (with critical value $z_{1-\alpha/2\cdot 1/n}$).

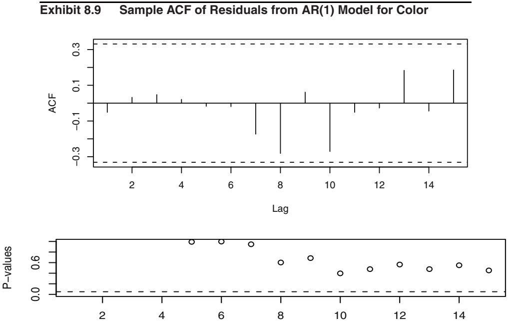
Checking normality for color data

Exhibit 8.4 Quantile-Quantile Plot: Residuals from AR(1) Color Model



- QQ-plot: If the residuals are approximately normally distributed, we expect that the points lie on a straight line.
- Test: Shapiro-Wilk normality test, with $H_0 =$ "Sample is drawn from a normal distribution".
  - P-value for Shapiro-Wilk normality test for residuals from color data around 0.6: We don't have evidence to reject the normality assumption.

Example of checking autocorrelation: Color data



- Check sample ACF (critical values based $\text{Var}(r_k) \approx 1/n$).
  - Approximate variance for r_k for residuals is generally smaller than $1/n$ for lower k and r_k's are autocorrelated, so make sure there is no significant autocorrelation up to lag K with LB test.
- Ljung-Box test with H_0: Autocorrelations up to and including lag K are zero.
  - P-values in plot are those for the LB test (no problems detected).

Forecasting

- Given $Y_1, Y_2, \ldots, Y_t$, forecast Y_{t+g}:
  - t is called the forecast origin,
  - g is called the lead time (referred to as l in the book, I use g to improve readability).
- The minimum mean square error forecast for Y_{t+g} given $Y_1, \ldots, Y_t$, is given by

$$\hat{Y}_t(g) = E(Y_{t+g}|Y_1, \ldots, Y_t),$$

  this forecast $\hat{Y}_t(g)$ is the function $h(Y_1, Y_2, \ldots, Y_t)$ which minimizes:

$$E\left[(Y_{t+g} - h(Y_1, Y_2, \ldots, Y_t))^2\right].$$

- How to obtain the forecast for ARIMA models?

Forecasting ARIMA models

- To find $\hat{Y}_t(g) = E(Y_{t+g}|Y_1, \ldots, Y_t)$, start by plugging in the expression for Y_{t+g}, as given by the ARIMA model, in the conditional expectation.
- You'll end up with a combination of $E(Y_{t+j}|Y_1, \ldots, Y_t)$'s and $E(e_{t+j}|Y_1, \ldots, Y_t)$'s.
- Then use that

$$E(Y_{t+j}|Y_1, \ldots, Y_t) = \begin{cases} Y_{t+j} & \text{for } j \leq 0, \\ \hat{Y}_t(j) \text{ (true forecast)} & \text{for } j > 0, \end{cases}$$

and

$$E(e_{t+j}|Y_1, \ldots, Y_t) = \begin{cases} 0 & \text{for } j > 0, \\ e_{t+j} & \text{for } j \leq 0. \end{cases}$$

Forecasting: example

Suppose Y_t is given by

$$Y_t = 1 + e_t - 0.4e_{t-1} + 0.1e_{t-2}$$

with $\sigma_e^2 = 1$ and the most recent Y_t's and e_t's as displayed below:

t	95	96	97	98	99	100
Y_t	-0.30	2.40	1.50	2.80	0.70	0.60
e_t	-1.10	0.90	1.00	2.10	0.40	-0.50

What is the 95% PI for Y_{102}?

- Start with point forecast $\hat{Y}_{100}(2)$.

Forecasting: example

$$Y_t = 1 + e_t - 0.4e_{t-1} + 0.1e_{t-2}$$

with $\sigma_e^2 = 1$ and the most recent Y_t's and e_t's as displayed below:

t	95	96	97	98	99	100
Y_t	-0.30	2.40	1.50	2.80	0.70	0.60
e_t	-1.10	0.90	1.00	2.10	0.40	-0.50

Start with point forecast $\hat{Y}_{100}(2)$, with $t = 100$:

$$\begin{aligned}
\hat{Y}_t(2) &= E(Y_{t+2}|Y_t, Y_{t-1}, \ldots, Y_1), \\
&= E(1 + e_{t+2} - 0.4e_{t+1} + 0.1e_t|Y_t, Y_{t-1}, \ldots, Y_1), \\
&= 1 + E(e_{t+2}|Y_t, Y_{t-1}, \ldots, Y_1) - 0.4E(e_{t+1}|Y_t, Y_{t-1}, \ldots, Y_1) \\
&\quad +0.1E(e_t|Y_t, Y_{t-1}, \ldots, Y_1), \\
&= 1 + 0.1e_t \\
&= 1 + 0.1 \cdot (-0.5) = 0.95.
\end{aligned}$$

How to get prediction intervals?

- For an invertible ARIMA model:
  $e_t(g) = Y_{t+g} - \hat{Y}_t(g) = e_{t+g} + \psi_1 e_{t+g-1} + \ldots + \psi_{g-1} e_{t+1}$,
  where the coefficients follow from the representation
  $Y_t - \mu = \sum_{j=0}^{\infty} \psi_j e_{t-j}$.
- Then $e_t(g) \sim N(0, Var(e_t(g)))$, where $Var(e_t(g)) = \sigma_e^2 \sum_{j=0}^{g-1} \psi_j^2$.
- We can construct prediction intervals (PI) for the future observation Y_{t+g}, using $e_t(g) = Y_{t+g} - \hat{Y}_t(g) \sim N(0, Var(e_t(g)))$ thus

$$P\left(-z_{1-\alpha/2} \leq \frac{Y_{t+g} - \hat{Y}_t(g)}{\sqrt{Var(e_t(g))}} \leq z_{1-\alpha/2}\right) = 1 - \alpha$$

- E.g. the 95% PI for Y_{t+g} is given by $\hat{Y}_t(g) \pm 1.96\sqrt{\widehat{Var}(e_t(g))}$ (ignoring additional uncertainty that follows from estimating model parameters).

Forecasting: example

Suppose Y_t is given by

$$Y_t = 1 + e_t - 0.4e_{t-1} + 0.1e_{t-2}$$

with $\sigma_e^2 = 1$. What is the 95% PI for Y_{102}?

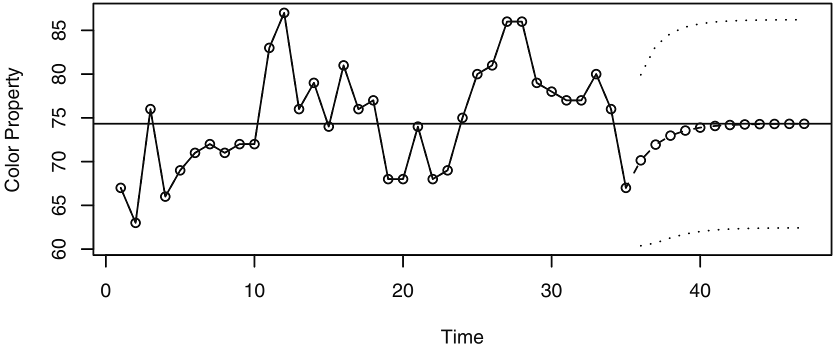
- The point forecast $\hat{Y}_{100}(2) = 0.95$.
- $Var(e_t(g)) = \sigma_e^2 \sum_{j=0}^{g-1} \psi_j^2$, so we need to find the ψ_j's but this turns out to be easy for an MA(q) process!
- $Y_t - 1 = e_t - 0.4e_{t-1} + 0.1e_{t-2} = \sum_{j=0}^{\infty} \psi_j e_{t-j}$, thus $\psi_0 = 1$, $\psi_k = -\theta_k$ for $k = 1, 2$ and $\psi_j = 0$ otherwise.
- Thus

$$\begin{aligned} Var(e_t(2)) &= (1 + \psi_1^2)\sigma_e^2, \\ &= (1 + \theta_1^2)\sigma_e^2 = (1 + 0.4^2) \cdot 1 = 1.16. \end{aligned}$$

The 95% PI is given by $\hat{Y}_t(2) \pm 1.96 sd(e_t(g))$, here (-1.2, 3.1).

Example of AR(1) projection with uncertainty

Exhibit 9.3 Forecasts and Forecast Limits for the AR(1) Model for Color



Some interesting forecast properties for a stationary ARMA processes

- For point forecasts:
  - For $g = 1, \ldots, q$ the point prediction is determined by AR and MA terms (past Y_t's and past white noise),
  - For $g = q+1, q+2, \ldots$ the point prediction is determined by AR terms (past Y_t's) but their influence decreases over time as the forecast gets closer to μ.
  - $Y_t(g) \to \mu$ as $g \to \infty$ (easy to show for AR(1), discussed for general ARMA processes).
- For the variance of the forecast errors:

$$Var(e_t(g)) \to \gamma_0 \text{ as } g \to \infty.$$

Overview of time series analysis steps

- We discussed in Ch 1 to 9 how to:
  - Analyze the properties of a time series.
  - Identify candidate (ARIMA) model(s).
  - Fit the model through (ML) estimation.
  - Check whether the model “fits well’.
  - Forecast future outcomes.
- Outlook:
  - How to choose between models and how to check systematically for non-stationarity?
  - How to include covariates?
  - ...