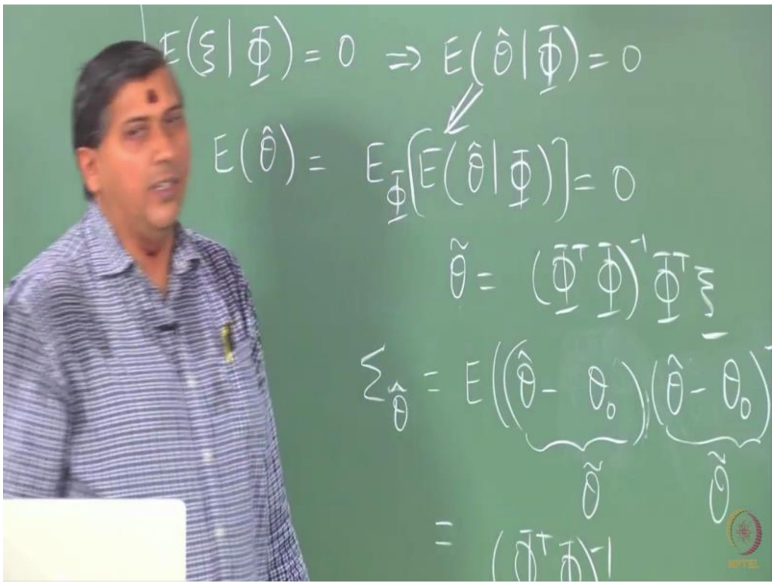
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 101
Lecture 44A - Estimation Methods 1 -8

Yesterday we looked at the bias in parameter estimates resulting from least squares. And we discussed two cases that case of the deterministic phi and then the stochastic phi.

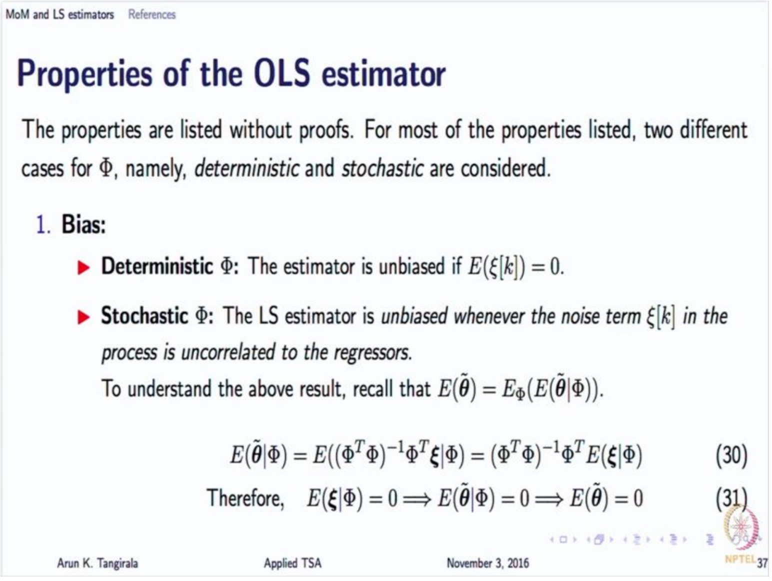
(Refer Slide Time: 00:28)


$E(\xi | \Phi) = 0 \Rightarrow E(\hat{\theta} | \Phi) = 0$
$E(\hat{\theta}) = E_{\Phi}[E(\hat{\theta} | \Phi)] = 0$
$\tilde{\theta} = (\Phi^T \Phi)^{-1} \Phi^T \xi$
$\Sigma_{\hat{\theta}} = E((\underbrace{\hat{\theta} - \theta_0}_{\tilde{\theta}})(\underbrace{\hat{\theta} - \theta_0}_{\tilde{\theta}})$
$= (\Phi^T \Phi)^{-1}$

And the stochastic is we show that if the left over or the residuals are uncorrelated; strictly speaking the conditional expectation should be 0, but that also implies that they should be uncorrelated if you recall from the properties. Then we can guarantee that the conditional expectation of theta hat is 0. Now the idea is that if this condition expectation is 0 then one could show using iterative expectation if you recall the iterative expectation essentially says expectation of x is expectation of the conditional expectation.

But, remember here the inner one is across z or you can say across theta hat and the outer one is across phi. If the inner one is 0 then you will get the total expectation to be 0 not the other way around necessarily. So, in this way we are ensuring that theta hat is unbias. So, this is not necessarily a two way condition, but if you guarantee that the regressors are uncorrelated with the residuals or whatever you have left out then you can obtain unbias estimates, and we have gone through a couple of demonstrations yesterday.

(Refer Slide Time: 02:05)



MoM and LS estimators References

Properties of the OLS estimator

The properties are listed without proofs. For most of the properties listed, two different cases for Φ, namely, *deterministic* and *stochastic* are considered.

1. **Bias:**
   - **Deterministic Φ:** The estimator is unbiased if $E(\xi[k]) = 0$.
   - **Stochastic Φ:** The LS estimator is *unbiased whenever the noise term $\xi[k]$ in the process is uncorrelated to the regressors.*
     To understand the above result, recall that $E(\tilde{\theta}) = E_{\Phi}(E(\tilde{\theta}|\Phi))$.

$$E(\tilde{\boldsymbol{\theta}}|\Phi) = E((\Phi^T\Phi)^{-1}\Phi^T\boldsymbol{\xi}|\Phi) = (\Phi^T\Phi)^{-1}\Phi^T E(\boldsymbol{\xi}|\Phi) \qquad (30)$$

$$\text{Therefore,} \quad E(\boldsymbol{\xi}|\Phi) = 0 \Longrightarrow E(\tilde{\boldsymbol{\theta}}|\Phi) = 0 \Longrightarrow E(\tilde{\boldsymbol{\theta}}) = 0 \qquad (31)$$

Arun K. Tangirala Applied TSA November 3, 2016 NPTEL 37

So, this is also going to be the case for yesterday we discussed the case of fitting an AR 1 model to an AR 2 process. The bias can also arise for example if you are fitting an AR 2 model to an ARMA model. Although your auto regressive orders may match the moving average component will render the regressors correlated with residuals. And that example is also included in the marked down file, I will also upload the mark down file that I demonstrated yesterday.

So, you will have a chance to play around with the ARMA example as well, where I simulate an ARMA 2 1 and show that even though I fit an AR 2, so the orders of the auto regressive components match, but because I leave out and m a component and remember the m a component would have at least e k and e k minus 1 and e k minus 1 would be correlated with the regressors v k minus 1. So, as a result you would get bias estimates even in those situations.

In plain terms, you should make sure that the residuals are white that is all; if you do not understand any of this that is important. So, now we move on to the calculation of variance of theta hat which is very important. We know for two reasons: one it gives us an idea whether we obtain efficient estimates; that is minimum variance estimates, and two with the help of the distribution of theta hat we will be able to construct confidence regions. So, this has got to do now with the precision of theta hat. Again, the starting point is your theta tilde. This is your starting point, but now unlike in the bias case the expression for the variance is the bit more complicated. Remember the variance theta hat is general p by 1 vector; therefore we cannot talk of variance we talk of covariance matrix and that is what is denoted as sigma theta tilde.

You can also says sigma theta tilde is nothing but sigma theta hat. What is sigma theta hat expectation of? Theta hat minus expectation of theta hat times theta hat minus expectation of theta hat transpose. We use this expression because theta hat is a p by 1 vector. If theta hat was a scalar we would have simply said expectation of theta hat minus mu theta hat to the whole square. So now, assuming that theta hat is an unbiased estimator, assuming that we have ensured that is true by ensuring whiteness of residuals we can replace the expectation of theta hats with the respective true values. But the starting point is always expression that I wrote before. That is the definition of sigma theta hat.

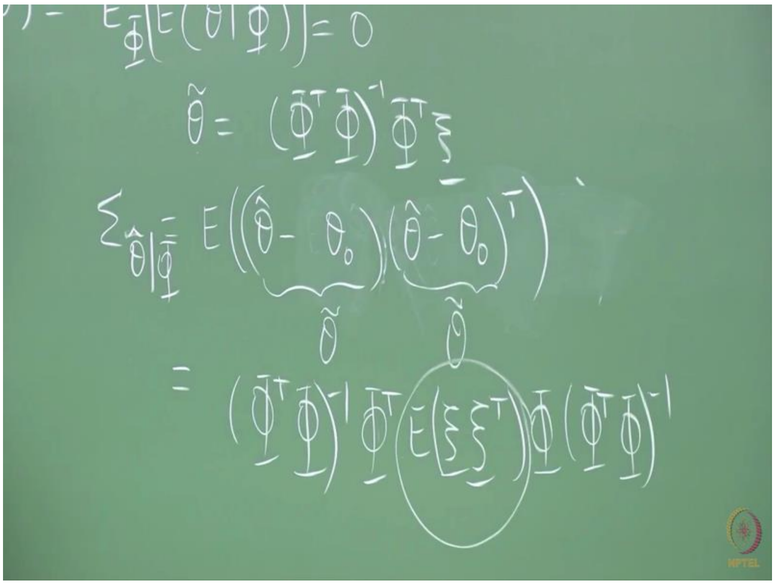
Now, this is sigma theta hat, once again when phi is deterministic things are fairly easy. In a sense easy, in the sense the sigma theta hat only depends on the properties of the equation error that is the error in your data. But when phi is stochastic then evaluating in these expectations becomes very difficult. Notice that this is your theta tilde.

(Refer Slide Time: 05:50)

Therefore, as in the case of bias we always; now we go back to the conditional covariance. So, we say fix phi for a given set of regressors what is the variation in theta hat that I see. And remember although it fix the regresors the equation errors are the errors in your data are going to change and that is why we expect to see a variation in theta hat.

So now, this brings us back more or less to the deterministic case, because we are fixing phi then evaluating this expression becomes very easy relatively, you can say, you can debate whether it is very easy or not. But relatively it becomes much simpler because all you have to do is now plug in this expression here and notice that when you take the transpose things get flit; the products get flit by the matrix property. As a result you get this long looking expression you may thing this all theory its useless no it is not; all the variance calculations that the numbers that are being reported to you by AR dot ols or lmn and so on, the rest on this fundamental result.

(Refer Slide Time: 07:00)



$$) = E_{\Phi}[E(\theta|\Phi)] = 0$$

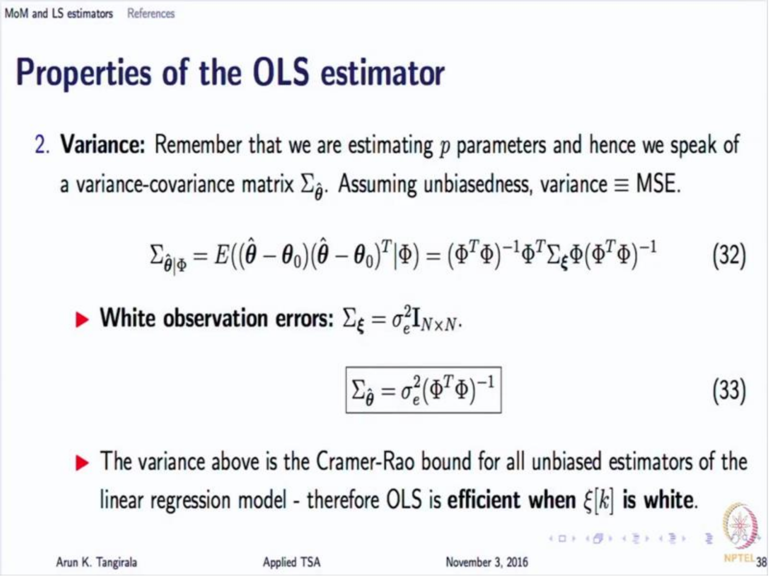
$$\tilde{\theta} = (\Phi^T\Phi)^{-1}\Phi^T\xi$$

$$\Sigma_{\hat{\theta}|\Phi} = E((\underbrace{\hat{\theta} - \theta_0}_{\tilde{\theta}})(\underbrace{\hat{\theta} - \theta_0}_{\tilde{\theta}})^T)$$

$$= (\Phi^T\Phi)^{-1}\Phi^T E(\xi\xi^T)\Phi(\Phi^T\Phi)^{-1}$$

So, now that we are going to evaluate the conditional covariance phi becomes fixed of deterministic and I can take the expectation past this phi is with the result that I have this expression. So now everything depends on this quantity here, what is that quantity? Expectation of z times z transpose I do not know how many of your able to see, but you can also look at the slide.

(Refer Slide Time: 07:39)



MoM and LS estimators References

Properties of the OLS estimator

2. **Variance:** Remember that we are estimating p parameters and hence we speak of a variance-covariance matrix $\Sigma_{\hat{\theta}}$. Assuming unbiasedness, variance ≡ MSE.

$$\Sigma_{\hat{\theta}|\Phi} = E((\hat{\theta}-\theta_0)(\hat{\theta}-\theta_0)^T|\Phi) = (\Phi^T\Phi)^{-1}\Phi^T\Sigma_{\xi}\Phi(\Phi^T\Phi)^{-1} \qquad (32)$$

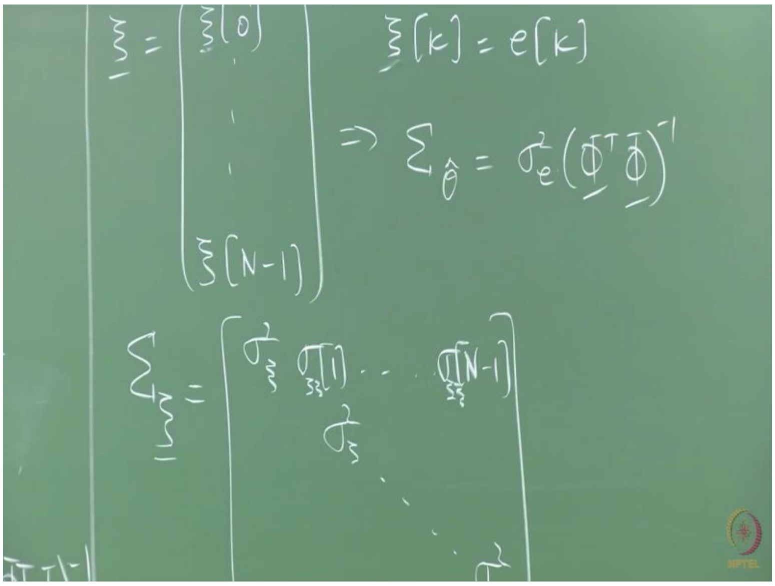
- **White observation errors:** $\Sigma_{\xi} = \sigma_e^2 \mathbf{I}_{N\times N}$.

$$\boxed{\Sigma_{\hat{\theta}} = \sigma_e^2(\Phi^T\Phi)^{-1}} \qquad (33)$$

- The variance above is the Cramer-Rao bound for all unbiased estimators of the linear regression model - therefore OLS is **efficient when** $\xi[k]$ **is white.**

Arun K. Tangirala Applied TSA November 3, 2016 38

(Refer Slide Time: 07:50)

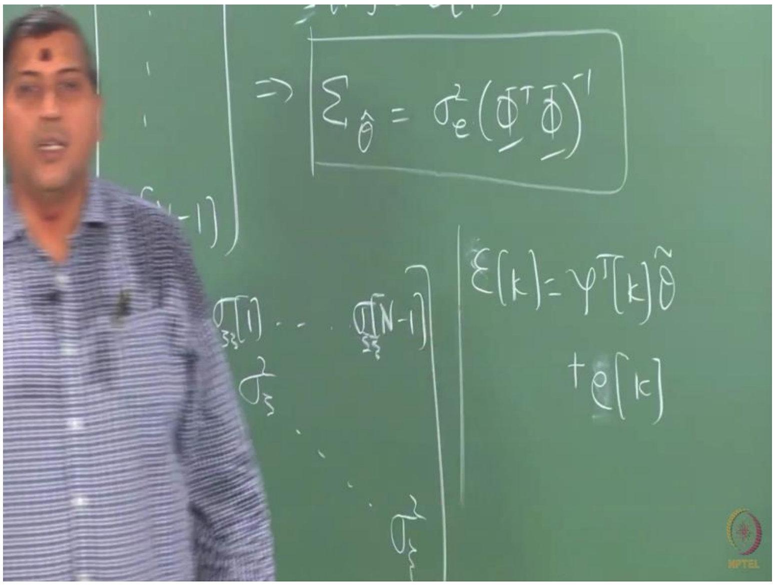


$\underline{\xi} = \begin{pmatrix} \xi[0] \\ \vdots \\ \xi[N-1] \end{pmatrix}$
$\xi[k] = e[k]$
$\Rightarrow \Sigma_{\hat{\theta}} = \sigma_e^2(\underline{\Phi}^T\underline{\Phi})^{-1}$
$\Sigma_{\underline{\xi}} = \begin{bmatrix} \sigma_\xi^2 & \sigma_{\xi\xi}[1] & \cdots & \sigma_{\xi\xi}[N-1] \\ & \sigma_\xi^2 & & \\ & & \ddots & \\ & & & \sigma^2 \end{bmatrix}$

So, it that expectation of z times z transpose, what is your z? It is a vector of errors. First you should recognize that this is a vector of errors starting from time 0 to n minus 1. Therefore, this expectation of z z transpose would be the variance covariance of the vector of errors; it is like your vector of random variables, at each instant z is a random variable. So, if you are thorough with the theory of random variables then all of this should be easy to follow.

So, here you have z as a vector, and therefore the expectation of z z transpose is nothing but your variance covariance matrix of this z. What will sigma z contain? What will the diagonals of sigma z contain let us say? Variance of z which need not be constant unless z is a stationary signal is a stationary noise.

(Refer Slide Time: 08:58)


$\Rightarrow \Sigma_{\hat{\theta}} = \sigma_e^2 (\Phi^T \Phi)^{-1}$
$\varepsilon[k] = \varphi^T[k]\tilde{\theta} + e[k]$

If z is stationary then the diagonals of sigma z, so this is your sigma z diagonals of sigma z in for the stationary case would contain the variance of the noise. And what about the half diagonal terms, auto covariance; it will it would contain the auto covariance’s at the respective lag. So, here would be sigma z at lag 1 and up to sigma at n minus 1 and it is going to be a symmetric matrix it is a positive definite matrix. Now, I am sorry.

Student: (Refer Time: 09:37).

Because z can be correlated in time; no I never said it white, it is just some error. When it is white then things are light, otherwise it becomes heavy; so now evaluating this expression become simpler when the equation errors are white. Now, always remember this z that we had introduced through the data generating process equation should always be interpreted, can best interpret as the residuals of your model. So, what we have been assuming is that structurally the model and the process are identical. Therefore, we are treating z as the errors in the data, but strictly speaking your z are the residuals from your model in practice.

Now if we guarantee that z is white, I cannot guarantee through the data generating process, but at least through my modelling I can guarantee. Then things become a lot simpler the off diagonal turns out be 0 and then sigma z becomes a diagonal matrix and therefore the sigma theta hat; so when z is white then straight away I can write sigma theta hat as sigma square e times phi transpose phi inverse, which is such a simple expression. And this is the expression that you will find in almost all initial what you say text on least squares, because this is the case that gives you some nice results that you can right by hand. Beyond that you have to calculate if z is coloured, but then the question is for coloured z of interest was. And it turns out that it is not so much of interest because, if the residuals are coloured or if your equation errors are coloured then least squares gives you in efficient estimates.

We have already seen that see coloured z does not mean that there is going to be a correlation between the residuals and regressors that need not be true. For example I can have a completely uncorrelated set of regressors and residuals, but still have coloured z. So, I can obtain unbiased estimates even when z is coloured, do not assume that whenever z is coloured it becomes correlated with the regressors; need not be at all. It all depends on your model, your application and so on.

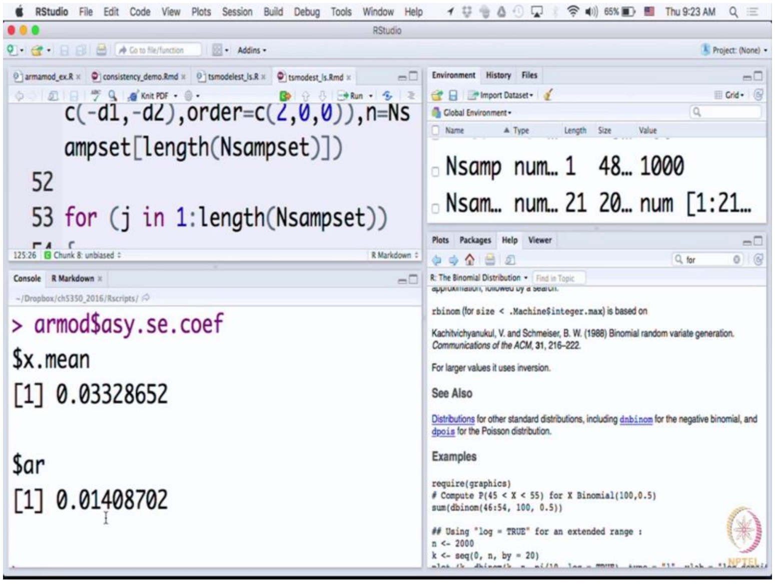
So, imagine a situation where the regressors are uncorrelated with the residuals, but the residuals are coloured. So, it is like you are fitting some simple linear model your using a sensor and the senor is actually producing coloured noise, it is not producing white noise. In such cases it turns out that least squares gives us inefficient estimates. What does inefficiency mean? That means there exists another estimator which can give me estimates with lower error. So, how do I know up front? I do not know. Sometimes I do know; sometimes if I have the luxury of performing experiments only on the senor not perturbing the process at all I just take the sensor I analyse the noise characteristics of the sensor and I figure out that the sensor gives me correlated noise coloured noise then its tells me up front I should be careful in using least squares method for estimating the parameters.

So what do I do? We will talk about that a bit later. Somehow, we should ensure that the residuals are white either through by modelling or some other operation. Suppose I guarantee, suppose I ensure that the residuals are white by a suitable modelling and that happens own that can happen either when you choose a proper regressors set and or that you have to model the residuals as well. So, to summarize z can be coloured either because of your sensor characteristics what nothing to do with a choice of regressors.

Like for example, if you take an fitting an AR model to an ARMA process, what happens? Suppose the AR component you have gotten it right process is ARMA 2 1 and the model is AR 2 So, the AR components have been whitely modelled. However, what you have left out is a coloured noise. So, there z is coloured, but additionally z is also correlated with the regressors. In such cases you will not only get inefficient estimates but also biased estimates.

On the other hand, there are situations where the noise in the data is coloured, but not correlated necessarily with the regressors. Then you get biased estimates, unbiased estimates, but you will get inefficient estimates. So, what you do? Either you turn to waited least squares which will discuss shortly or you turn to MLA. Somewhere you have to take remedies, what this result says is that first of all when the residuals or equation errors are white you can calculate sigma theta hat. So the expression, the numbers that you saw in yesterdays demo used these calculations; this result to calculate the variance of parameter estimates.

(Refer Slide Time: 15:33)



```
c(-d1,-d2),order=c(2,0,0)),n=Ns
ampset[length(Nsampset)])

for (j in 1:length(Nsampset))
```

```
> armod$asy.se.coef
$x.mean
[1] 0.03328652

$ar
[1] 0.01408702
```

Name	Type	Length	Size	Value
Nsamp	num..	1	48...	1000
Nsam...	num..	21	20...	num [1:21...

rbinom (for size < .Machine$integer.max) is based on

Kachitvichyanukul, V. and Schmeiser, B. W. (1988) Binomial random variate generation. *Communications of the ACM*, 31, 216–222.

For larger values it uses inversion.

See Also

Distributions for other standard distributions, including dnbinom for the negative binomial, and dpois for the Poisson distribution.

Examples

```
require(graphics)
# Compute P(45 < X < 55) for X Binomial(100,0.5)
sum(dbinom(46:54, 100, 0.5))

## Using "log = TRUE" for an extended range :
n <- 2000
k <- seq(0, n, by = 20)
```

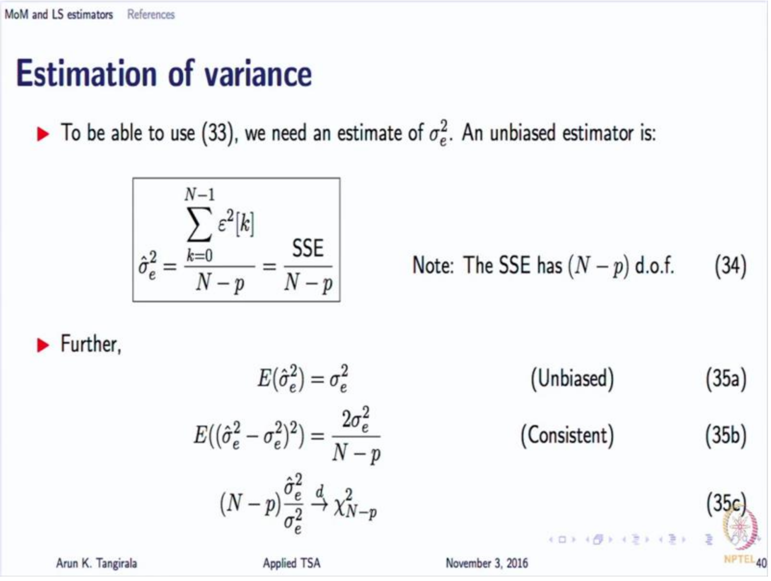
Remember your AR dot OLS if you know, so if you recall yesterday we had fit you can pick any of the AR models and you can see in one of the attributes of this model you have asymptotic, this is asymptotic dot standard error dot coef. That means, it has used the large sample expressions for calculating standard error and of each coefficients. So, this is what it is reporting. Yesterday remember we had fit and AR 1 model with the intercept term 1. So, ignore the top 1. So, this is the standard error that it has calculated. This calculation here is based on the result that you see here. But you must ask a question now, how does it calculate sigma theta hat? Do you know phi? In any of this linear modelling do we know phi or not? Yes or no, you should know it right otherwise there is no modelling.

What we do not know is sigma square e. We do not know sigma square e, I do not know the variance of the noise of the residuals or equation errors whatever you want to call. So, we need a method to estimate sigma square e. Remember we have always said in time series modelling as well the goal is not just to estimate model parameters, but also to estimate sigma square e. So, I have to estimate sigma square e from data and the way I do it is I turn to the residuals epsilon the prediction errors let me use a term prediction errors now or the approximation errors.

Yesterday I made this point that the approximation error contains two components: one the noise parts itself and then a contribution due to the difference between the truth and estimate. But that is going to be small that component is going to be small, so if you recall yesterdays expression where written this is the approximation error is psi transpose k times theta tilde plus you are noise term and now this is white, so will replace this with white noise.

In order to get an estimate of sigma square e we turn to the prediction, because this is the only accessible instrument that I have; the variable that I have. And it turns out fortunately that I can obtain an unbiased estimate of sigma square e by calculating this expression here sum square error by n minus p;

(Refer Slide Time: 18:17)



MoM and LS estimators References

Estimation of variance

- To be able to use (33), we need an estimate of σ_e^2. An unbiased estimator is:

$$\boxed{\hat{\sigma}_e^2 = \frac{\sum_{k=0}^{N-1} \varepsilon^2[k]}{N-p} = \frac{\text{SSE}}{N-p}} \qquad \text{Note: The SSE has } (N-p) \text{ d.o.f.} \qquad (34)$$

- Further,

$$E(\hat{\sigma}_e^2) = \sigma_e^2 \qquad \text{(Unbiased)} \qquad (35a)$$

$$E((\hat{\sigma}_e^2 - \sigma_e^2)^2) = \frac{2\sigma_e^2}{N-p} \qquad \text{(Consistent)} \qquad (35b)$$

$$(N-p)\frac{\hat{\sigma}_e^2}{\sigma_e^2} \xrightarrow{d} \chi^2_{N-p} \qquad (35c)$$

Arun K. Tangirala Applied TSA November 3, 2016 40

Why do we have n minus p here, because we have lost p degrees of freedom in estimating the p parameter, although there are n terms in this expression. If you recall yesterday it said if your estimate in AR 1, sorry the first residual is not available. Likewise, if you are estimating p parameters in the time series model the first p terms are not available. So, you can look at it in different ways the net effect is that the sum square error has n minus p degrees of freedom. And degree of freedom has got to do with the number of independent sources of randomness or in linear sense also uncorrelated sources of randomness.

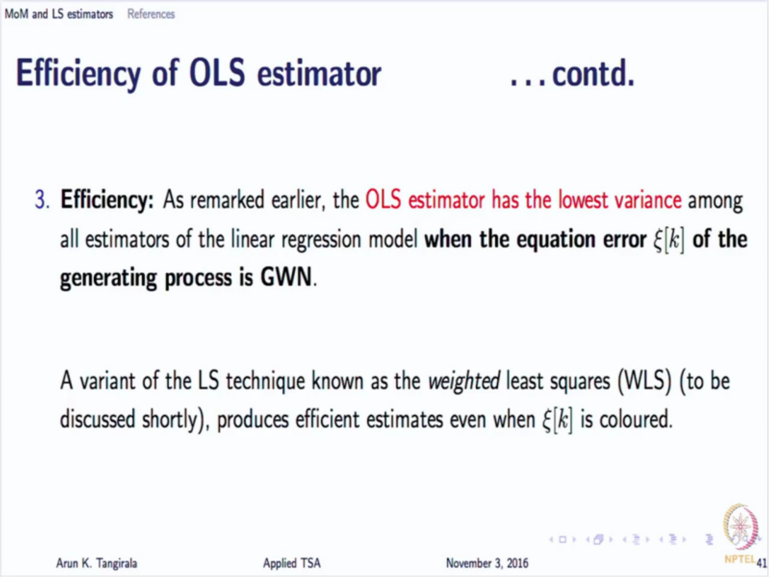
So, this expression now in conjunction with the equation 3 is used for calculating sigma theta hat. Once you get sigma theta hat then you have the diagonals and half diagonals with you, you can use this diagonal terms to calculate the respective standard errors. Although sigma theta hat is not necessarily a diagonal matrix we somehow tend to ignore the half diagonal terms, so we only calculate approximate standard errors.

And what you should do is for the AR 1 modelling I have also made that note in the marked down file, whatever calculations a the AR dot OLS is making for you in computing the standard errors you should cross check with the theoretical expression. So, for the AR 1 model that we fate or AR 2 model that we feet yesterday run the mark down file it will report the object will report also the method will report the standard errors keep that a side make your hand calculations, and calculations in with the help of r separately compute this standard errors using first computing sigma square e. So, the first step is to fit the model, second step is to compute the prediction errors, third step is to compute the sigma square e hat, and then the final step is a sigma theta hat. You know your phi, and c if it matches with what the expression gives you.

And of course, as I said this expression gives you an unbiased estimate of sigma square e it also is a consistent estimate of sigma square e. That means as n goes to infinity this estimate here will recover sigma square e for you, the true sigma square e. And of course, we will ignore the last result. It basically says that this estimate here sigma square e hat follows the chi square distribution, which we do not use at this moment it is actually used in calculating the f statistics. Remember you are l m when we looked at the output of the l m there was an f statics that was reported and that f statics is the ratio of two variances; one variance is that of the y hat. That is contribution due to the predictions and other variance is due to the errors.

So, this tells you that the sigma square e hat is the chi square distribution. We will not proceed further with that.

(Refer Slide Time: 21:42)



MoM and LS estimators References

Efficiency of OLS estimator ...contd.

3. **Efficiency:** As remarked earlier, the OLS estimator has the lowest variance among all estimators of the linear regression model **when the equation error $\xi[k]$ of the generating process is GWN**.

A variant of the LS technique known as the *weighted* least squares (WLS) (to be discussed shortly), produces efficient estimates even when $\xi[k]$ is coloured.

Arun K. Tangirala Applied TSA November 3, 2016 41

Now we have talked about OLS efficiency I just want to emphasis again; that if the OLS estimator has the lowest variance among all estimators in the linear regression of the linear regression model when the equation error is not only white but also Gaussian. This is also known as Gauss-Markov theorem and there are many names to this result, but this is a very important result in parameter estimation; that your OLS gives you most efficient estimates when the equation error is Gaussian white noise. And you have to keep asking yourself how do I know apriori it is Gaussian white noise I do not necessarily know.

One way to check is after you fit your model you check if it is white and if it has a Gaussian distribution. How would you check if it has a Gaussian distribution, the residuals? Just look the histogram or do a even more advance q q plot are you conduct hypothesis test on the distribution. Then only you should be convinced that you are working with the most efficient estimates. The weighted least squares method that will shortly discuss. We will overcome some of the limitations of this least squares, these requirements for efficiency.