# Eco 4306 Economic and Business Forecasting

## Chapter 6: Forecasting with Moving Average (MA) Processes

# Outline

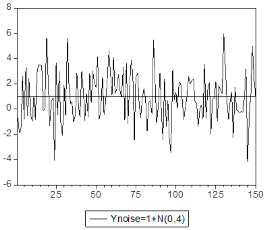
- we now start building time series models
- first, we introduce **white noise process**, characterized by absence of linear time dependence
- after that we will develop a **moving average model** for processes which do exhibit some linear time dependence

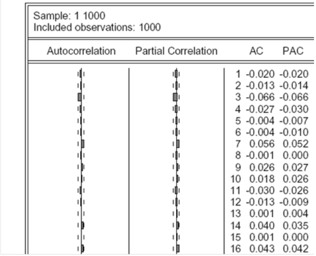
# 6.1 A Model with No Dependence: White Noise

- a stationary stochastic process $\{\varepsilon_t\}$ is called **white noise process** if $\rho_k = 0$ for $k \geq 1$, and $r_k = 0$ for $k \geq 1$, that is if autocorrelation and partial autocorrelation functions are zero
- no linear dependence, autocorrelations are zero - no link between past and present observations, no link between present and future observations
- no dependence to exploit so we cannot predict future realizations of the process - $\varepsilon_t$ is unpredictable shock, residual in our time series models

# 6.1 A Model with No Dependence: White Noise

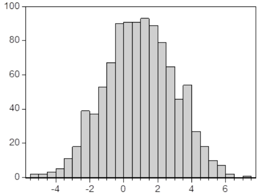
- consider stochastic process $Y_t = 1 + \varepsilon_t$ where $\varepsilon_t \sim N(0, 4)$
- theoretical unconditional mean and variance of $\{Y_t\}$ are $E(Y_t) = E(1 + \varepsilon_t) = 1$ and $var(Y_t) = var(1 + \varepsilon_t) = 4$
- first two population moments are thus time invariant



Sample: 1 1000
Included observations: 1000

| Autocorrelation | Partial Correlation | | AC | PAC |
|---|---|---|---|---|
| | | 1 | -0.020 | -0.020 |
| | | 2 | -0.013 | -0.014 |
| | | 3 | -0.066 | -0.066 |
| | | 4 | -0.027 | -0.030 |
| | | 5 | -0.004 | -0.007 |
| | | 6 | -0.004 | -0.010 |
| | | 7 | 0.056 | 0.052 |
| | | 8 | -0.001 | 0.000 |
| | | 9 | 0.026 | 0.027 |
| | | 10 | 0.018 | 0.026 |
| | | 11 | -0.030 | -0.026 |
| | | 12 | -0.013 | -0.009 |
| | | 13 | 0.001 | 0.004 |
| | | 14 | 0.040 | 0.035 |
| | | 15 | 0.001 | 0.000 |
| | | 16 | 0.043 | 0.042 |




Series: Ynoise
Sample 1 1000
Observations 1000

| | |
|---|---|
| Mean | 0.957014 |
| Median | 0.941058 |
| Maximum | 7.235267 |
| Minimum | -5.443433 |
| Std. Dev. | 2.034487 |
| Skewness | -0.029960 |
| Kurtosis | 2.784223 |
| Jarque-Bera | 2.089597 |
| Probability | 0.351763 |

# 6.1 A Model with No Dependence: White Noise

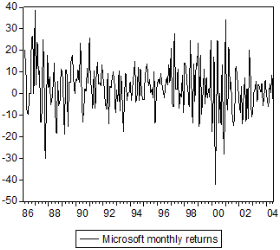
- time series $\{y_t\}$ looks very ragged
- histogram for $\{y_t\}$ its has the expected bell shape corresponding to a normal distribution
- skewness is approximately zero, kurtosis approximately 3, Jarque-Bera test indicates that normality is not rejected $(p = 0.351)$
- first two sample moments are very close to the population moments - sample mean is 0.96, sample standard deviation is 2.03
- time series plot shows that realizations bounce around a mean value of 1 and volatility does not appear to change significantly
- AC function and PAC function at all lags are not significantly different from 0 at 5% level
- time series $\{y_t\}$ is thus a white noise process

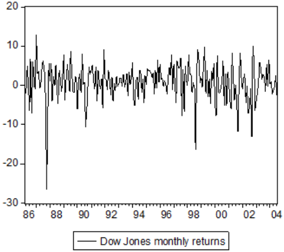
# 6.1 A Model with No Dependence: White Noise

- in business and economics some data behave very similarly to a white noise process
- white noise processes are especially common among financial series
- this is the reason why these data are so difficult to predict - they do not exhibit any temporal linear dependence that could be consistently exploited
- for example: returns for individual stocks and for stock market indices have correlograms that resemble a white noise process

# 6.1 A Model with No Dependence: White Noise

- returns, Microsoft and DJ Index, 1986M4-2004M7, **Figure06_02_MSFT_DJ.xls**
- all lags of AC and PAC functions are not significantly different from 0 at 5% level




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

| Lag | 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

| Lag | 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 |

# 6.3 Forecasting with Moving Average Models

- a **moving average process of order** $q$, referred to as MA$(q)$, has the form

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

  where $\varepsilon_t$ is a zero-mean white noise process
- order of the model is given by the largest lag, not by the number of lag variables in the right-hand side
- for instance

$$Y_t = \mu + \varepsilon_t + \theta_3 \varepsilon_{t-3}$$

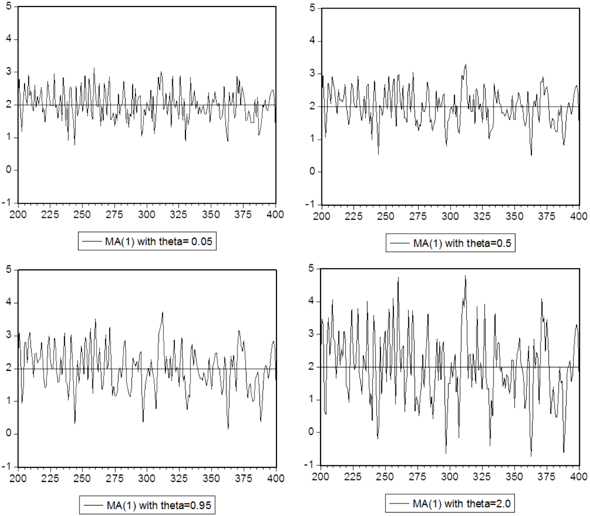
  is an MA(3) because the largest lag is 3 although there is only one lagged variable

# 6.3 Forecasting with Moving Average Models

- we will next look at the statistical properties of MA models
- our ultimate objective is constructing the optimal forecast
- we will analyze the lowest order process, MA$(1)$, generalization to MA$(q)$ is straightforward
- three questions we want to answer
  1. What does a time series of an MA process look like?
  2. How do the AC and PAC functions for MA process look like?
  3. What is the optimal forecast for an MA process?

## 6.3.1 MA(1) Process

- consider MA(1) process $Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$
- four simulations, each 200 observations of MA(1), with different values of $\theta \in \{0.05, 0.5, 0.95, 2.0\}$, but with same $\mu = 2$, and $\varepsilon_t \sim N(0, 0.25)$



# 6.3.1 MA(1) Process

- four time series seem to be weakly stationary
- unconditional population mean is time invariant

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

- unconditional variance is also time invariant

  $$var(Y_t) = E(Y_t - E(Y_t))^2 = E(\varepsilon_t + \theta\varepsilon_{t-1})^2 = E(\varepsilon_t^2 + 2\theta\varepsilon_t\varepsilon_{t-1} + \theta^2\varepsilon_{t-1}^2) = (1+\theta^2)\sigma_\varepsilon^2$$

  and it is increasing with $\theta$
- we still need to verify that the autocorrelation function does not depend on time to claim that the process is covariance stationary

# 6.3.1 MA(1) Process

- only $\hat{\rho}_1$ in the sample AC function is significantly different from zero
- $\hat{\rho}_1$ is proportional to $\theta$ for $|\theta| < 1$ and its sign is the same as the sign of $\theta$

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 |

# 6.3.1 MA(1) Process

- population autocovariance of order 1

$$\gamma_1 = E[(Y_t - \mu)(Y_{t-1} - \mu)] = E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] = \theta\sigma_\varepsilon^2$$

- population autocorrelation of order 1

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta\sigma_\varepsilon^2}{(1+\theta^2)\sigma_\varepsilon^2} = \frac{\theta}{1+\theta^2}$$

- population autocovariance of order $k > 1$

$$\gamma_k = E[(Y_t - \mu)(Y_{t-k} - \mu)] = E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-k} + \theta\varepsilon_{t-k-1})] = 0$$

  thus for population autocorrelation of order $k > 1$ we have

$$\rho_k = \frac{\gamma_k}{\gamma_0} = 0$$

- so autocorrelation function really does not depend on time, and thus MA$(1)$ is covariance stationary process

# 6.3.1 MA(1) Process

- note that AC function and PAC function for the MA(1) processes with $\theta = 0.5$ and $\theta = 2$ are identical
- this is due to the fact that for the MA(1) with parameter $\hat{\theta} = \frac{1}{\theta}$ we get

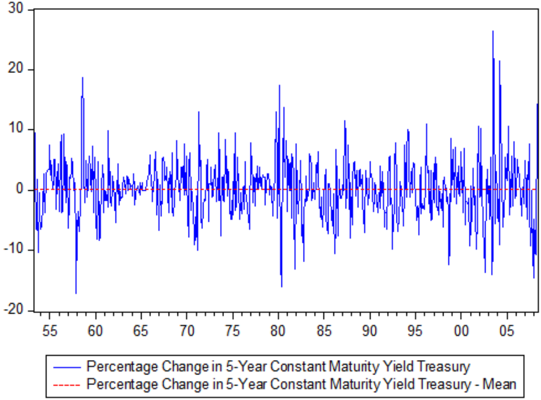
$$\rho_1 = \frac{\hat{\theta}}{1 + \hat{\theta}^2} = \frac{\frac{1}{\theta}}{1 + \frac{1}{\theta^2}} = \frac{\theta}{\theta^2 + 1} = \frac{\theta}{1 + \theta^2}$$

- an MA(1) process is called **invertible** if $|\theta| < 1$
- if an MA process is invertible, we can always find an autoregressive representation in which the present $Y_t$ is a function of the past $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots$

# 6.3.1 MA(1) Process

- we will next analyze and forecast the percentage change in 5-Year Constant Maturity Yield on Treasury Securities, using April 1953 to April 2008 sample
- data available at FRED **https://fred.stlouisfed.org/graph/?g=mXGI** and **Figure06_05_Table6_1_treasury.xls**
- U.S. Treasury securities are considered to be the least risky assets
- they constitute an asset of reference to monitor the level of risk of other fixed-income securities such as grade bonds and certificates of deposit
- risk spread - difference between the yield of the fixed-income security and the yield of a corresponding Treasury security with the same maturity

# 6.3.1 MA(1) Process



## 6.3.1 MA(1) Process

- AC function and PAC function similar to those for MA(1)
- AC function has only one positive spike at $\hat{\rho}_1$, remaining autocorrelations are not significantly different from zero
- PAC function alternating signs, decreasing toward zero

Date: 02/07/18 Time: 22:03
Sample: 1953M05 2008M04
Included observations: 660

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.332 | 0.332 | 72.915 | 0.000 |
| | | 2 | -0.053 | -0.183 | 74.786 | 0.000 |
| | | 3 | 0.002 | 0.097 | 74.789 | 0.000 |
| | | 4 | 0.022 | -0.027 | 75.108 | 0.000 |
| | | 5 | -0.042 | -0.042 | 76.278 | 0.000 |
| | | 6 | -0.060 | -0.029 | 78.664 | 0.000 |
| | | 7 | -0.070 | -0.058 | 81.958 | 0.000 |
| | | 8 | 0.025 | 0.075 | 82.362 | 0.000 |
| | | 9 | 0.080 | 0.038 | 86.667 | 0.000 |
| | | 10 | 0.036 | 0.005 | 87.563 | 0.000 |
| | | 11 | 0.029 | 0.033 | 88.126 | 0.000 |
| | | 12 | -0.044 | -0.089 | 89.448 | 0.000 |

## 6.3.1 MA(1) Process

- recall: under quadratic loss function the optimal point forecast is conditional mean, $f_{t,h} = \mu_{t+h|t} = E(Y_{t+h}|I_t)$
- we next analyze this optimal forecast under quadratic loss function for $h = 1, 2, \dots$
- we will see that forecasting with an MA(1) is rather limited by the very short memory of the process - for $h > 1$ the optimal forecast is identical to the unconditional mean of the process

## 6.3.1 MA(1) Process

- for MA(1) model and forecasting horizon $h = 1$ we have

i. optimal point forecast

$$f_{t,1} = \mu_{t+1|t} = E(Y_{t+1}|I_t) = E(\mu + \varepsilon_{t+1} + \theta\varepsilon_t) = \mu + \theta\varepsilon_t$$

ii. 1-period-ahead forecast error

$$e_{t,1} = Y_{t+1} - f_{t,1} = \mu + \varepsilon_{t+1} + \theta\varepsilon_t - \mu - \theta\varepsilon_t = \varepsilon_{t+1}$$

iii. uncertainty associated with the forecast is summarized by the variance of the forecast error

$$\sigma^2_{t+1|t} = var(e_{t+1}) = \sigma^2_\varepsilon$$

iv. density forecast is the conditional probability density function of the process at the future date; assuming $\varepsilon_t$ is normally distributed we have

$$f(Y_{t+1}|I_t) \sim N(\mu + \theta\varepsilon_t, \sigma^2_\varepsilon)$$

v. using the density forecast we can construct interval forecasts - since for $Z \sim N(0,1)$ we have $P(-1.96 \leq Z \leq 1.96) = 0.95$, the 95% interval forecast for $Y_{t+1}$ is

$$\mu_{t+1|t} \pm 1.96\sigma_{t+1|t} = (\mu_{t+1|t} - 1.96\sigma_{t+1|t}, \mu_{t+1|t} + 1.96\sigma_{t+1|t})$$

## 6.3.1 MA(1) Process

- for MA(1) model and forecasting horizon $h = 2$ we have

i. optimal point forecast

$$f_{t,2} = \mu_{t+2|t} = E(Y_{t+2}|I_t) = E(\mu + \varepsilon_{t+2} + \theta\varepsilon_{t+1}) = \mu$$

ii. 2-period-ahead forecast error

$$e_{t,2} = Y_{t+2} - f_{t,2} = \mu + \varepsilon_{t+2} + \theta\varepsilon_{t+1} - \mu = \varepsilon_{t+2} + \theta\varepsilon_{t+1}$$

iii. uncertainty associated with the forecast is summarized by the variance of the forecast error

$$\sigma^2_{t+2|t} = var(e_{t+2}) = (1 + \theta^2)\sigma^2_\varepsilon$$

iv. density forecast is the conditional probability density function of the process at the future date; assuming $\varepsilon_t$ is normally distributed we have

$$f(Y_{t+2}|I_t) \sim N(\mu, (1 + \theta^2)\sigma^2_\varepsilon)$$

# 6.3.1 MA(1) Process

- for MA(1) model and forecasting horizon $h = s$ we have

i. optimal point forecast

$$f_{t,s} = \mu_{t+s|t} = E(Y_{t+s}|I_t) = \mu$$

ii. $s$-period-ahead forecast error

$$e_{t,s} = Y_{t+s} - f_{t,s} = \mu + \varepsilon_{t+s} + \theta\varepsilon_{t+s-1} - \mu = \varepsilon_{t+s} + \theta\varepsilon_{t+s-1}$$

iii. uncertainty associated with the forecast is summarized by the variance of the forecast error

$$\sigma^2_{t+s|t} = var(e_{t+s}) = (1 + \theta^2)\sigma^2_\varepsilon$$

iv. density forecast is the conditional probability density function of the process at the future date; assuming $\varepsilon_t$ is normally distributed we have

$$f(Y_{t+s}|I_t) \sim N(\mu, (1 + \theta^2)\sigma^2_\varepsilon)$$

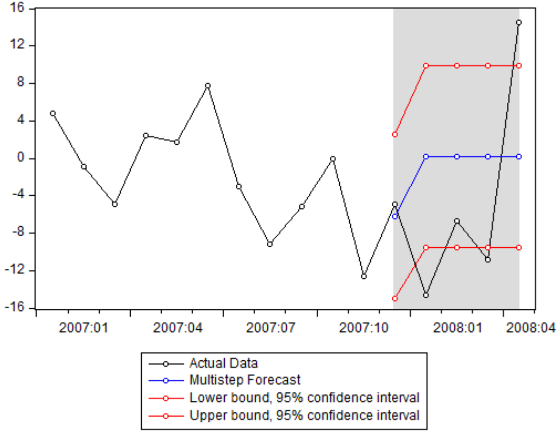
# 6.3.1 MA(1) Process

Forecasting 5-year Constant Maturity Yield on Treasury Securities:

- AC and PAC suggest that the Percentage Change in 5-year Constant Maturity Yield on Treasury Securities follows an MA(1) process
- we will use 1953M5-2007M11 as estimation sample and 2007M12-2008M4 as prediction sample
- we will thus construct forecast for $h = 1, 2, \ldots, 5$ so 1-step to 5-step ahead forecasts
- to estimate $\theta$ in EViews choose **Object** $\rightarrow$ **New Object** $\rightarrow$ **Equation**, in equation specification write **dy c MA(1)**, and in sample 1953M5-2007M11
- afterwards to create a multistep forecast in EViews open the equation and choose **Proc** $\rightarrow$ **Forecast**, enter name dyf_se for standard deviation $\sigma_{t+h|t}$ into "S.E. (optional)", change forecast sample to 2007M12-2008M4, and select "Dynamic forecast" method in the forecast window
- to construct the lower and the upper bounds of the 95% confidence interval $(\mu_{t+h|t} - 1.96\sigma_{t+h|t}, \mu_{t+h|t} + 1.96\sigma_{t+h|t})$ choose **Object** $\rightarrow$ **Generate series** set sample to 2007M12-2008M4 and enter first **dyf_lb = dyf - 1.96*dyf_se** and then the second time **dyf_ub = dyf + 1.96*dyf_se**

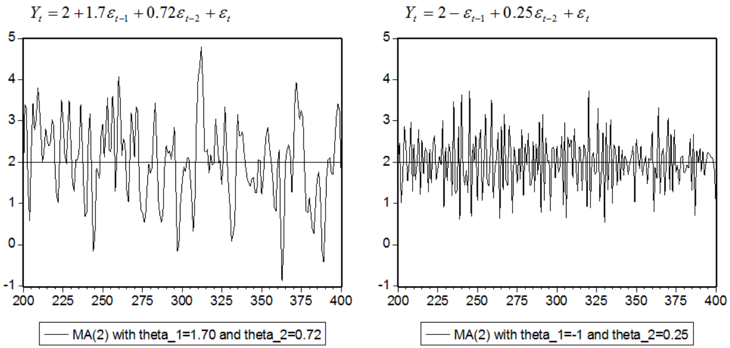
## 6.3.1 MA(1) Process

Forecasting 5-year Constant Maturity Yield on Treasury Securities:



# 6.3.2 MA(2) Process

- consider now an MA(2) process $Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$
- two simulations, each 200 observations of MA(2), with $\theta_1 = 1.70, \theta_2 = 0.72$ and with $\theta_1 = -1, \theta_2 = 0.25$, in addition to $\mu = 2$ and $\varepsilon_t \sim N(0, 0.25)$




- unconditional population mean is time invariant

$$E(Y_t) = E(\mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}) = \mu$$

- unconditional variance is also time invariant

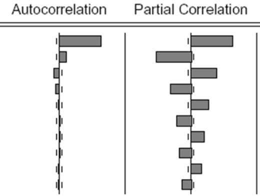
$$var(Y_t) = E(Y_t - E(Y_t))^2 = E(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2})^2 = (1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2$$

## 6.3.2 MA(2) Process

- first two components in sample AC function are different from zero $\hat{\rho}_1 \neq 0$, $\hat{\rho}_2 \neq 0$, remaining autocorrelations are equal to zero, $\hat{\rho}_k = 0$ for $k > 2$
- sample PAC function decreases toward zero

Sample: 1 2000
Included observations: 1998

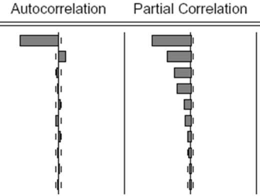
$$Y_t = 2 + 1.7\varepsilon_{t-1} + 0.72\varepsilon_{t-2} + \varepsilon_t$$




| | 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 |

## 6.3.2 MA(2) Process

- population autocovariance of order 1

$$\gamma_1 = E[(Y_t - \mu)(Y_{t-1} - \mu)] = (\theta_1 + \theta_1\theta_2)\sigma_\varepsilon^2$$

thus for population autocorrelation of order 1 we have

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

- population autocovariance of order $k = 2$

$$\gamma_2 = E[(Y_t - \mu)(Y_{t-2} - \mu)] = \theta_2\sigma_\varepsilon^2$$

thus for population autocorrelation of order $k = 2$ we have

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

- autocorrelations of higher order are all equal to zero

$$\gamma_k = E[(Y_t - \mu)(Y_{t-k} - \mu)] = 0$$

thus for population autocorrelation of order $k > 2$ we have

$$\rho_k = \frac{\gamma_k}{\gamma_0} = 0$$

- since autocorrelation function does not depend on time, MA$(2)$ is covariance stationary process

## 6.3.2 MA(2) Process

- for MA(2) model and forecasting horizon $h = 1$ we have

i. optimal point forecast

$$f_{t,1} = \mu_{t+1|t} = E(Y_{t+1}|I_t) = E(\mu + \varepsilon_{t+1} + \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1}) = \mu + \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1}$$

ii. 1-period-ahead forecast error

$$e_{t,1} = Y_{t+1} - f_{t,1} = \mu + \varepsilon_{t+1} + \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1} - \mu - \theta_1\varepsilon_t - \theta_2\varepsilon_{t-1} = \varepsilon_{t+1}$$

iii. uncertainty associated with the forecast is summarized by the variance of the forecast error

$$\sigma^2_{t+1|t} = var(e_{t+1}) = \sigma^2_\varepsilon$$

iv. density forecast is the conditional probability density function of the process at the future date; assuming $\varepsilon_t$ is normally distributed we have

$$f(Y_{t+1}|I_t) \sim N(\mu + \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1}, \sigma^2_\varepsilon)$$

v. using the density forecast we can construct interval forecasts - since for $Z \sim N(0,1)$ we have $P(-1.96 \leq Z \leq 1.96) = 0.95$, the 95% interval forecast for $Y_{t+1}$ is

$$\mu_{t+1|t} \pm 1.96\sigma_{t+1|t} = (\mu_{t+1|t} - 1.96\sigma_{t+1|t}, \mu_{t+1|t} + 1.96\sigma_{t+1|t})$$

# 6.3.2 MA(2) Process

- for MA(2) model and forecasting horizon $h = 2$ we have

i. optimal point forecast

$$f_{t,2} = \mu_{t+2|t} = E(Y_{t+2}|I_t) = E(\mu + \varepsilon_{t+2} + \theta_1\varepsilon_{t+1} + \theta_2\varepsilon_t) = \mu + \theta_2\varepsilon_t$$

ii. 2-period-ahead forecast error

$$e_{t,2} = Y_{t+2} - f_{t,2} = \mu + \varepsilon_{t+2} + \theta_1\varepsilon_{t+1} + \theta_2\varepsilon_t - \mu - \theta_2\varepsilon_t = \varepsilon_{t+2} + \theta_1\varepsilon_{t+1}$$

iii. uncertainty associated with the forecast is summarized by the variance of the forecast error

$$\sigma^2_{t+2|t} = var(e_{t+2}) = (1 + \theta_1^2)\sigma^2_\varepsilon$$

iv. density forecast is the conditional probability density function of the process at the future date; assuming $\varepsilon_t$ is normally distributed we have

$$f(Y_{t+2}|I_t) \sim N(\mu + \theta_2\varepsilon_t, (1 + \theta_1^2)\sigma^2_\varepsilon)$$

# 6.3.2 MA(2) Process

- for MA(2) model and forecasting horizon $h = s$ with $s > 2$ we have

i. optimal point forecast

$$f_{t,s} = \mu_{t+s|t} = E(Y_{t+s}|I_t) = \mu$$

ii. $s$-period-ahead forecast error

$$e_{t,s} = Y_{t+s} - f_{t,s} = \mu + \varepsilon_{t+s} + \theta\varepsilon_{t+s-1} - \mu = \varepsilon_{t+s} + \theta_1\varepsilon_{t+s-1} + \theta_2\varepsilon_{t+s-2}$$

iii. uncertainty associated with the forecast is summarized by the variance of the forecast error

$$\sigma^2_{t+s|t} = var(e_{t+s}) = (1 + \theta_1^2 + \theta_2^2)\sigma^2_\varepsilon$$

iv. density forecast is the conditional probability density function of the process at the future date; assuming $\varepsilon_t$ is normally distributed we have

$$f(Y_{t+s}|I_t) \sim N(\mu, (1 + \theta_1^2 + \theta_2^2)\sigma^2_\varepsilon)$$

- forecasting with an MA(2) is thus limited by the short memory - for $h > 2$ the optimal forecast is identical to the unconditional mean of the process

# 6.3.2 MA($q$) Process

- for MA($q$) the AC and PAC functions satisfy similar properties as those for MA(2) process
- first $q$ components in sample AC function are different from zero $\hat{\rho}_k \neq 0$ for $k = 1, 2, \ldots, q$
- remaining autocorrelations are equal to zero $\hat{\rho}_k = 0$ for $k > q$
- sample PAC function decreases toward zero (in exponential or in oscillating pattern)
- forecasting with an MA($q$) is quite limited - for $h > q$ the optimal forecast is identical to the unconditional mean of the process

## 6.3.2 MA($q$) Process

Forecasting Growth of Employment in Nonfarm Business Sector

- download PRS85006013.csv obtained from fred.stlouisfed.org/series/PRS85006013, import it into EViews as time series `emp`
- generate time series `gemp` as percentage change of `emp`
- AC and PAC suggest that MA(3) process $y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3}$ can be used to model `gemp`
- we will use 1947Q2-2014Q4 as estimation sample and 2015Q1-2018Q4 as prediction sample
- we will thus construct forecast for $h = 1, 2, \ldots, 16$ so 1-step to 16-step ahead forecasts

# 6.3.2 MA($q$) Process

- first, to estimate parameters $\mu, \theta_1, \theta_2, \theta_3$ of the MA(3) process

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3}$$

  in EViews choose **Object** $\rightarrow$ **New Object** $\rightarrow$ **Equation**, in equation specification write **gemp c MA(1) MA(2) MA(3)**, and in sample write 1947Q2-2014Q4

- afterwards to create a multistep forecast in EViews open the equation and choose **Proc** $\rightarrow$ **Forecast**, enter name gempf_se for standard deviation $\sigma_{t+h|t}$ into "S.E. (optional)", change forecast sample to 2015Q1-2018Q4, and select "Dynamic forecast" method in the forecast window

- to construct the lower and the upper bounds of the 95% confidence interval $(\mu_{t+h|t} - 1.96\sigma_{t+h|t}, \mu_{t+h|t} + 1.96\sigma_{t+h|t})$ choose **Object** $\rightarrow$ **Generate series** set sample to 2015Q1-2018Q4 and enter first **gempf_lb = gempf - 1.96*gempf_se** and then the second time **gempf_ub = gempf + 1.96*gempf_se**

- to construct time series with unconditional mean of gemp choose **Object** $\rightarrow$ **Generate Series** and enter **gemp_mean = @mean(gemp)**