Economics 144
Economic Forecasting

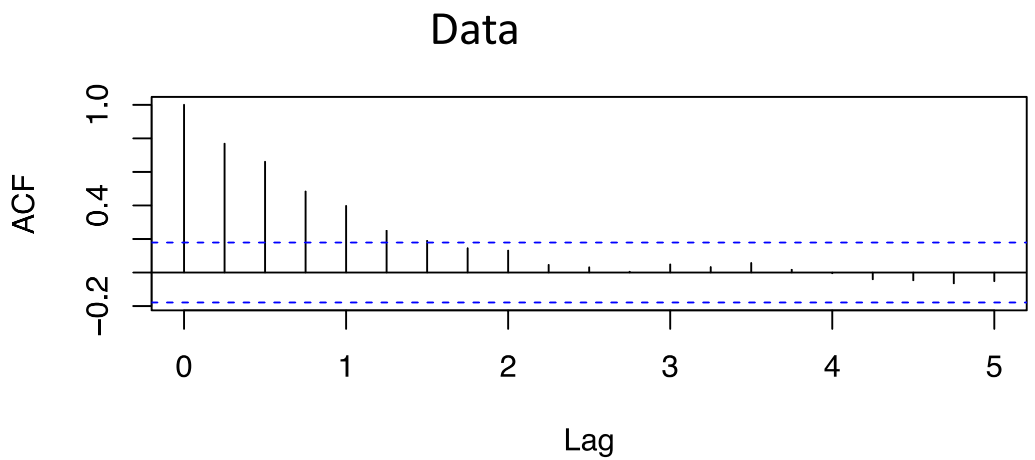
Lecture 6
Characterizing Cycles
Moving Average Models

Dr. Randall R. Rojas

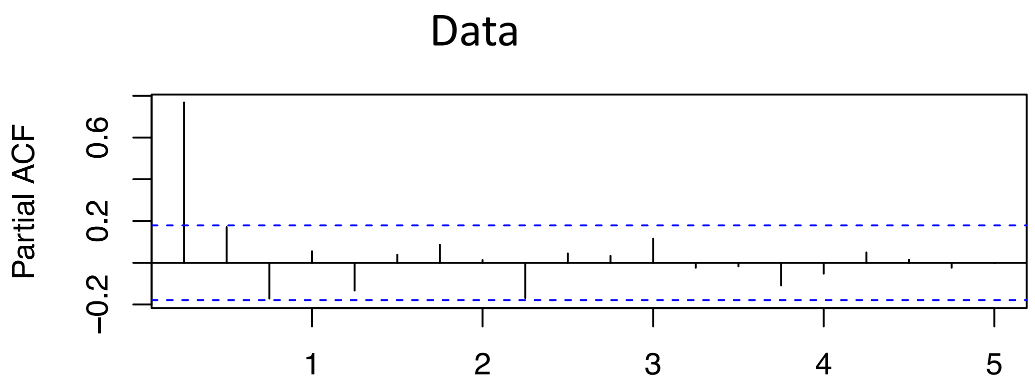
Today's Class

- Covariance Stationary Time Series
- White Noise
- The Lag Operator
- Wold's Theorem
- Characteristics of the MA(q) Process
  – Example: MA(1) Process

Covariance Stationary Time Series



In R: `acf(data)`



In R: `pacf(data)`

Recall that covariance stationary processes have $\rho(k)$ and $p(k)$ that →0 as k →infinity.

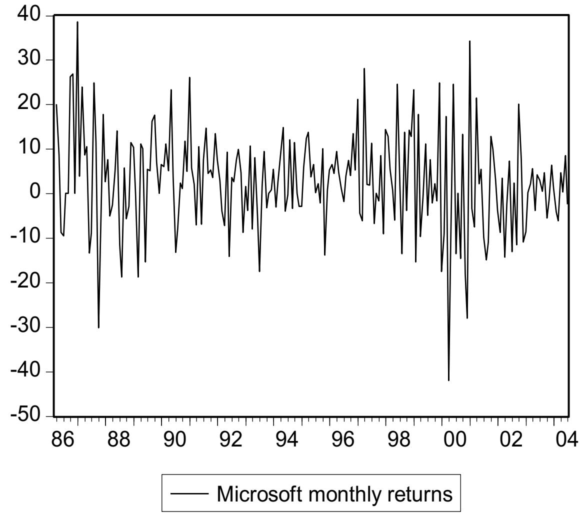
White Noise 1 of 3

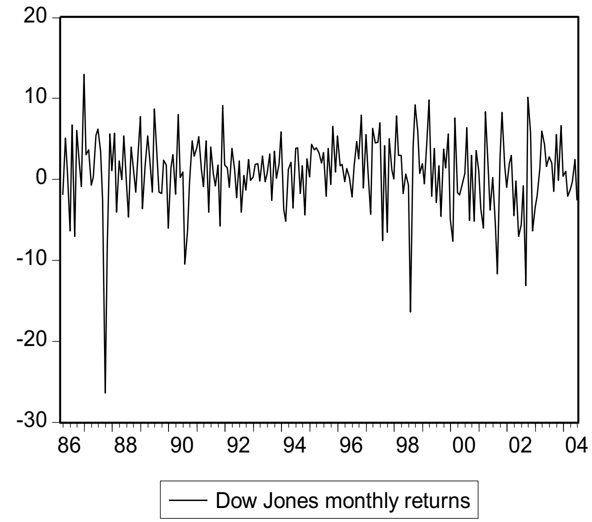
- White Noise: Time series process with zero mean, constant variance, and no serial correlation.

- Since $\rho(k) = 0$ and $p(k) = 0$ for $k \geq 1$, there is no link between past and present observations.
  → Cannot predict the future.

- Examples: S&P500 returns, interest rates, etc.

Example: Autocorrelation Functions of Monthly Returns for Microsoft and the Dow Jones Index





Sample: 1986:03 2004:07
Included observations: 220

Autocorrelation	Partial Correlation		AC	PAC
		1	-0.081	-0.081
		2	-0.094	-0.101
		3	0.132	0.117
		4	-0.017	-0.006
		5	0.008	0.030
		6	-0.013	-0.029
		7	0.106	0.113
		8	0.015	0.023
		9	-0.006	0.024
		10	0.131	0.112
		11	0.013	0.035
		12	-0.016	0.005
		13	-0.020	-0.045
		14	0.030	0.013
		15	-0.075	-0.091
		16	0.064	0.068
		17	0.085	0.051
		18	-0.094	-0.064
		19	-0.049	-0.079
		20	0.046	0.005

Sample: 1986:03 2004:07
Included observations: 220

Autocorrelation	Partial Correlation		AC	PAC
		1	-0.021	-0.021
		2	-0.044	-0.044
		3	-0.056	-0.058
		4	-0.126	-0.131
		5	0.048	0.037
		6	-0.031	-0.045
		7	0.092	0.081
		8	-0.044	-0.057
		9	-0.043	-0.030
		10	0.043	0.035
		11	-0.012	0.006
		12	0.015	-0.006
		13	-0.003	0.003
		14	-0.034	-0.034
		15	-0.059	-0.060
		16	0.039	0.042
		17	0.031	0.012
		18	0.079	0.076
		19	-0.026	-0.028
		20	-0.016	0.005

Wold's Theorem (Part I)

- Q: What's left after filtering the trend and seasonal components?
- A: Covariance stationary (short memory) residuals! How should we model them?→ Wold's Theorem!

- Wold's Representation Theorem: Let $\{y_t\}$ be any zero-mean covariance-stationary process. Then

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i} \text{ and } \varepsilon_t \sim WN(0, \sigma^2)$$

(Innovations)

where $\sum_{i=0}^{\infty} b_i^2 < \infty$ and b_o=1.

Moving Average Models

- Def: MA(q) = Moving Average process or order q≥0: $Y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$, where $\varepsilon_t \sim (0, \sigma^2)$

- Examples:
  - MA(1): $Y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t$
  - MA(5): $Y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_5 \varepsilon_{t-5} + \varepsilon_t$
  - MA(10): $Y_t = \mu + \theta_{10} \varepsilon_{t-10} + \varepsilon_t$

Moving Average Models

For every MA(q) process, we need to address the following 3 questions:

1. What does a time series of an MA process look like?
2. What do the corresponding ACFs and PACFs look like?
3. What is the optimal forecast?

Moving Average Models
Example: MA(1) Process

- (1) What does a time series of an MA process look like?

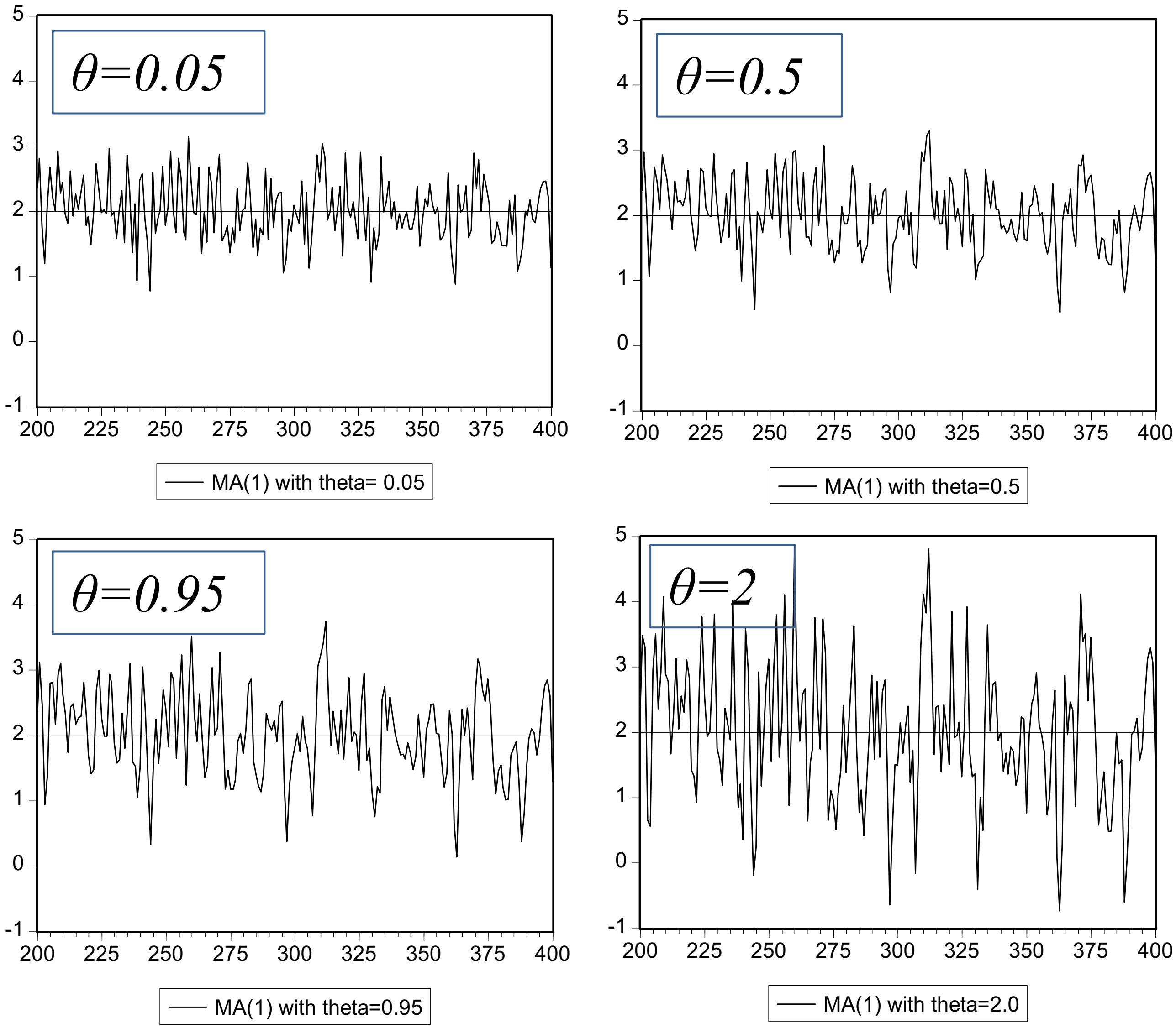
- Consider the following MA(1) process:

  $Y_t = \mu + \theta\, \varepsilon_{t-1} + \varepsilon_t$

- We can plot this process for $\mu = 2$, and different values of θ, e.g., for $\theta=0.05$, $\theta=0.5$, $\theta=0.95$, and $\theta=2$.

- We can show that $E(Y_t) = \mu$, and $\sigma^2(Y_t) = (1+\theta^2)\sigma^2_{\varepsilon}$

MA(1): $Y_t = \mu + \theta\, \varepsilon_{t-1} + \varepsilon_t$



Moving Average Models
Example: MA(1) Process

- (2) What do the corresponding ACFs and PACFs look like?

- ACF:
  - (i) We would expect to see only 1 spike different from zero, i.e., $\rho_1 \neq 0$, and all others equal to zero ($\rho_k = 0$, $k > 1$).
  - (ii) The magnitude of the spike should be proportional to θ for $|\theta| < 1$.
  - Note: You can show (please fill in the steps) that
    $\rho_1 = \theta / (1 + \theta^2)$
  - (iii) Given the expression above, the sign of the ACF is the same as the sign of θ.

Moving Average Models
Example: MA(1) Process

- (2) What do the corresponding ACFs and PACFs look like?

- PACF:
  - Note: You can show (please fill in the steps) that the autocovariance of order 1 is given by: $\gamma_1 = \theta\sigma^2_\varepsilon$, and all other orders are equal to 0.
  - (i) The PACF decreases to zero in an alternating fashion, according to $p_k > 0$ (k=odd), and $p_k < 0$ (k =even).
  - (ii) We can also show that $p_1 = \rho_1$.

Autocorrelation Functions of Simulated MA(1) Processes

Sample: 1 2000
Included observations: 1999

$\theta = 0.05$

Autocorrelation	Partial Correlation		AC	PAC
		1	0.072	0.072
		2	-0.027	-0.032
		3	-0.061	-0.057
		4	-0.024	-0.017
		5	0.016	0.016
		6	-0.004	-0.011
		7	0.022	0.022
		8	0.003	0.001
		9	0.000	0.000
		10	-0.011	-0.009

Sample: 1 2000
Included observations: 1999

$\theta = 0.5$

Autocorrelation	Partial Correlation		AC	PAC
		1	0.406	0.406
		2	-0.038	-0.243
		3	-0.076	0.053
		4	-0.038	-0.039
		5	0.006	0.029
		6	0.009	-0.015
		7	0.020	0.030
		8	0.011	-0.012
		9	-0.003	0.002
		10	-0.010	-0.009

Sample: 1 2000
Included observations: 1999

$\theta = 0.95$

Autocorrelation	Partial Correlation		AC	PAC
		1	0.498	0.498
		2	-0.042	-0.386
		3	-0.081	0.220
		4	-0.042	-0.183
		5	0.003	0.157
		6	0.013	-0.127
		7	0.019	0.131
		8	0.013	-0.112
		9	-0.004	0.092
		10	-0.010	-0.088

Sample: 1 2000
Included observations: 1999

$\theta = 2$

Autocorrelation	Partial Correlation		AC	PAC
		1	0.406	0.406
		2	-0.039	-0.243
		3	-0.076	0.053
		4	-0.038	-0.039
		5	0.006	0.030
		6	0.009	-0.016
		7	0.020	0.030
		8	0.010	-0.013
		9	-0.004	0.002
		10	-0.010	-0.008

Can you distinguish them?

Moving Average Models
Example: MA(1) Process

- From our previous example, based on the ACFs alone, we could not distinguish the MA(1) process with $\theta=0.5$ from the MA(1) process with $\theta=2$. This property is known as invertibility.
- Def: Invertibility

  An MA(1) process is invertible if $|\theta| < 1$. Otherwise, if $|\theta| \geq 1$, the process is noninvertible.

Moving Average Models
Example: MA(1) Process

- Meaning of Invertibility: You can transform the MA(1) process to an ‘autoregressive’ function of its own past (lagged values), such that the recent past has more weight than the distant past.
- Example: $Y_t = \mu + \theta\varepsilon_{t-1} + \varepsilon_t = \mu + (1\text{-}(\text{-}\theta)\mathrm{L})\, \varepsilon_t$
- Solve for ε_t → $\boxed{\varepsilon_t = \dfrac{1}{1-(-\theta)L}(Y_t - \mu)}$

Moving Average Models
Example: MA(1) Process

- Since $|\theta|<1$, we can perform a Taylor series expansion of the denominator:

$$\frac{1}{1-(-\theta)L} = \lim_{i\to\infty}\left(1+(-\theta)L+(-\theta)^2L^2+\cdots+(-\theta)^iL^i+\cdots\right)$$
$$= \left(1-\theta L+\theta^2L^2+\cdots\right)$$

- Therefore, we can re-express $\varepsilon_t = \frac{1}{1-(-\theta)L}(Y_t-\mu)$ as: $\varepsilon_t = \left(1-\theta L+\theta^2L^2+\cdots\right)(Y_t-\mu)$

$$\rightarrow Y_t-\mu = \theta(Y_{t-1}-\mu)-\theta^2(Y_{t-2}-\mu)+\cdots+\varepsilon_t$$

= Autoregressive Process

Moving Average Models
Example: MA(1) Process

- If an MA process is invertible, you can always find an Autoregressive representation.

- To predict the future, we need the information contained in the past.

- As noted earlier, the recent past has more weight than the distant past since: $\lim_{t \to \infty} \theta^t \to 0$

Moving Average Models
Example: MA(1) Process

- Q: What can we conclude for the case when $|\theta| \geq 1$?
- A: We cannot perform the Taylor series expansion on θ but we can on 1/θ.
- Since θ>1, consider expanding 1/(1-θL) as follows:

$$\frac{1}{1-\theta L} = -\frac{\frac{1}{\theta L}}{1-\frac{1}{\theta L}} = -\frac{1}{\theta L}\underbrace{\left(1+\frac{1}{\theta L}+\left(\frac{1}{\theta L}\right)^2+\cdots\right)}_{\text{Infinite lag polynomial}}$$

Moving Average Models
Example: MA(1) Process

- Def: Forward Operator (F) = 1/L, where

  $FY_t = Y_{t+1} \rightarrow F(L(Y_t)) = F(Y_{t-1}) = Y_t$. The forward operator is inverse of the lag operator. It delivers the process at a future date.

$$\frac{1}{1-\theta L} = -\frac{1}{\theta}F\left(1+\frac{1}{\theta}F+\frac{1}{\theta^2}F^2+\cdots\right)$$

$$\varepsilon_t = \frac{1}{1-\theta L}Y_t = -\frac{1}{\theta}F\left(1+\frac{1}{\theta}F+\frac{1}{\theta^2}F^2+\cdots\right)Y_t$$

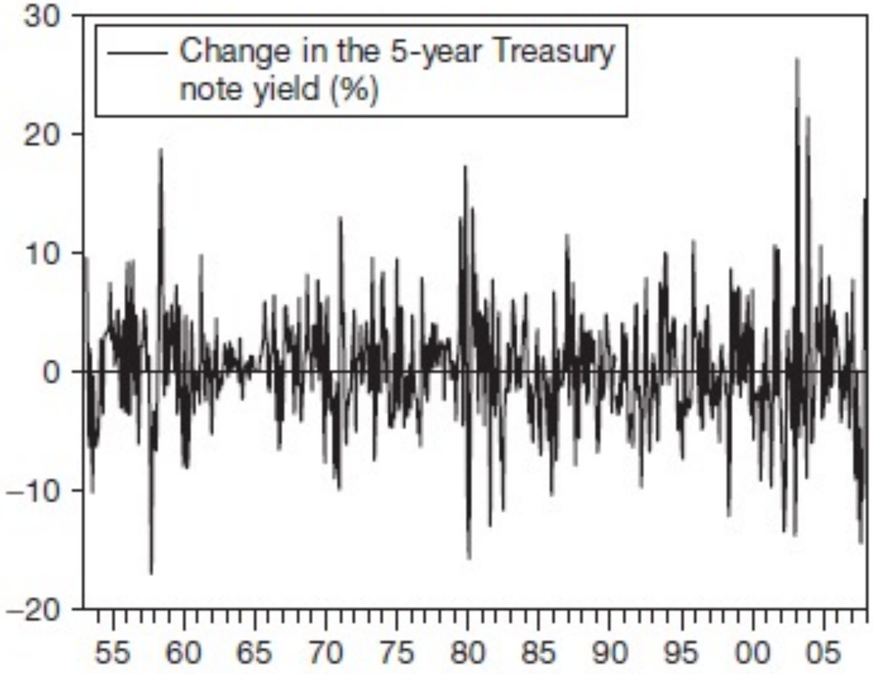
$$\boxed{-\frac{1}{\theta}Y_{t+1} = +\frac{1}{\theta^2}Y_{t+2}+\frac{1}{\theta^3}Y_{t+3}\cdots+\varepsilon_t}$$

The present is a function of the future.

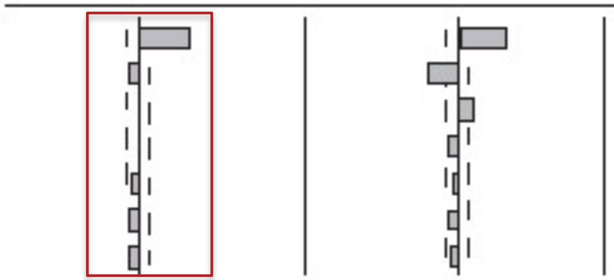
Autoregressive representation of a noninvertible MA(1) process.

Moving Average Models
Example: MA(1) Process

- What is the practical use of all this?



Sample: 1953M04 2008M04
Included observations: 660



Autocorrelation	Partial Correlation		AC	PAC
		1	0.339	0.339
		2	-0.073	-0.213
		3	0.007	0.129
		4	0.014	-0.063
		5	-0.043	-0.017
		6	-0.073	-0.060
		7	-0.069	-0.035

What process would you suggest?

$$\widehat{\rho}_1 = 0.339 \rightarrow \widehat{\theta}_1 \approx 0.34$$

$\boxed{\rightarrow \text{MA(1)}}$

Moving Average Models
Forecasting in an MA(1) Process

- Consider first the 1-step-ahead forecast, h = 1:
  MA(1): $Y_t = \mu + \theta\varepsilon_{t-1} + \varepsilon_t \rightarrow Y_{t+1} = \mu + \theta\varepsilon_t + \varepsilon_{t+1}$

- Optimal Point Forecast: $f_{t,1} = \mathrm{E}(Y_{t+1}|I_t) = \mu + \theta\varepsilon_t$
  - Note: We can write ε_t in terms of lags of Y_t

- One-period-ahead Forecast Error: $e_{t,1} = Y_{t+1} - f_{t,1} = \varepsilon_{t+1}$

- Uncertainty of the Forecast: $\sigma^2_{t+1|t} = var(Y_{t+1}|I_t) = \sigma^2_\varepsilon$

- Density Forecast: $f(Y_{t+1}|I_t) \sim N(\mu + \theta\varepsilon_t\,,\ \sigma^2_\varepsilon)$
  - Note: We can compute the confidence intervals from the density forecast.

Please go over the steps for h=1 and h=2 (Section 6.2[a])

Moving Average Models
Forecasting in an MA(1) Process

- Consider first the k-step-ahead forecast, *h* = k:

  MA(1): $Y_t = \mu + \theta\varepsilon_{t-1} + \varepsilon_t \rightarrow Y_{t+1} = \mu + \theta\varepsilon_t + \varepsilon_{t+1}$

- Optimal Point Forecast: $f_{t,k} = \mathrm{E}(Y_{t+k}|I_t) = \mu$

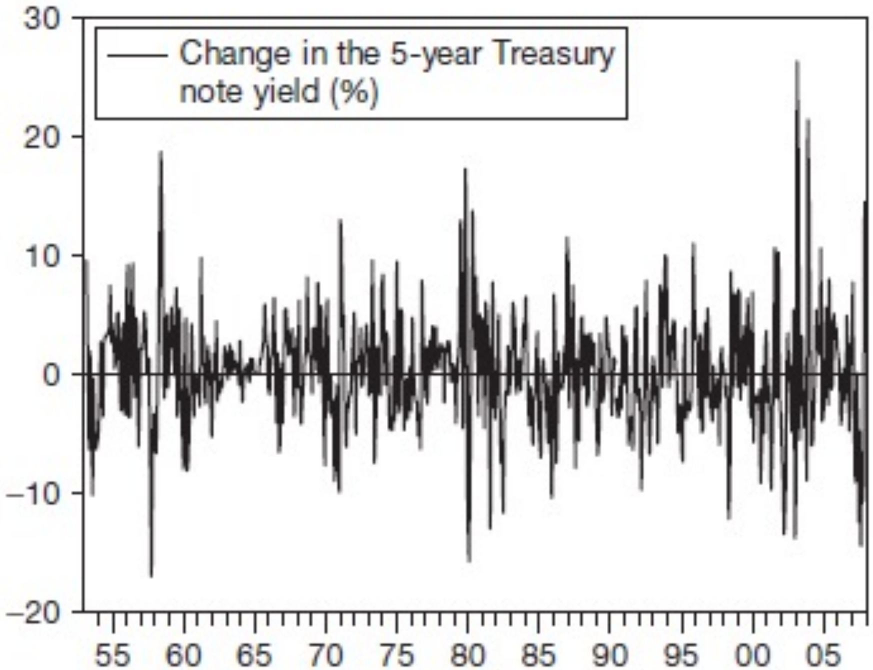
- k-period-ahead Forecast Error: $e_{t,k} = Y_{t+k} - f_{t,k} = \varepsilon_{t+k} + \theta\varepsilon_{t+k-1}$

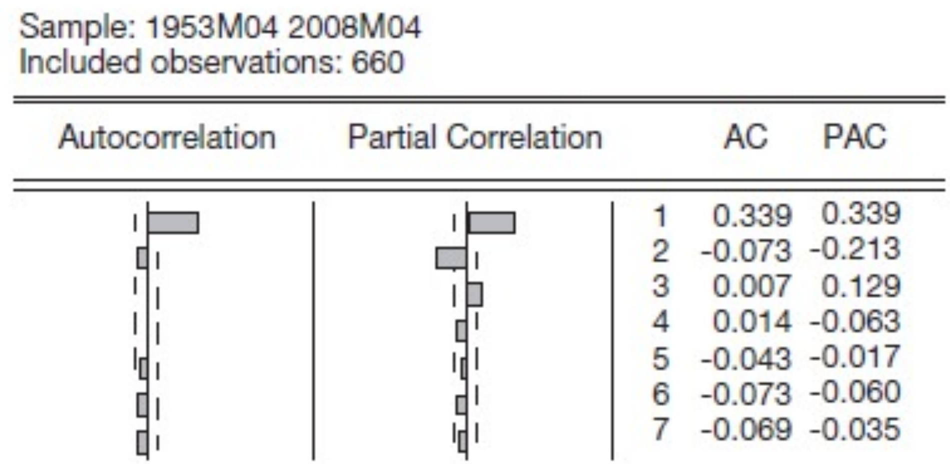
- Uncertainty of the Forecast: $\sigma^2_{t+k|t} = \sigma^2_{\varepsilon}(1+\theta^2) = \sigma^2_Y$

- Density Forecast: $f(Y_{t+k}|I_t) \sim N(\mu, \sigma^2_Y)$

Moving Average Models
Example: MA(1) Process

- Forecasting the 5-year Constant Maturity Yield





Sample: 1953M04 2008M04
Included observations: 660

Autocorrelation	Partial Correlation		AC	PAC
		1	0.339	0.339
		2	-0.073	-0.213
		3	0.007	0.129
		4	0.014	-0.063
		5	-0.043	-0.017
		6	-0.073	-0.060
		7	-0.069	-0.035

Model: $Y_t = \mu + \theta\varepsilon_{t-1} + \varepsilon_t$

Estimated Model: $Y_t = 0.160 + 0.485\varepsilon_t + \varepsilon_t$

where $\widehat{\sigma}_Y^2 = 23.683$ and $\widehat{\sigma}_\varepsilon = 4.449$

Estimation Output: 5-Year Treasury Yield (Monthly Percentage Changes)

Dependent Variable: DY
Method: Least Squares
Sample (adjusted): 1953M05 2007M11
Included observations: 655 after adjustments
Convergence achieved after 7 iterations
Backcast: 1953M04

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.160159	0.258095	0.620544	0.5351
MA(1)	0.485011	0.034468	14.07130	0.0000

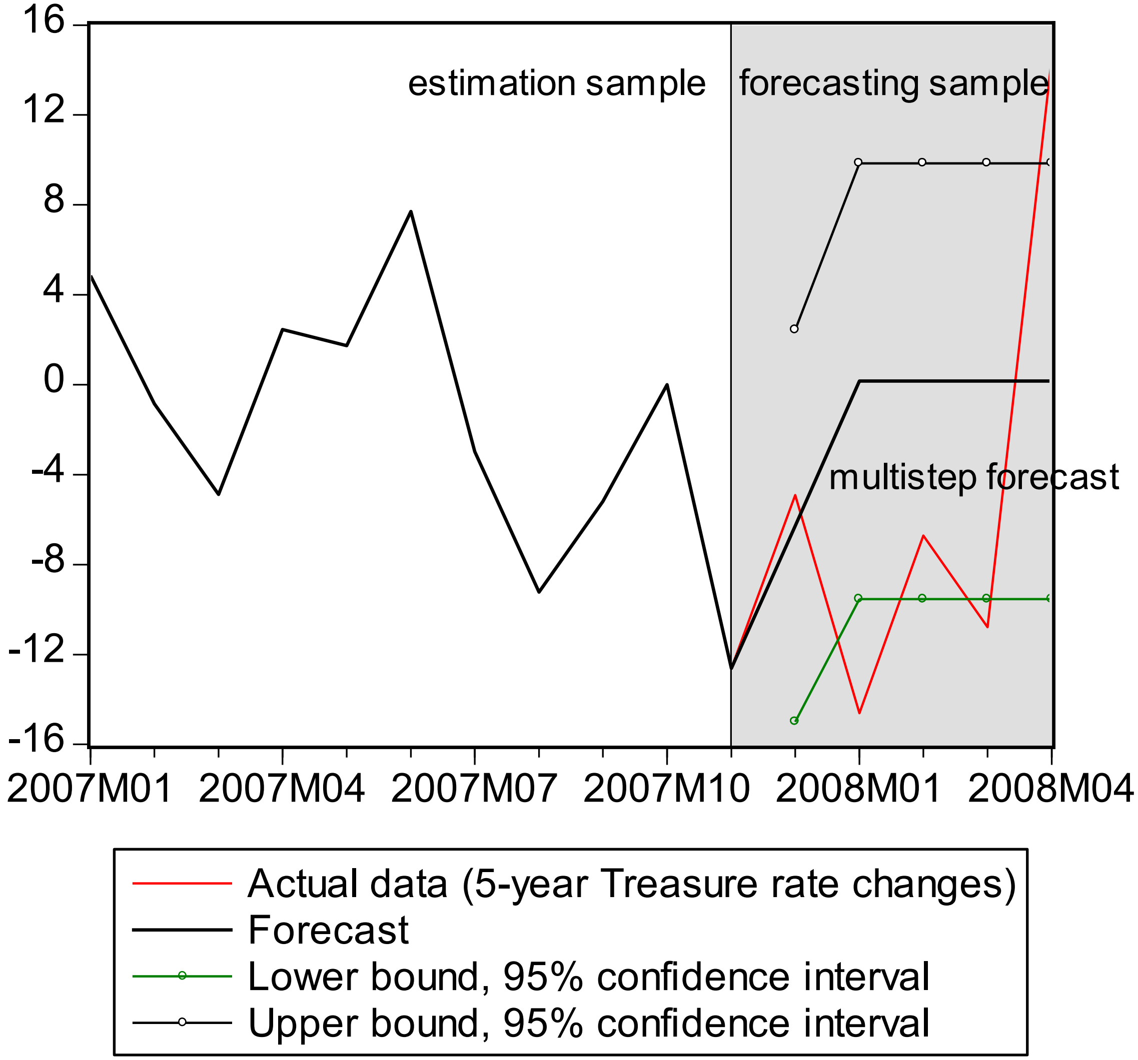
R-squared	0.165370	Mean dependent var	0.168613
Adjusted R-squared	0.164092	S.D. dependent var	4.866609
S.E. of regression	4.449443	Akaike info criterion	5.826484
Sum squared resid	12927.79	Schwarz criterion	5.840177
Log likelihood	-1906.173	F-statistic	129.3829
Durbin-Watson stat	2.055799	Prob(F-statistic)	0.000000

Inverted MA Roots	-.49

December 2007-April 2008 Forecasts of 5-year Treasure Yield Changes

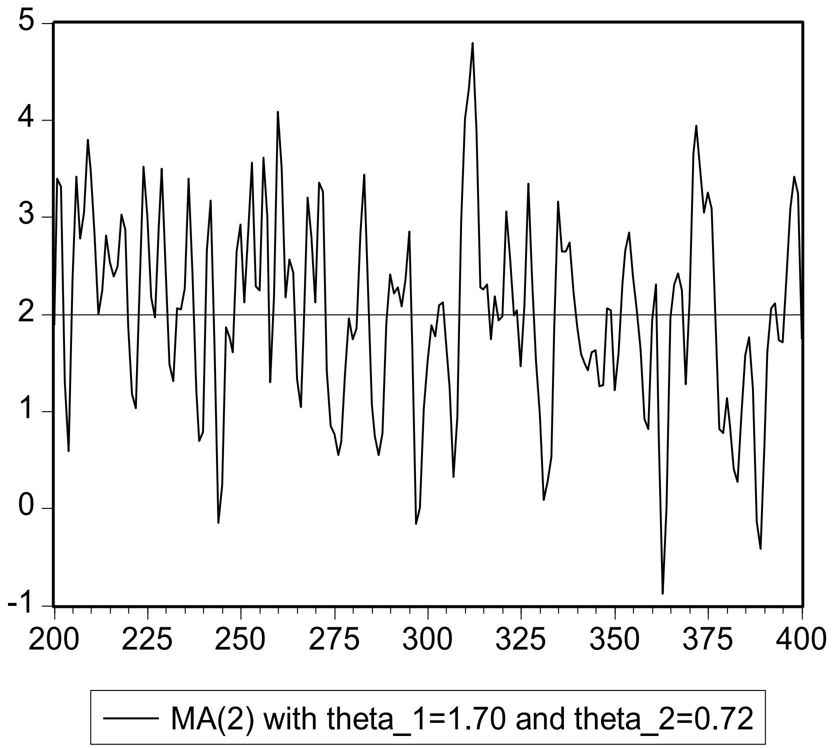
$h=1$ 12/2007	$f_{t,1}=\hat{\mu}+\hat{\theta}\varepsilon_t=$ $=0.160+0.485\hat{\varepsilon}_t$ $=-6.276\%$	$\sigma^2_{t+1\|t}=\hat{\sigma}^2_\varepsilon=4.449^2$	$f(Y_{t+1}\|I_t)\rightarrow N(\mu_{t+1\|t},\sigma^2_{t+1\|t})$ $=N(-6.276,4.449^2)$
$h=2$ 1/2008	$f_{t,2}=\hat{\mu}=0.160\%$	$\sigma^2_{t+2\|t}=\hat{\sigma}^2_\varepsilon(1+\hat{\theta}^2)$ $=4.449^2(1+0.485^2)$ $=23.683=\hat{\sigma}^2_Y$	$f(Y_{t+2}\|I_t)\rightarrow N(0.16,23.683)$
$h=3$ 2/2008	$f_{t,3}=\hat{\mu}=0.160\%$	$\sigma^2_{t+3\|t}=23.683=\hat{\sigma}^2_Y$	$f(Y_{t+3}\|I_t)\rightarrow N(0.16,23.683)$
$h=4$ 3/2008	$f_{t,4}=\hat{\mu}=0.160\%$	$\sigma^2_{t+4\|t}=23.683=\hat{\sigma}^2_Y$	$f(Y_{t+4}\|I_t)\rightarrow N(0.16,23.683)$
$h=5$ 4/2008	$f_{t,5}=\hat{\mu}=0.160\%$	$\sigma^2_{t+5\|t}=23.683=\hat{\sigma}^2_Y$	$f(Y_{t+5}\|I_t)\rightarrow N(0.16,23.683)$

Moving Average Models
Example: MA(1) Process



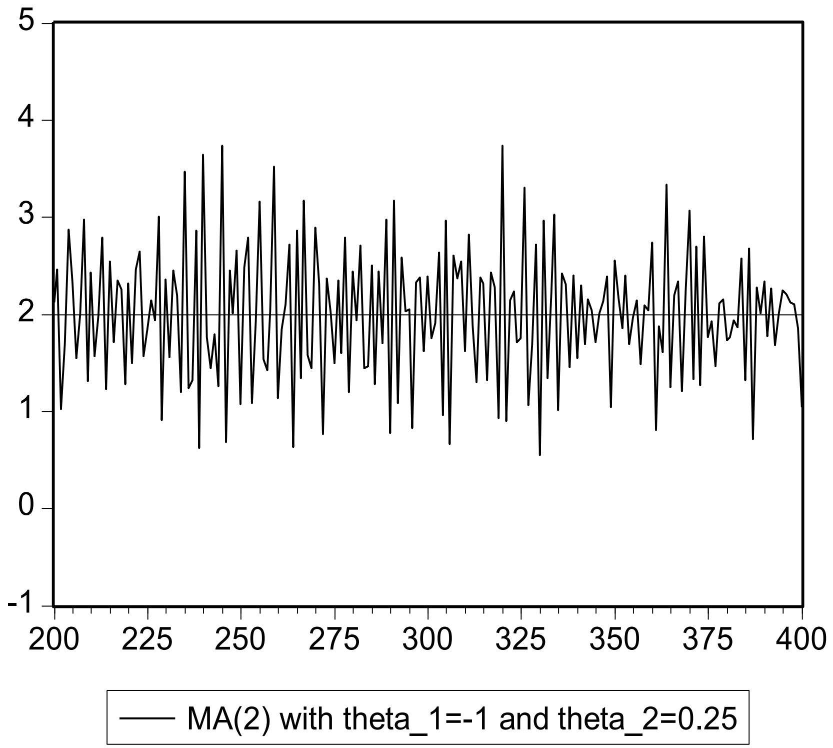
Moving Average Models
Example: MA(2) Process

$$Y_t = 2 + 1.7e_{t-1} + 0.72e_{t-2} + e_t$$



(a)

$$Y_t = 2 - \varepsilon_{t-1} + 0.25\varepsilon_{t-2} + \varepsilon_t$$

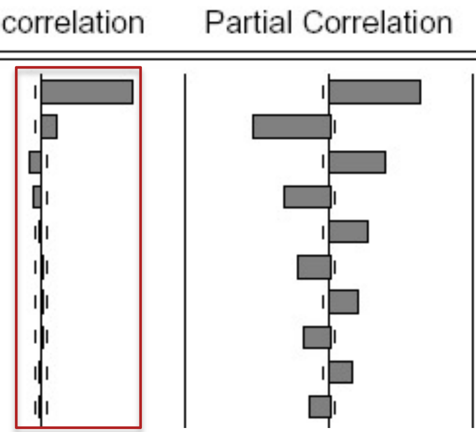


(b)

Both ACFs show
2 spikes only
→ MA(2)

Sample: 1 2000
Included observations: 1998

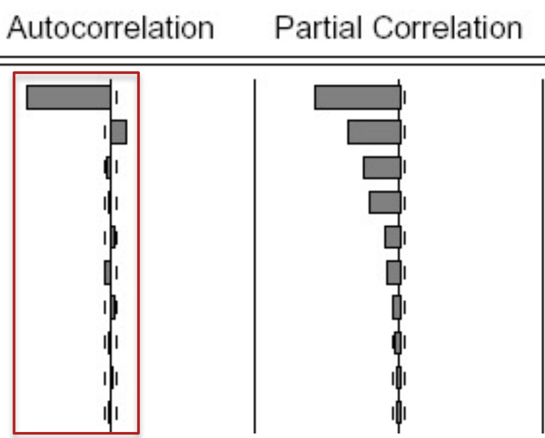
$$Y_t = 2 + 1.7\varepsilon_{t-1} + 0.72\varepsilon_{t-2} + \varepsilon_t$$



	AC	PAC
1	0.649	0.649
2	0.109	-0.542
3	-0.082	0.394
4	-0.054	-0.315
5	-0.007	0.271
6	0.016	-0.215
7	0.022	0.199
8	0.014	-0.184
9	-0.002	0.157
10	-0.010	-0.143

Sample: 1 2000
Included observations: 1998

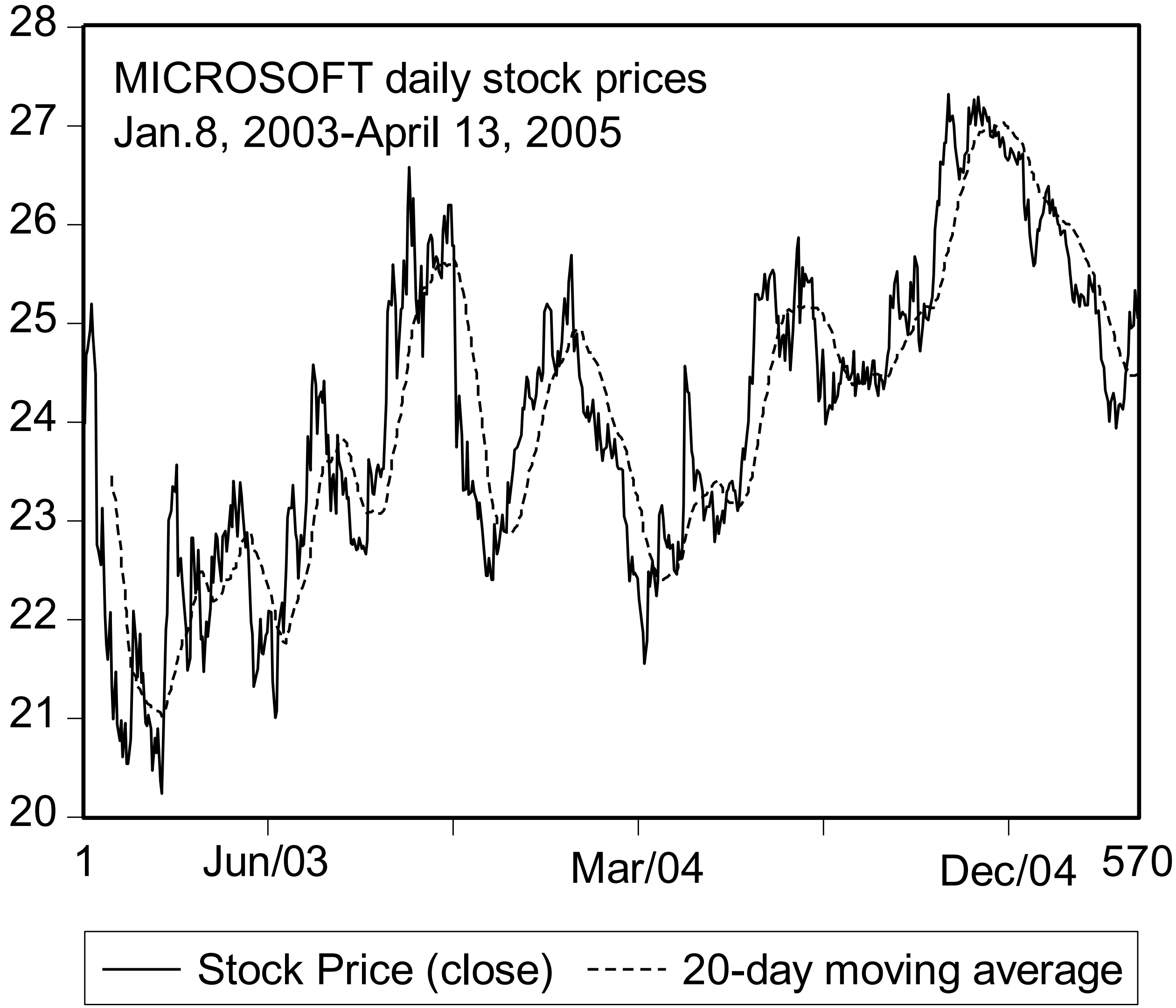
$$Y_t = 2 - \varepsilon_{t-1} + 0.25\varepsilon_{t-2} + \varepsilon_t$$



	AC	PAC
1	-0.592	-0.592
2	0.119	-0.356
3	-0.026	-0.250
4	-0.001	-0.199
5	0.031	-0.106
6	-0.034	-0.088
7	0.028	-0.043
8	-0.014	-0.022
9	0.008	-0.001
10	-0.012	-0.008

Moving Average Models

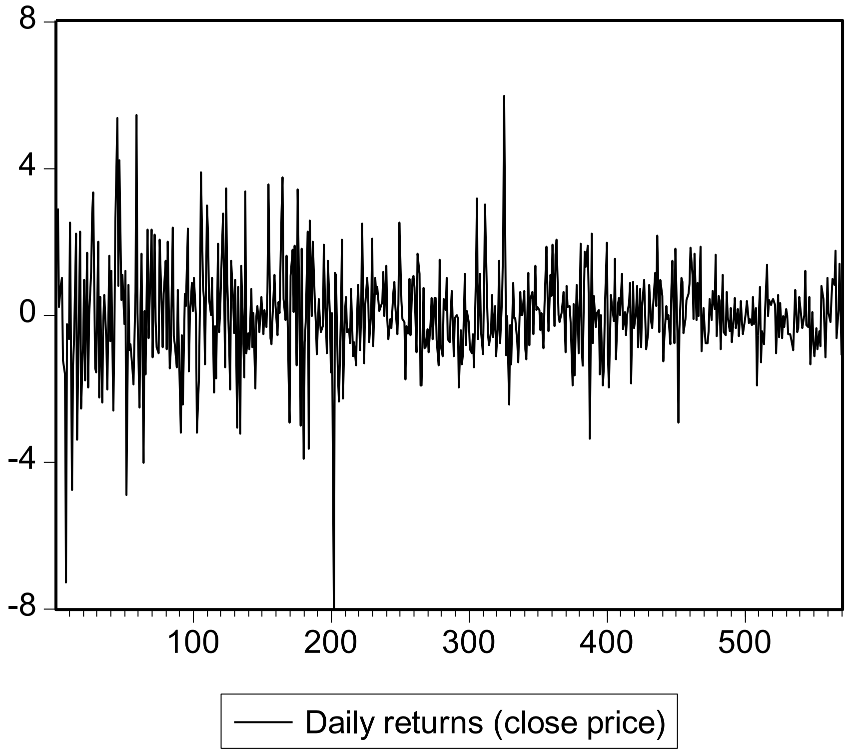
Example: MA(?) Process

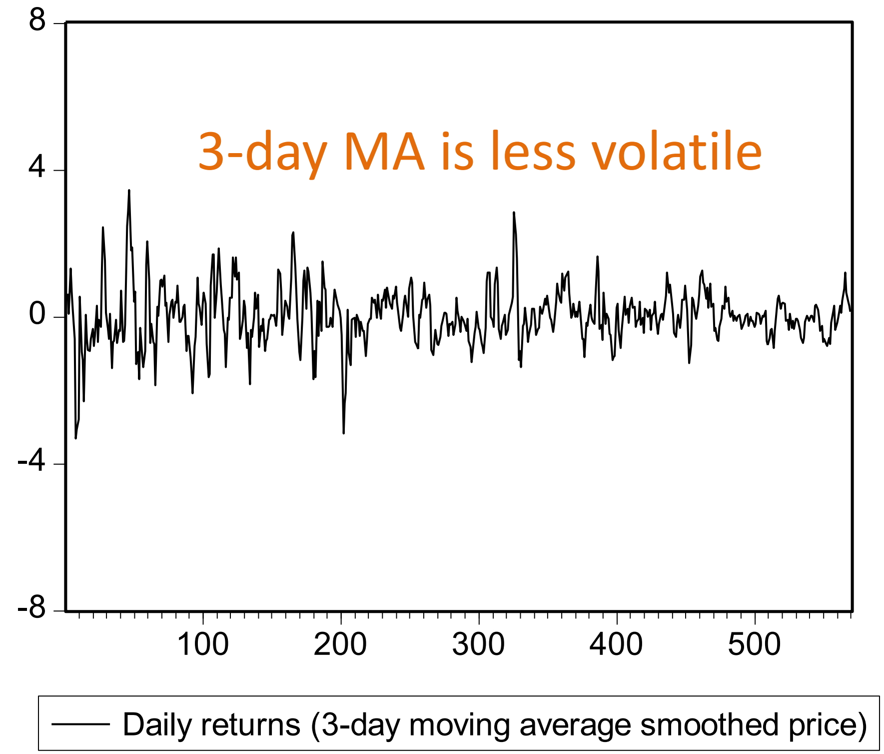


Moving Average Models

Example: MA(?) Process

Daily Returns to Microsoft

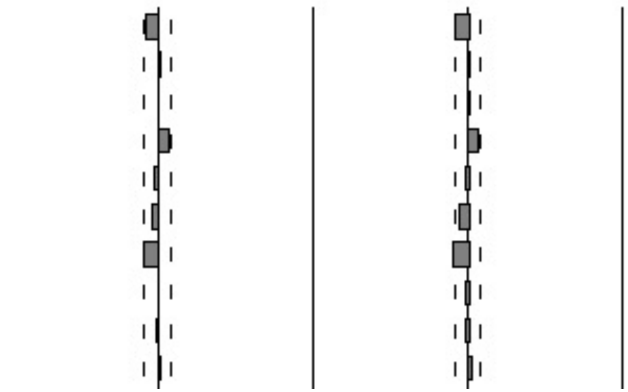




Moving Average Models
Example: MA(2) Process

Returns

Sample: 1 570
Included observations: 569

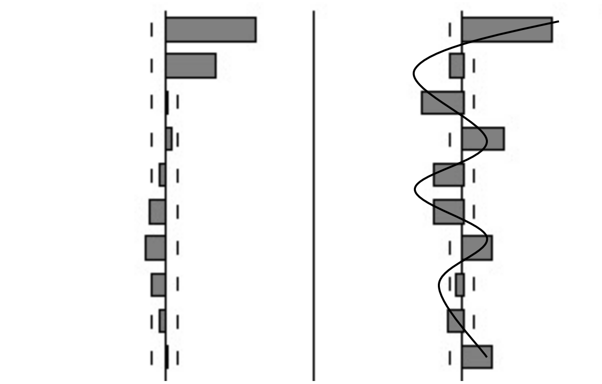


Lag	AC	PAC
1	-0.076	-0.076
2	0.012	0.006
3	0.002	0.004
4	0.067	0.068
5	-0.026	-0.016
6	-0.048	-0.053
7	-0.091	-0.100
8	0.003	-0.015
9	-0.014	-0.010
10	0.015	0.022

Looks like white-noise

3-day MA

Sample: 1 570
Included observations: 569



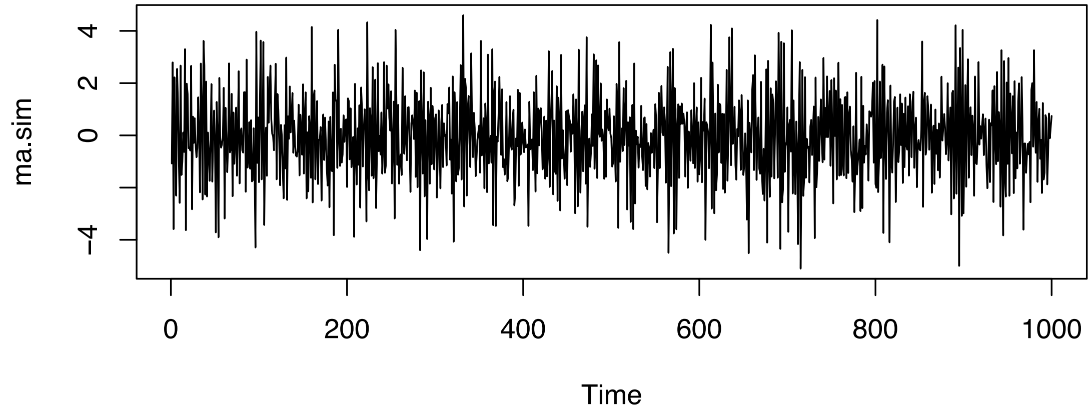
Lag	AC	PAC
1	0.629	0.629
2	0.350	-0.076
3	0.023	-0.277
4	0.047	0.300
5	-0.038	-0.193
6	-0.102	-0.194
7	-0.138	0.198
8	-0.084	-0.042
9	-0.037	-0.103
10	0.023	0.198

Looks like an MA(2) process

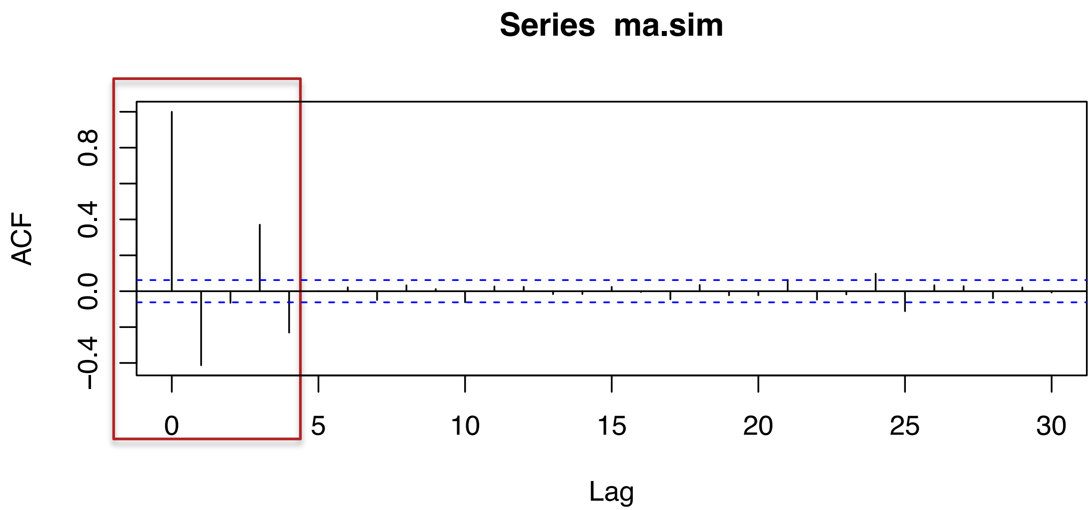
$$\boxed{Y_t = 0.95\varepsilon_{t-1} + 0.94\varepsilon_{t-2} + \varepsilon_t}$$

Moving Average Models
Example: MA(q) Process

What type of a process is this?



MA(4) Process



In general, for any MA(q) process, $\rho_k = 0$ for any $k>q$.

For Next Class

- Readings about today's class:

  Chapter 6[a], 8[b]

- Review Exercises / Problems:

  Chapter 7[b]: 1, 2, 3

- Readings for next class:

  6[a] ,7[a], 8[b], 9[b]