# Times Series and Forecasting (VII)

## Chapter 7. Estimation

Jianhua Hu

hu.jianhua@mail.shufe.edu.cn

Spring 2013, SHUFE

# 7.1. Introduction

- In the last chapter, we were concerned with model selection, that is, identifying values of $p$, $d$, and $q$ which were consistent with the observed (or suitably transformed) data.
- In this chapter, our efforts are directed towards estimating parameters in this class of models.
- In doing so, it suffices to restrict attention to stationary ARMA$(p, q)$ models.

# Preview

- If $d > 0$ (which corresponds to a nonstationary process), the methodology described herein can be applied to the suitably differenced process $(1 - B)^d Y_t = \nabla^d Y_t$.
  - Therefore, when we write $Y_1, Y_2, \cdots, Y_n$ to represent our "data", it is understood that $Y_1, Y_2, \cdots, Y_n$ may denote a set of differenced observations, a set of transformed responses (e.g., log-transformed, etc.), or possibly both.
- We will discuss three estimation techniques: method of moments, least squares including conditional least squares and unconditional least squares, and maximum likelihood.

# 7.2.1. Method of moments

The method of moments (MOM) approach to estimation consists of the following two parts:

- equating sample moments to the corresponding population (theoretical) moments and
- solving the resulting system of equations for any unknown parameters.

# MOM applied to AR(1) models

- Consider the stationary AR(1) model

$$Y_t = \phi Y_{t-1} + e_t.$$

- In this model, there are two parameters, namely, $\phi$ and $\sigma_e^2$.
- The MOM estimator of $\phi$ is obtained by setting the population lag one autocorrelation $\rho_1$ equal to the sample lag one autocorrelation $r_1$, and solving for $\phi$, that is,

$$\rho_1 \stackrel{\text{set}}{=} r_1.$$

- For this model, of course, $\rho_1 = \phi$ (see Chapter 4). Therefore, the MOM estimator of $\phi$ is

$$\widehat{\phi} = r_1.$$

# MOM applied to AR(2) models

- For the AR(2) model,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t,$$

there are three parameters, namely, $\phi_1$, $\phi_2$, and $\sigma_e^2$.

- To find the MOM estimators of $\phi_1$ and $\phi_2$, recall the Yule-Walker equations

$$\rho_1 = \phi_1 + \rho_1\phi_2, \quad \rho_2 = \rho_1\phi_1 + \phi_2.$$

- Setting $\rho_1 = r_1$ and $\rho_2 = r_2$, we have

$$r_1 = \phi_1 + r_1\phi_2, \quad r_2 = r_1\phi_1 + \phi_2.$$

- Solving this system for $\phi_1$ and $\phi_2$ produces the MOM estimators

$$\widehat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}, \quad \widehat{\phi}_2 = \frac{(r_2 - r_1^2)}{1-r_1^2}.$$

# MOM applied to AR(p) models

- For the general AR($p$) model,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t,$$

there are $p+1$ parameters, namely, $\phi_1, \phi_2, \cdots, \phi_p$, and $\sigma_e^2$.

- The Yule-Walker equations from Chapter 4, namely, is

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \cdots + \phi_p\rho_{p-1} \\ &\cdots \\ \rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \cdots + \phi_p. \end{aligned}$$

- Set $\rho_1 = r_1$, $\rho_2 = r_2$, $\cdots$, $\rho_p = r_p$ to obtain

$$\begin{aligned} r_1 &= \phi_1 + \phi_2 r_1 + \phi_3 r_2 + \cdots + \phi_p r_{p-1} \\ &\cdots \\ r_p &= \phi_1 r_{p-1} + \phi_2 r_{p-2} + \phi_3 r_{p-3} + \cdots + \phi_p. \end{aligned}$$

- The MOM estimators $\widehat{\phi}_1, \widehat{\phi}_2, \cdots, \widehat{\phi}_p$ solve this system of equations.

# Remarks

- It should be noted that the MOM approach may produce estimates $\widehat{\phi}_1, \widehat{\phi}_2, \cdots, \widehat{\phi}_p$ that fall "outside" the stationarity region, even if the process is truly stationary!

- That is, the estimated AR($p$) polynomial

$$\widehat{\phi}_{\mathsf{MOM}}(x) = 1 - \widehat{\phi}_1 x - \widehat{\phi}_2 x^2 - \cdots - \widehat{\phi}_p x^p$$

  may possess roots which do not exceed 1 in absolute value (or modulus).

# MOM applied to MA(1) models

- Consider the invertible MA(1) process

$$Y_t = e_t - \theta e_{t-1}.$$

- In this model, there are two parameters, namely, $\theta$ and $\sigma_e^2$.
- To find the MOM estimator of $\theta$, we solve

$$\rho_1 = \frac{-\theta}{1+\theta^2} \stackrel{\text{set}}{=} r_1 \text{ for } \theta.$$

- If $|r_1| > 0.5$, then no real solutions for $\theta$ exist.
- If $|r_1| = 0.5$, then the solutions for $\theta$ are $\pm 1$, which correspond to an MA(1) model that is not invertible.
- If $|r_1| < 0.5$, the invertible solution for $\theta$ is (only one)

$$\widehat{\theta} = \frac{-1 + \sqrt{1 - 4r_1^2}}{2r_1}.$$

# MOM applied to MA(q) models

- For the MA($q$) case, we are left to solve the highly nonlinear system

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2} \overset{\text{set}}{=} r_k, \;\; k = 1, 2, \cdots, q-1$$

$$\rho_q = \frac{-\theta_q}{1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2} \overset{\text{set}}{=} r_q,$$

  for $\theta_1, \theta_2, \cdot, \theta_q$.

- Just as in the MA(1) case, there will likely be multiple solutions, only of which at most one will correspond to a fitted invertible model.
- For this and other reasons, the MOM approach is not recommended for use with moving average processes.

# MOM applied to ARMA(1,1)

- Consider the ARMA(1,1) process

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1},$$

  where $\{e_t\} \sim (0, \sigma_e^2)$.
- There are three parameters, namely, $\phi$, $\theta$ and $\sigma_e^2$.
- Recall from Chapter 4 that

$$\rho_k = \left[ \frac{(1-\theta\phi)(\phi-\theta)}{1-2\theta\phi+\theta^2} \right] \phi^{k-1},$$

  so that $\rho_2 = \rho_1 \phi$.
- Therefore, the MOM estimator of $\phi$ is given by

$$\widehat{\phi} = \frac{r_2}{r_1},$$

  and the MOM estimator of $\theta$ solves

$$r_1 = \frac{(1-\theta\widehat{\phi})(\widehat{\phi}-\theta)}{1-2\theta\widehat{\phi}+\theta^2}.$$

# Estimation for white noise variance

- We wish to estimate the variance $\sigma_e^2$ associated with the white noise process $\{e_t\}$.
- To do this, for any stationary ARMA model, the (process) variance of $Y_t$, $\gamma_0 = \text{var}(Y_t)$, can be estimated by the sample variance

$$S^2 = \frac{1}{n-1}\sum_{t=1}^{n}(Y_t - \overline{Y})^2.$$

- This estimate will be needed to estimate $\sigma_e^2$, as we see now.

# Estimation for white noise variance in AR(p)

- For the AR($p$) model

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t,$$

we recall from Chapter 4 that

$$\gamma_0 = \frac{\sigma_e^2}{1 - \sum_{i=1}^{p} \phi_i \rho_i} \Longrightarrow \sigma_e^2 = \left(1 - \sum_{i=1}^{p} \phi_i \rho_i\right) \gamma_0.$$

- Therefore, the MOM estimator of $\sigma_e^2$ is obtained by substituting in $\widehat{\phi}_k$ for $\phi_k$ and $r_k$ for $\rho_k$, for $k = 1, 2, \cdots, p$, and $S^2$ for $\gamma_0$. We obtain

$$\widehat{\sigma}_e^2 = \left(1 - \widehat{\phi}_1 r_1 - \cdots - \widehat{\phi}_p r_p\right) S^2.$$

# Estimation for white noise variance in MA(q)

- For a general MA($q$) process,

$$Y_t = e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q},$$

- we recall from Chapter 4 that

$$\gamma_0 = \left(1 + \theta_1^2 + \cdots + \theta_q^2\right) \sigma_e^2 \Longrightarrow \sigma_e^2 = \frac{\gamma_0}{1 + \theta_1^2 + \cdots + \theta_q^2}.$$

- Therefore, the MOM estimator of $\sigma_e^2$ is obtained by substituting in $\widehat{\theta}_k$ for $\theta_k$, for $k = 1, 2, \cdots, q$, and $S^2$ for $\gamma_0$. We obtain

$$\widehat{\sigma}_e^2 = \frac{S^2}{1 + \widehat{\theta}_1^2 + \widehat{\theta}_2^2 + \cdots + \widehat{\theta}_q^2}.$$

# Estimation for white noise variance in ARMA(1,1)

- For an ARMA(1,1) process,

$$\gamma_0 = \left(\frac{1-2\phi\theta+\theta^2}{1-\phi^2}\right)\sigma_e^2 \Longrightarrow \sigma_e^2 = \left(\frac{1-\phi^2}{1-2\phi\theta+\theta^2}\right)\gamma_0.$$

- The MOM estimator of $\sigma_e^2$ is obtained by substituting in $\widehat{\theta}$ for $\theta$, $\widehat{\phi}$ for $\phi$ and $S^2$ for $\gamma_0$. We obtain

$$\widehat{\sigma}_e^2 = \left(\frac{1-\widehat{\phi}^2}{1-2\widehat{\phi}\widehat{\theta}+\widehat{\theta}^2}\right)S^2.$$

# Remarks

- The MOM approach to estimation in stationary ARMA models is not always satisfactory.
- In fact, some would recommend (including authors of CC!) to avoid MOM estimation in any model with moving average components.
- We therefore consider other approaches, starting with conditional least squares (CLS).

# CLS applied to AR(1)

- Consider the first order autoregressive model

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t,$$

  where a nonzero mean $\mu = \mathsf{E}(Y_t)$ has been added for flexibility.

- For this model, the conditional sum of squares function is

$$S_C(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2$$

  with a sample $Y_1, Y_2, \cdots, Y_n$. It is worth noting that there are $n-1$ items instead of $n$ items in the sum $S_C(\phi, \mu)$.

- The principle of (conditional) least squares says to choose the values of $\phi$ and $\mu$ that make $S_C(\phi, \mu)$ as small as possible.

# CLS applied to AR(1)

- For this AR(1) model, this amounts to solving

$$\frac{\partial S_C(\hat{\phi}, \hat{\mu})}{\partial \phi} \stackrel{\text{set}}{=} 0$$

$$\frac{\partial S_C(\hat{\phi}, \hat{\mu})}{\partial \mu} \stackrel{\text{set}}{=} 0$$

for $\hat{\phi}$ and $\hat{\mu}$.

# CLS applied to AR(1)

- Note that

$$\frac{\partial S_C(\hat{\phi}, \hat{\mu})}{\partial \mu} = \sum_{i=2}^{n} 2[(Y_t - \hat{\mu}) - \hat{\phi}(Y_{t-1} - \hat{\mu})](-1 + \hat{\phi}) = 0.$$

- We have

$$\hat{\mu} = \frac{1}{(n-1)(1-\hat{\phi})} \left[ \sum_{t=2}^{n} Y_t - \hat{\phi} \sum_{t=2}^{n} Y_{t-1} \right] \approx \bar{Y}.$$

# CLS applied to AR(1)

- Note that

$$\frac{\partial S_C(\hat{\phi}, \hat{\mu})}{\partial \phi} = \sum_{i=2}^{n} 2[(Y_t - \bar{Y}) - \hat{\phi}(Y_{t-1} - \bar{Y})](Y_t - \bar{Y}) = 0.$$

- The conditional least squares estimators are taken to be

$$\widehat{\phi} = \frac{\sum_{t=2}^{n}(Y_t - \overline{Y})(Y_{t-1} - \overline{Y})}{\sum_{t=2}^{n}(Y_t - \overline{Y})^2}.$$

# Remarks for CLS to AR(1)

- The CLS estimator $\widehat{\phi}$ is approximately equal to $r_1$, the lag one sample autocorrelation (the only difference is that the denominator does not include the $t = 1$ term). We would expect the difference between the $\widehat{\phi}$ and $r_1$ (the MOM estimator) to be negligible when the sample size $n$ is large.

- The CLS estimator $\widehat{\mu}$ is only approximately equal to the sample mean $\overline{Y}$, but the approximation should be adequate when the sample size $n$ is large.

# CLS applied to AR(p)

- Consider the p-order autoregressive model

$$Y_t - \mu = \sum_{i=1}^{p} \phi_i (Y_{t-i} - \mu) + e_t,$$

where a nonzero mean $\mu = \mathrm{E}(Y_t)$ has been added for flexibility.

- For AR($p$) model, the conditional sum of squares function is

$$S_C(\phi_1, \cdots, \phi_p, \mu) = \sum_{t=p+1}^{n} \left[ (Y_t - \mu) - \sum_{i=1}^{p} \phi_i (Y_{t-i} - \mu) \right]^2,$$

a function of $p+1$ parameters, with the sample $Y_1, \cdots, Y_p$, $Y_{p+1}, \cdots, Y_n$. It is worth noting that there are $n-p$ items rather than $n$ items in the sum $S_C(\phi_1, \cdots, \phi_p, \mu)$.

- Despite being more complex, $\phi_1, \phi_2, \cdots, \phi_p$ and $\mu$ are chosen so that $S_C(\phi_1, \phi_2, \cdots, \phi_p, \mu)$ is minimized.

# CLS applied to AR(p)

- For this AR(p) model, this amounts to solving

$$\frac{\partial}{\partial \phi_i} S_C(\hat{\phi}_1, \cdots, \hat{\phi}_p, \hat{\mu}) =^{\text{set}} 0$$
$$\frac{\partial}{\partial \mu} S_C(\hat{\phi}_1, \cdots, \hat{\phi}_p, \hat{\mu}) =^{\text{set}} 0$$

for $\hat{\phi}_1, \cdots, \hat{\phi}_p$ and $\hat{\mu}$.

# CLS applied to AR(1)

- Note that

$$\frac{\partial S_C}{\partial \mu} = \sum_{i=p=1}^{n} 2[(Y_t - \hat{\mu}) - \sum_{i=1}^{p} \hat{\phi}_i (Y_{t-i} - \hat{\mu})](-1 + \sum_{i=1}^{p} \hat{\phi}_i) = 0.$$

- We have

$$\hat{\mu} = \frac{1}{(n-p)(1 - \sum_{i=1}^{p} \hat{\phi}_i)} \left[ \sum_{t=p+1}^{n} Y_t - \sum_{i=1}^{p} \hat{\phi}_i \sum_{t=p+1}^{n} Y_{t-i} \right] \approx \bar{Y},$$

an approximation when $n$ is large (i.e., much larger than $p$).

# CLS applied to AR(p)

- Note that, for $i = 1, \cdots, p$,

$$\frac{\partial S_C}{\partial \phi_i} = \sum_{i=p+1}^{n} 2[(Y_t - \bar{Y}) - \sum_{i=1}^{n} \phi_i (Y_{t-i} - \bar{Y})](\bar{Y} - Y_{t-1}) = 0.$$

- The conditional least squares estimators for $\phi_1, \phi_2, \cdots, \phi_p$ are well approximated by the solutions to the sample Yule-Walker equations

$$\begin{aligned} r_1 &= \phi_1 + \phi_2 r_1 + \phi_3 r_2 + \cdots + \phi_p r_{p-1} \\ &\cdots \\ r_p &= \phi_1 r_{p-1} + \phi_2 r_{p-2} + \phi_3 r_{p-3} + \cdots + \phi_p. \end{aligned}$$

- Therefore, in stationary AR models, the MOM and conditional least squares estimates should be approximately equal.

# CLS applied to MA(1)

- Consider the zero mean MA(1) model

$$Y_t = e_t - \theta e_{t-1}.$$

- Invertible by assumption, recall from Chapter 4 that we can rewrite the MA(1) as an infinite-order AR model

$$Y_t = \underbrace{-\theta Y_{t-1} - \theta^2 Y_{t-2} - \theta^3 Y_{t-3} - \cdots}_{\text{"AR}(\infty)\text{"}} + e_t.$$

- In theory, we are left to find the value of $\theta$ which minimizes

$$S_C(\theta) = \sum \left(Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \cdots\right)^2.$$

- Unfortunately, this is not a practical exercise, because we have only an observed sample $Y_1, Y_2, \cdots, Y_n$ to do the estimation.

# CLS applied to MA(1) (continued)

- Take $e_0 \equiv 0$, rewrite the MA(1) model as

$$e_t = Y_t + \theta e_{t-1}.$$

- We have, conditional on $e_0 \equiv 0$,

$$e_1 = Y_1, \quad e_2 = Y_2 + \theta e_1, \quad \cdots \quad , e_n = Y_n + \theta e_{n-1}.$$

- We can find the value of $\theta$ that minimizes

$$S_C(\theta) = \sum_{t=1}^{n} e_t^2 = \sum_{i=1}^{n} (Y_t + \theta Y_{t-1} + \cdots + \theta^{t-2} Y_2 + \theta^{t-1} Y_1)^2.$$

- This minimization problem can be carried out numerically, e.g., Gauss-Newton or Nelder-Mead, searching over a grid of $\theta$ values in $(-1, 1)$ and selecting the value that produces the smallest possible $S_C(\theta)$.

# CLS applied to MA(q)

- The technique just described for MA(1) estimation via CLS can be carried out for MA($q$) models in the same fashion.
- When $q > 1$, the problem becomes to find the values of $\theta_1, \theta_2, \cdots, \theta_q$ such that

$$S_C(\theta_1, \cdots, \theta_q) = \sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} (Y_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q})^2 ,$$

  is minimized, subject to conditions $e_0 = \cdots = e_{-q} = 0$.
- This can be performed using a grid search (a multivariate numerical method) over all possible values of $\theta_1, \theta_2, \cdots, \theta_q$ which yield an invertible solution.

# CLS applied to ARMA(1,1)

- Consider the zero mean ARMA(1,1) process

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}.$$

- Rewrite the model as

$$e_t = Y_t - \phi Y_{t-1} + \theta e_{t-1},$$

  with the goal of minimizing

$$S_C(\phi, \theta) = \sum_{t=1}^{n} e_t^2.$$

- Two "startup" problems appear, specifying values for $e_0$,$Y_0$.
- CC recommend avoiding specifying $Y_0$ and minimizing

$$S_C^*(\phi, \theta) = \sum_{t=2}^{n} e_t^2.$$

# CLS applied to ARMA(p,q)

- Consider the zero mean ARMA(p,q) process

$$e_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}$$

  with $e_p = \cdots = e_{p+1-q} = 0$.
- Similar modification is recommended for ARMA models when $p > 1$ and/or when $q > 1$.
- Minimizing

$$S_C(\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q) = \sum_{t=p+1}^{n} e_t^2.$$

# Estimation for white noise variance

- For the AR($p$),

$$\widehat{\sigma}_e^2 = \left(1 - \widehat{\phi}_1 r_1 - \widehat{\phi}_2 r_2 - \cdots - \widehat{\phi}_p r_p\right) S^2.$$

- For the MA($q$),

$$\widehat{\sigma}_e^2 = \frac{S^2}{1 + \widehat{\theta}_1^2 + \widehat{\theta}_2^2 + \cdots + \widehat{\theta}_q^2}.$$

- For an ARMA(1,1),

$$\widehat{\sigma}_e^2 = \left(\frac{1 - \widehat{\phi}^2}{1 - 2\widehat{\phi}\widehat{\theta} + \widehat{\theta}^2}\right) S^2.$$

- Nothing changes with our formulae for the white noise variance estimates. The only difference is that now CLS estimates for the $\phi$'s and $\theta$'s can be used in place of MOM estimates.

# Example 7.1

## Example

We use SAS to simulate data from four autoregressive (AR) processes

- AR(1) with $\phi = 0.7$
- AR(2) with $\phi_1 = 0.7$ and $\phi_2 = -0.5$
- AR(3) with $\phi_1 = 0.7$, $\phi_2 = -0.5$ and $\phi_3 = 0.4$
- AR(4) with $\phi_1 = 0.7$, $\phi_2 = -0.5$, $\phi_3 = 0.4$ and $\phi_4 = -0.2$.

In each case, $n = 200$ and the white noise distribution is $N(0, 1)$ so that $\sigma_e^2 = 1$.

# Figure 7.1. Four AR simulations with $n = 200$ and $\sigma_e^2 = 1$

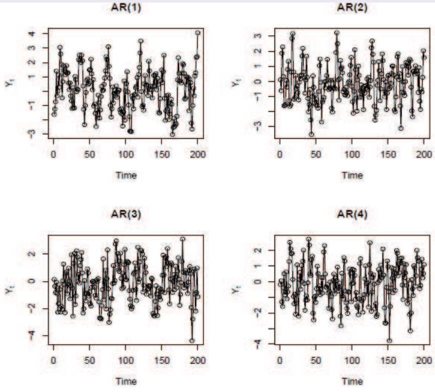



Figure 7.1: *Four autoregressive simulations with* $n = 200$ *and* $\sigma_e^2 = 1$.

# CLS for Example 7.1

| Model | $\widehat{\phi}_1$ | $\widehat{\phi}_2$ | $\widehat{\phi}_3$ | $\widehat{\phi}_4$ | $\widehat{\sigma}_e^2$ |
|---|---|---|---|---|---|
| AR(1) | 0.57248 | | | | 1.02983 |
| AR(2) | 0.59821 | -0.47593 | | | 1.01215 |
| AR(3) | 0.56770 | -0.51881 | 0.36404 | | 1.015139 |
| AR(4) | 0.58088 | -0.51412 | 0.35331 | -0.19080 | 1.037916 |
| True | $\phi_1 = .7$ | $\phi_2 = -.5$ | $\phi_3 = .4$ | $\phi_4 = -.2$ | $\sigma_e^2 = 1$ |

# Discussion of Example 7.1

- The four simulated series are depicted in Figure 7.1, and the CLS estimates from the simulations are given in the table above (along with the true values of $\phi_1$, $\phi_2$, $\phi_3$, and $\phi_4$).
- The estimates for all parameters are generally on target although they are different from the true values.
- You can use ML method and ULS method to repeat the above estimates. Check their AIC or BIC.

# 7.4. Maximum likelihood estimation

- The method of maximum likelihood is the most commonly-used technique to estimate model parameters (not just in time series, but in other applications, too).
- Parameter estimates are based on the observed sample $Y_1, Y_2, \cdots, Y_n$, so we don't have to worry about "start up" values (which could be of great consequence in small samples).
- The main disadvantage is that we have to specify a joint probability distribution for the random variables in the sample.
- This makes the method more mathematical.

# Likelihood function

- The likelihood function $L$ is a function that describes the joint distribution of $Y_1, Y_2, \cdots, Y_n$.
- However, it is viewed as a function of the model parameters with the "data" $Y_1, Y_2, \cdots, Y_n$ being fixed.
- From this perspective, if we maximize the likelihood function as a function of the model parameters, we are finding the values of the parameters (i.e., the estimates) that are most consistent with the observed data.

# MLE applied to AR(1)

- Consider the AR(1) model

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t,$$

where $e_t \sim N(0, \sigma_e^2)$ and $\mu = \mathsf{E}(Y_t)$ is the overall mean.

- There are three parameters in this model: $\phi$, $\mu$ and $\sigma_e^2$.
- The probability density function (pdf) of $e_t$ is

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

for all $-\infty < x_t < \infty$.

# MLE applied to AR(1)

- The joint pdf of $e_2, e_3, \cdots, e_n$ is given by

$$f(x_2, \cdots, x_n) = \prod_{t=2}^{n} f(x_t) = \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi}\sigma_e} \text{exp}\left(-x_t^2/2\sigma_e^2\right)$$

$$= (2\pi\sigma_e^2)^{-(n-1)/2} \text{exp}\left(-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n} x_t^2\right).$$

# MLE applied to AR(1) (continued)

- To write out the joint pdf of $Y = (Y_1, Y_2, \cdots, Y_n)$, we perform a multivariate transformation

$$\begin{aligned} Y_2 &= \mu + \phi(Y_1 - \mu) + e_2 \\ &\cdots \\ Y_n &= \mu + \phi(Y_{n-1} - \mu) + e_n, \end{aligned}$$

  with $Y_1 = y_1$ (fixed) to obtain the (conditional) joint distribution of $Y_2, Y_3, \cdots, Y_n$, given $Y_1 = y_1$.

- Applying the laws of conditioning, the joint pdf of $Y$, and, hence, the likelihood function $L \equiv L\left(\phi, \mu, \sigma_e^2 | y\right)$, is given by

$$L\left(\phi, \mu, \sigma_e^2 | y\right) = f\left(y_2, y_3, \cdots, y_n | y_1\right) f\left(y_1\right).$$

# MLE applied to AR(1) (continued)

- The details on pp 159 (CC) show that

$$f(y_2, \cdots, y_n | y_1) \propto \exp\left\{ -\frac{1}{2\sigma_e^2} \sum_{t=2}^{n} [(y_t - \mu) - \phi(y_{t-1} - \mu)]^2 \right\}$$

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

- Multiplying these pdfs and simplifying, we get

$$L(\phi, \mu, \sigma_e^2 | y) = \left(2\pi\sigma_e^2\right)^{-n/2} (1-\phi^2)^{1/2} \exp\left[ -\frac{S(\phi, \mu)}{2\sigma_e^2} \right],$$

where

$$S(\phi, \mu) = (1-\phi^2)(y_1 - \mu)^2 + \sum_{t=2}^{n} [(y_t - \mu) - \phi(y_{t-1} - \mu)]^2.$$

# Result for MLE of AR(1) model

- For this AR(1) model, the maximum likelihood estimators (MLEs) of $\phi$, $\mu$ and $\sigma_e^2$ are the values $\hat{\phi}$, $\hat{\mu}$ and $\hat{\sigma}_e^2$ such that

$$L(\hat{\phi}, \hat{\mu}, \hat{\sigma}_e^2 | y) = \max_{\phi,\mu,\sigma^2} L(\phi, \mu, \sigma_e^2 | y).$$

- In practice, MLEs are found numerically.

# Result for MLE of AR(1) model

- The function $S(\phi, \mu)$ is called the unconditional sum-of-squares function.

- Note that when $S(\phi, \mu)$ is viewed as random,

$$S(\phi, \mu) = S_C(\phi, \mu) + (1 - \phi^2)(Y_1 - \mu),$$

where $S_C(\phi, \mu)$ is the conditional sum of squares function defined in Section 7.3 (notes).

# CLS vs ULS

- The conditional least squares (CLS) estimates of $\phi$ and $\mu$ are found by minimizing $S_C(\phi, \mu)$.
- The unconditional least squares (ULS) estimates of $\phi$ and $\mu$ are found by minimizing $S(\phi, \mu)$.
- ULS estimation is regarded as a "compromise" between CLS estimation and the method of maximum likelihood. We will not pursue the ULS approach.

# Other ARMA models

- Conceptually, the approach to finding maximum likelihood estimates in any stationary ARMA($p, q$) model is the same as the approach that we have just outlined in the special AR(1) case.
- However, as you might suspect, the likelihood function $L$ becomes notably more complex.
- Because we will use software to maximize $L$, this turns out not to be a big deal.

# Comparisons among CLS, ML and ULS for simulated AR(4)

- The comparison among the CLS, ML, ULS and OLS estimates for simulated series of AR(4) with $(.7, -.5, .4, -.2)$ is given below.

| Method | $\hat{\phi}_1$ | $\hat{\phi}_2$ | $\hat{\phi}_1$ | $\hat{\phi}_2$ | $\widehat{\sigma}_e^2$ | BIC |
|---|---|---|---|---|---|---|
| CLS | .58086 | -.5141 | .35327 | -.19079 | 1.0379 | 592.1711 |
| ML | .58883 | -.5169 | .36182 | -.20512 | 1.0309 | 591.5099 |
| ULS | .59192 | -.5222 | .36742 | -.20949 | 1.0308 | 591.5170 |
| OLS | .58066 | -.5136 | .35217 | -.19028 | 1.4211 | 637.8657 |

  - The estimates are very close.
  - The estimates of ML are between those of CLS and ULS.

# Comparisons among YW, ITYW, ML and ULS for Lake Huron data

- The comparison among the YW, ITYW, ML and ULS estimates for Lake Huron data is given below.

| Method | intercept | $\hat{\beta}$ | $\hat{\phi}_1$ | $\hat{\phi}_2$ | AIC | $\widehat{\sigma}_e^2$ |
|---|---|---|---|---|---|---|
| YW | 50.8914 | -0.0218 | 0.9714 | -0.2754 | 210.543 | 0.47716 |
| ITYW | 50.8178 | -0.0217 | 0.9714 | -0.2754 | 210.495 | 0.47686 |
| ML | 50.5109 | -0.0216 | 1.0048 | -0.2913 | 210.397 | 0.47605 |
| ULS | 50.4111 | -0.0215 | 1.0153 | -0.2974 | 210.409 | 0.47599 |

  - The estimates are very close.

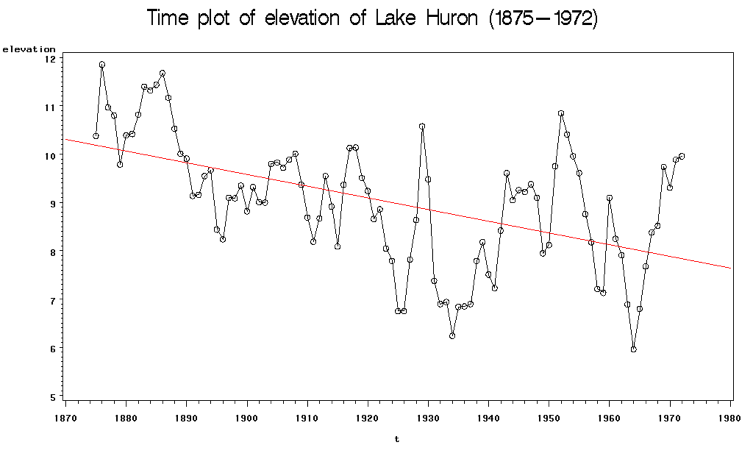
# Example 7.2. The Lake Huron Data

## Example

Lake Huron is the second largest of the Great Lakes, including Lake Superior, Lake Huron, Lake Michigan, Lake Erie and Lake Ontario.

Lake Huron has 59,596 square km, 332 km long and 295 km at its greatest width. Lake Huron sits between Ontario, Canada and Michigan.

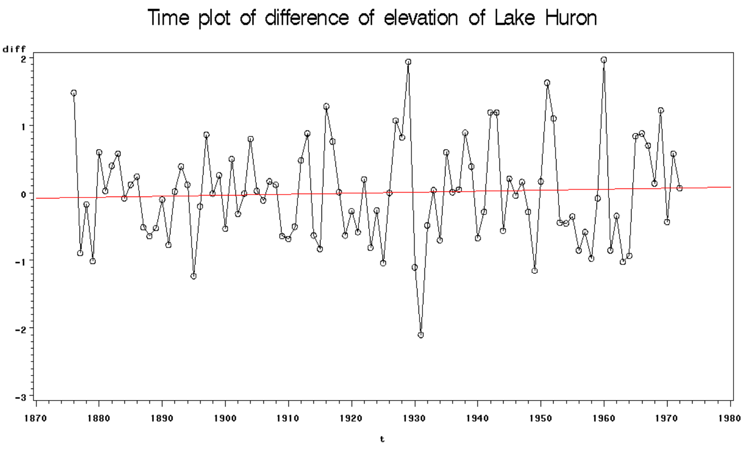
Figure 7.2 displays the average July water surface elevation (measured in feet from 1875-1972) for Harbor Beach, Michigan, on Lake Huron, Station 5014.


Time plot of elevation of Lake Huron (1875–1972)
elevation
12
11
10
9
8
7
6
5
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
t

# Example 7.2. The Lake Huron Data

- Use proc arima procedure to identify the stationarity and p, d, q parameters. The data does not pass the ADF unit root test. See sas program example72LakeHuron .
- By Box-Cox transformation, it is unnecessary to take a transformation for the data. The time plot of difference of data does not suggest the nonconstant variance, see Figure 7.2-1.
- Therefore, the nonstationarity comes from the unit root, implying that differencing the data is needed.

# Figure 7.2-1. The time plot of difference of Lake Huron data



# Example 7.2. The Lake Huron Data

- For data $\Delta Y_t$, it pases ADF root test. ESACF (maybe the sample ACF) suggests a random walk model

$$Y_t = Y_{t-1} + e_t$$

with $\sigma_e^2 = 0.555$ and $BIC = 218.2158$. CLS and ML get same results. A problem comes from failing to pass the test for white noise of residuals.

# Example 7.2. Conclusions for the Lake Huron data

- BIC and SCAN suggest AR(2) model. Using CLS method yields the fitted AR(2) process

$$\Delta Y_t = 0.15939\Delta Y_{t-1} - 0.20933\Delta Y_{t-2} + e_t$$

is suggested to be appropriate with $\sigma^2_{CLS} = 0.533$ and $BIC = 220.886$. $\phi_1$ fails to pass T-test.

- Using ML method (see Section 7.4) yields the fitted AR(2) process

$$\Delta Y_t = 0.17262\Delta Y_{t-1} - 0.22337\Delta Y_{t-2} + e_t$$

is suggested to be appropriate with $\sigma^2_{ML} = 0.5297$ and $BIC = 221.34$. $\phi_1$ fails to pass T-test.

# Example 7.2. Conclusions for the Lake Huron data

- If we try Proc autoreg procedure, for data $Y_t$, we find that AR(2) process with a trend

$$Y_t = 50.5109 - 0.0216t + X_t$$

and

$$X_t = 1.0048X_{t-1} - 0.2913X_{t-2} + e_t,$$

where $\sigma^2 = 0.47605$. This fitted model seems to be best due to $BIC = 220.7364$ and passing all tests. Note that $X_t$ is stationarity.

# 7.4.1. Large sample properties for a stationary ARMA$(p, q)$

- Consider a stationary ARMA$(p, q)$ process

$$\phi(B)Y_t = \theta(B)e_t,$$

where the AR and MA characteristic operators are

$$\begin{aligned}\phi(B) &= \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)\\ \theta(B) &= \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right),\end{aligned}$$

and where $\{e_t\}$ is a normal zero mean white noise process.

# Asymptotic normality

- Maximum likelihood estimators $\widehat{\phi}_j$ and $\widehat{\theta}_k$ satisfy

$$\sqrt{n}\left(\widehat{\phi}_j - \phi_j\right) \longrightarrow_{approx} N(0, \sigma^2_{\widehat{\phi}_j}), \quad \text{for} \quad j = 1, 2, \cdots, p,$$

and

$$\sqrt{n}\left(\widehat{\theta}_k - \theta_k\right) \longrightarrow_{approx} N(0, \sigma^2_{\widehat{\theta}_k}), \quad \text{for} \quad k = 1, 2, \cdots, q,$$

respectively, as $n \to \infty$.

# Asymptotic normality

- In other words, for large $n$,

$$\widehat{\phi}_j \sim AN\left(\phi_j, \sigma^2_{\widehat{\phi}_j}/n\right)$$
$$\widehat{\theta}_k \sim AN\left(\theta_k, \sigma^2_{\widehat{\theta}_k}/n\right),$$

for all $j = 1, 2, \cdots, p$ and $k = 1, 2, \cdots, q$.

- That is, maximum likelihood estimators are approximately unbiased and normally distributed in large samples.

# Asymptotic normality for special cases

- AR(1).

$$\widehat{\phi} \sim AN\left(\phi, \frac{1-\phi^2}{n}\right)$$

- AR(2).

$$\widehat{\phi}_1 \sim AN\left(\phi_1, \frac{1-\phi_2^2}{n}\right)$$

$$\widehat{\phi}_2 \sim AN\left(\phi_2, \frac{1-\phi_2^2}{n}\right)$$

- MA(1).

$$\widehat{\theta} \sim AN\left(\theta, \frac{1-\theta^2}{n}\right)$$

# Asymptotic normality for special cases (continued)

- MA(2).

$$\widehat{\theta}_1 \sim AN\left(\theta_1, \frac{1-\theta_2^2}{n}\right)$$
$$\widehat{\theta}_2 \sim AN\left(\theta_2, \frac{1-\theta_2^2}{n}\right)$$

- ARMA(1,1).

$$\widehat{\phi} \sim AN\left(\phi, \frac{c(\phi,\theta)(1-\phi_2^2)}{n}\right)$$
$$\widehat{\theta} \sim AN\left(\theta, \frac{c(\phi,\theta)(1-\theta_2^2)}{n}\right),$$

where $c(\phi,\theta) = [(1-\phi\theta)/(\phi-\theta)]^2$.

# Remarks for asymptotic normality

- In multi-parameter models; e.g., AR(2), MA(2), ARMA(1,1), etc., the MLEs are (asymptotically) correlated.

- Furthermore, this correlation can be large even when $n$ is large. See pp 161 (CC) for a further description.

# Large sample confidence intervals

- These results make getting large sample confidence intervals for model parameters easy.
- For example, a $100(1-\alpha)$ percent (approximate) confidence interval for $\phi$ in an AR(1) model is given by

$$\widehat{\phi} \pm z_{\alpha/2}\sqrt{\frac{1-\widehat{\phi}^2}{n}}.$$

- A $100(1-\alpha)$ percent (approximate) confidence interval for $\theta$ in an MA(1) model is given by

$$\widehat{\theta} \pm z_{\alpha/2}\sqrt{\frac{1-\widehat{\theta}^2}{n}}.$$

# Asymptotic normality for other methods' estimators

- In words, the form is

  "point extimate $\pm z_{\alpha/2}$ (estimated standard error)."

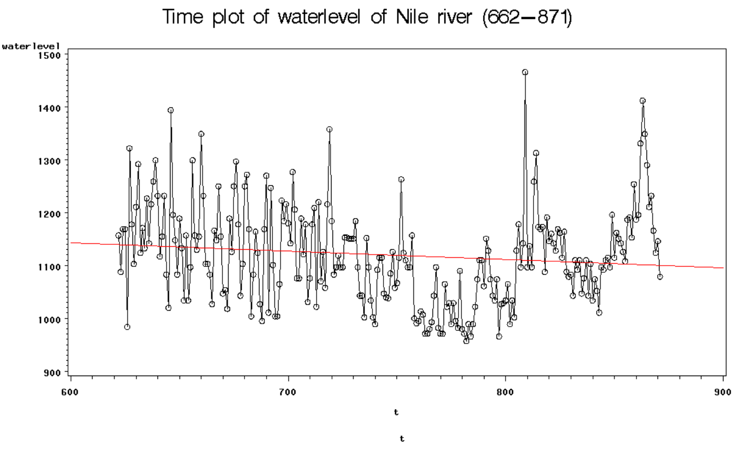
- Approximate confidence intervals for the other ARMA model parameters are computed in exactly the same way.
- Although it is perhaps not obvious, MLEs and estimates from CLS and ULS have the same large-sample distributions.
- On the other hand, large sample distributions of MOM estimates can be quite different for purely MA models (although they are the same for purely AR models). See pp 162 (CC).

# Example 7.3. Annual minimum water level data of the Nile

**Example**

The Nile is a major north-flowing river in North Africa, generally regarded as the longest river in the world. It is 6650 km long.

The data file NILE.TSM consists of the annual minimum water levels of the Nile river as measured at the Roda gauge near Cairo for the years 622-871. These values are plotted in Figure 7.3.


Time plot of waterlevel of Nile river (662—871)
waterlevel
1500
1400
1300
1200
1100
1000
900
600
700
800
900
t
t

# Example 7.3. The Nile river Data

- Use proc arima procedure to identify the stationarity and p, d, q parameters. See sas code example73Nile.
- The rather slow decay of the sample ACF suggests the possibility of a fractionally intergrated model for the mean-corrected series $Y_t - \mu$.
- ADF unit root test suggests that unit root hypothesis can not be rejected.
- The first order difference data passes the ADF test and the ACF suggests that MA(2) model

$$Y_t - Y_{t-1} = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

is appropriate.

# Example 7.3. The Nile river Data

- CLS, ULS and ML estimates of parameters in MA(2) are

| Method | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\widehat{\sigma}_e^2$ | BIC |
|---|---|---|---|---|
| CLS | 0.61931 | 0.21194 | 6127.574 | 2888.076 |
| ML | 0.61647 | 0.20943 | 6127.58 | 2888.103 |
| ULS | 0.61929 | 0.21256 | 6127.245 | 2888.119 |

- The fitted model is

$$Y_t = Y_{t-1} + e_t - 0.61931e_{t-1} - 0.21194e_{t-2},$$

an IMA(1,2) model with $BIC = 2888.076$ and $\hat{\sigma} = 78.27882$,
$95\%$ CI for $\theta_1$ is $0.61931 \pm 2(0.06239)$ and
$95\%$ CI for $\theta_2$ is $0.21194 \pm 2(0.06242)$.

# Example 7.3 (continued)

- Using proc autoreg procedure suggests

$$Y_t = 1119 + X_t$$

$$X_t = 0.3855X_{t-1} + 0.1323X_{t-3} + 0.1672X_{t-4} + e_t$$

with $BIC = 2905.043$ and $\hat{\sigma} = 77.78962$.

# More about Nile data

- Consider Box-Cox transformation. The transformations $1/Y_t$ is suggested , see sas code example73_Nile_transf.

- Taking transformation $1/Y_t$, and using proc arima procedure suggests

$$1/Y_t - 1/Y_{t-1} = (1 - 0.60508B - 0.20424B^2)e_t$$

  with $BIC = -4129.85$ and $\hat{\sigma} = 0.000059$.

- Taking transformation $1/Y_t$, and using proc autoreg procedure suggests

$$\frac{1}{Y_t} = 0.000899 + X_t$$

$$X_t = 0.4021X_{t-1} + 0.1388X_{t-3} + 0.1665X_{t-4} + e_t$$

  with $BIC = -4140.4736$ and $\hat{\sigma} = 0.0000591$.

# More about Nile data

- In practice, the transformations $log(Y_t)$ is suggested , see sas code example73_Nile_transf.

- Using proc arima procedure suggests

$$\log(Y_t) - \log(Y_{t-1}) = (1 - 0.61151B - 0.20867B^2)e_t$$

with $AIC = -630.763$ ($BIC = -623.528$) and $\hat{\sigma} = 0.067806$.

- Using proc autoreg procedure suggests

$$\log(Y_t) = 7.0174 + X_t$$

$$X_t = 0.3942X_{t-1} + 0.1363X_{t-3} + 0.1667X_{t-4} + e_t,$$

with $AIC = -634.45551$ ($BIC = -620.37$) and $\hat{\sigma} = 0.00454$, being nearly one-17240th of estimated variance of the untransformed IMA(1,2) model. Finally, it is preferred.

# Example 7.4. Crude oil monthly spot price data

**Example**

Crude oil prices behave much as any other commodity with wide price swings in times of shortage or oversupply. The crude oil price cycle may extend over several years responding to changes in demand as well as OPEC and non-OPEC supply.

The data in Figure 7.4 are monthly spot prices for crude oil (measured in U.S. dollars per barrel) from Cushing, OK, from 1/1986 to 1/2006.

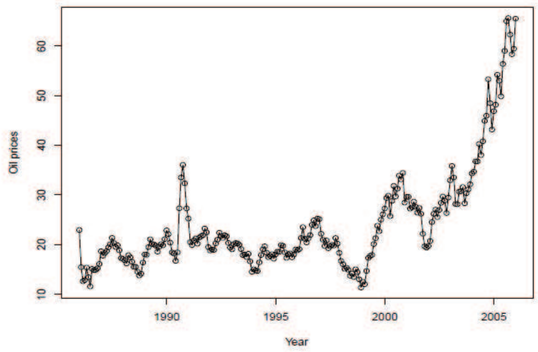



Figure 7.4: *Crude oil price data. Monthly spot prices in dollars from Cushing, OK, from 1/1986 to 1/2006.*

# Analysis for Example 7.4

- The plot of the data suggests that this process is not stationary. This is confirmed by the augmented Dickey-Fuller (ADF) test. The ADF test does not reject the null hypothesis of a unit root.
- Consider the first order difference for the data. $\nabla Y_t$ appears to be stationary in the mean, so no further differencing is warranted.
- ACF suggests that MA(4) are candidate models, further examination recommends to consider the models

$$Y_t - Y_{t-1} = e_t - \theta_1 e_{t-1} - \theta_4 e_{t-4}$$

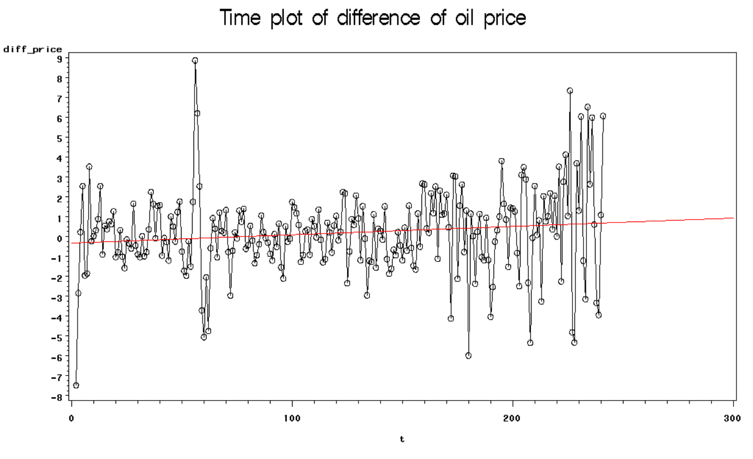
- But, residuals of the fitted model IMA(1,1) fail to pass test for white noise series. If we use proc autoreg procedure via ML, residuals of the fitted model also fail to pass test for white noise series.

# Analysis for Example 7.4

- We are worry about the ACF of the differenced data. Therefore we further examine the first differences $\nabla Y_t$, whose time plot is depicted in Figure 7.5. There is a pronounced increasing variance pattern towards the end of the series.

- To address the nonconstant variance problem, we examine the first differences of the logarithms, $\nabla \log Y_t$. Remember that the log transformation is applied before taking differences.

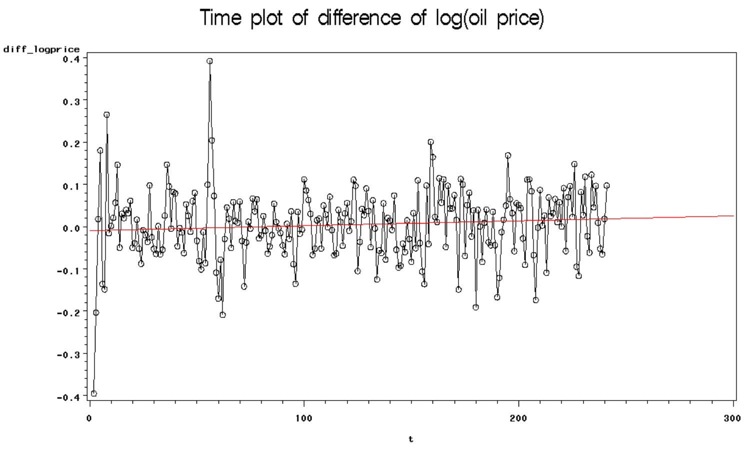
# Figure 7.5. Time plot of difference of oil price



# Analysis for Example 7.4

- Taking logarithm transformation to price comes from the background of economics rather than Box-Cox transformation. If taking Box-Cox transformation to price, transformation $1/\sqrt{Y_t}$ is suggested, see sas code example74Oilprice_transf.
- The transformation $1/\sqrt{Y_t}$ is not easily explained.
- The processes $\nabla \log Y_t$ is depicted in Figure 7.6, see sas code example74Oilprice_log.

# Figure 7.6. Time plot of difference of log(oil price)



# Analysis for Example 7.4 (continued)

- Taking logarithms helps to reduce the nonconstant variance. $\nabla \log Y_t$ process is stationary by ADF unit root test. And the sample ACF, ESACF and SCAN for the $\nabla \log Y_t$ process resemble that of an MA(1).

- We therefore (tentatively) adopt

$$\nabla \log Y_t = e_t - \theta e_{t-1},$$

an IMA(1,1) model for $\log Y_t$.

# Analysis for Example 7.4 (continued)

- For logprice data, using proc arima procedure via ML to get the fitted IMA(1,1) model

$$\nabla \log Y_t = \log(Y_t) - \log(Y_{t-1}) = e_t + 0.29372 e_{t-1}.$$

with AIC$= -518.934$ ($BIC = -515.453$) and $\hat{\sigma} = 0.006707$.

- For logprice data, using proc autoreg procedure via ML to get the fitted model

$$\nabla \log Y_t = 3.1029 - 0.006567t + 0.0000418t^2 + X_t$$

with

$$X_t = 1.1832 X_{t-1} - 0.2768 X_{t-2} + e_t.$$

with AIC$= -524.18$ ($BIC = -506.75606$) and $\hat{\sigma} = 0.00646$.

# Remarks for Example 7.4

- It is careful for us to use the quadratic item except the cases where the quadratic item is believed to exist.

- The log-transformed IMA(1,1) is preferred.

## chapter 7 is over

# Thank you for your attention!