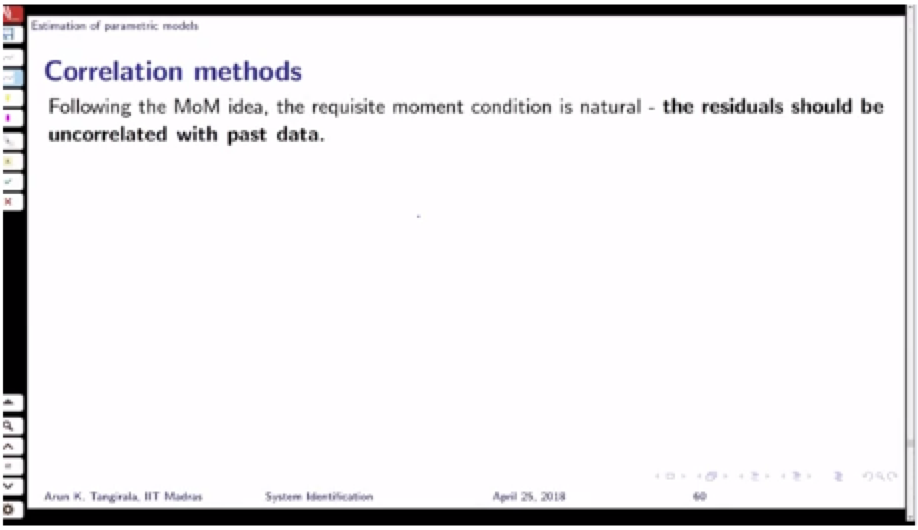
CH5230: System Identification

State-Space/Subspace Identification

Part 1

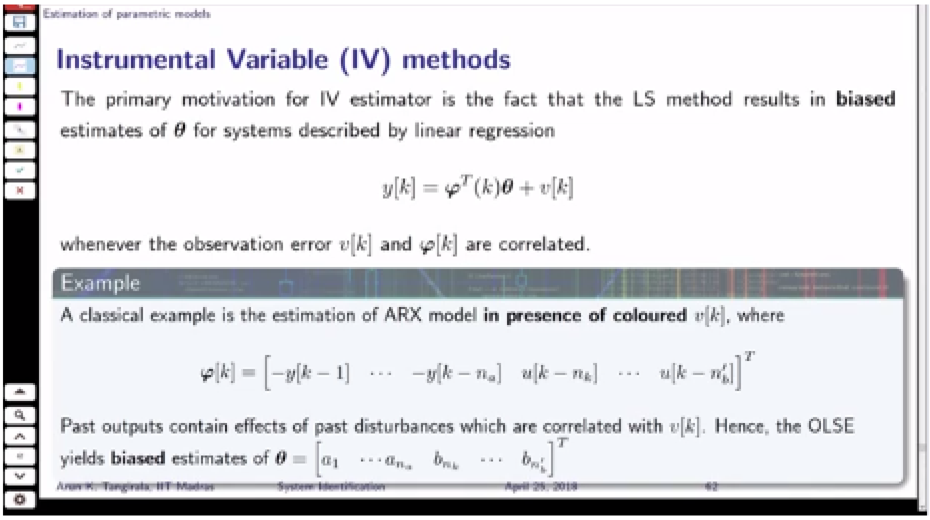
(Refer Slide Time: 00:18)



So, let's move on quickly to the second class of methods, namely the correlation methods, where the idea is to estimate parameters by imposing the condition that the residual should be uncorrelated with past data. So here we are not solving any optimization problem. It's a different approach like in a method of moments we do. So if you now state the problem formally, by first denoting the past data up to k minus 1 as soon as z k minus 1. And let zeta k be some function of this past data, for example, it could be a predictor, it could be any function, it could be predicted also. Then the way you want to estimate theta is by imposing these conditions, right? So this is the solution to this problem here. We have this equation, zeta is what you construct, that's your choice. And epsilon as usual is a prediction error.

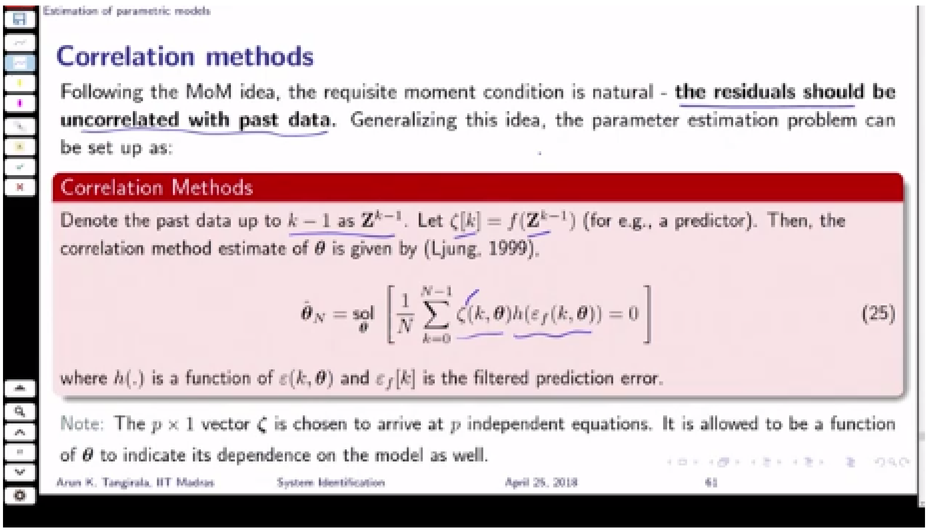
We want the residuals or that we want the function of the residuals to be uncorrelated with the function of the past data. So although we set out by saying the residual should be uncorrelated with past data, we are now imposing a more general condition, we are saying let the function, any function of residuals are some chosen function of residuals be uncorrelated with some chosen function of the past data. Typically, we choose h to be 1, that means just we work with the residuals and we choose zeta accordingly. And remember if there are p parameters I have to set up p such equations, that means I have to find p such zetas, okay? So, that is the idea. Now the instrumental variable method that we talked about belongs to this class of methods. You can now think of zetas as some kind of instruments also.

(Refer Slide Time: 02:18)



So, let's go back to this linear regression case. And if you look at IV methods, it may not necessarily start from this equation that we wrote here, it typically starts with this motivation that the least squares method results in biased estimates, whenever this v case are correlated with the Psi there.

(Refer Slide Time: 02:27)

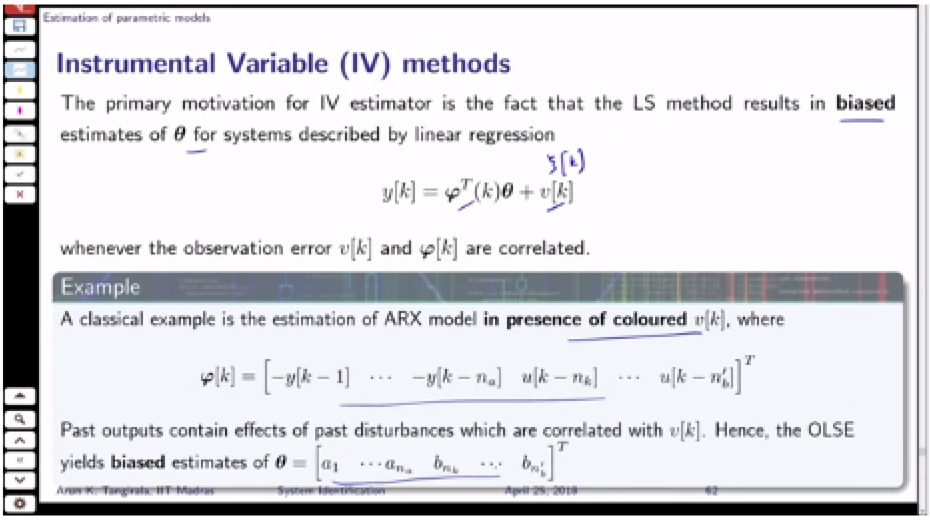


Okay? Earlier we might have used a notation Xi that doesn't matter just a dummy variable. So whenever the regressors are correlated with whatever you have left out, you know that least squares estimates result in biased estimates. Now, the objective of IV method is to obtain unbiased estimates. I'm not going to change my assumptions at all. I'm going to still work with the same equation. But I want unbiased estimates of theta.

I'm not going to change my model; I'm only going to change my estimation method. The bias that I obtain in theta is not because of the model necessarily. Always remember, when you have bias in parameter estimates, it can be due to two things: improper model specification or inappropriate method of estimation, or both. So in this case, I'm determined to use the same model structure but I want to change the method of estimation. Now, the idea-- so as an example, and then I'll explain the idea of IV method. As an example, suppose I'm fitting an ARX model in presence of coloured v k, we have discussed this case before, then the regressors are past outputs and past inputs.

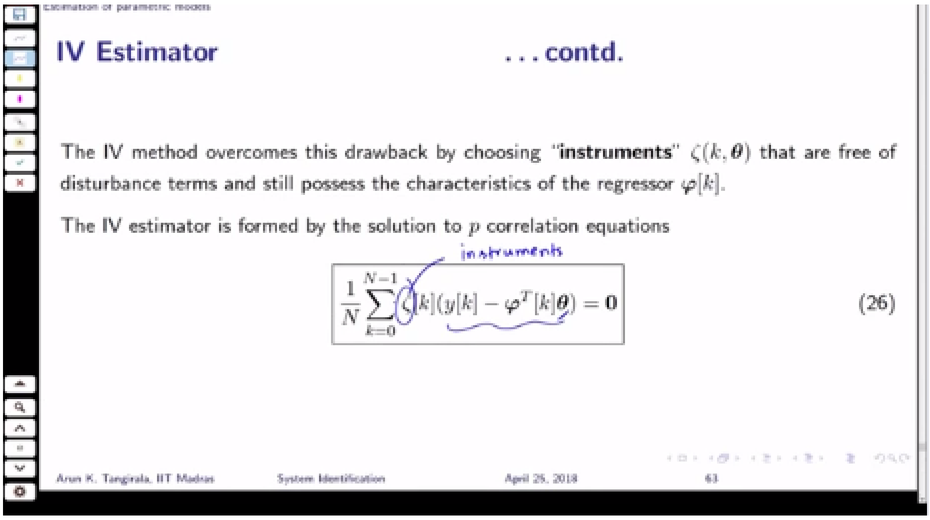
Then we know that the past outputs contain effects of past disturbances, so therefore the regressors are correlated with whatever I leave out and I should expect biased estimates of these parameters. It's pretty much like an OE.

(Refer Slide Time: 04: 16)



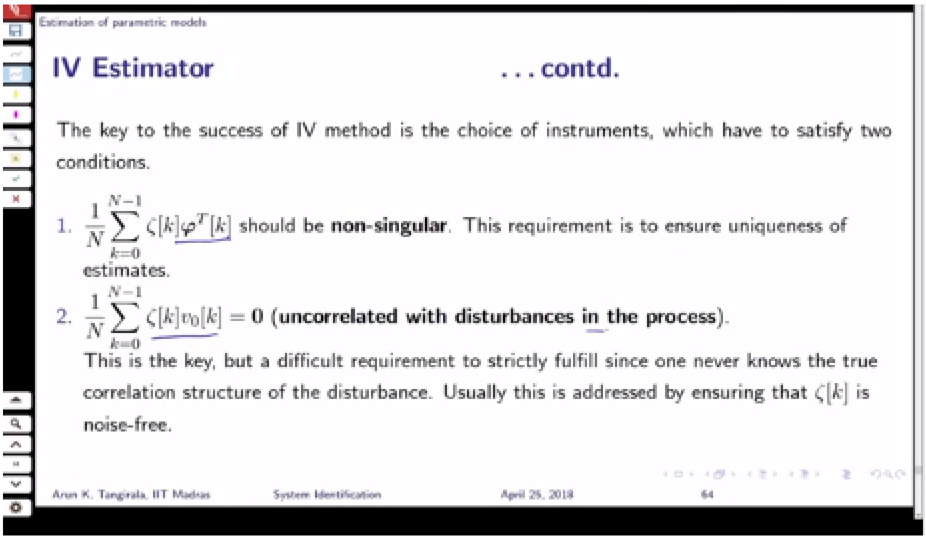
So what is idea in IV? In IV now instead of using least squares methods, we construct the zetas. Now you can see the strong connection with the correlation methods. So, we are going to know set up p equations to estimate those p parameters theta by requiring that the residuals of this model be uncorrelated with some function of past data. Okay, these zetas are essentially instruments. At each k depend, I have one instrument and I have to find p such instruments, so as to solve this equation. And we know very well from our previous discussion that we had in the estimation of FIR models that zetas have to satisfy two conditions. Right?

(Refer Slide Time: 5:10)



One that it should be strongly correlated to the regressors, otherwise, you can have a singularity issue. You may not be able to estimate theta uniquely. Then the other issue which is the other requirement is that it should be uncorrelated with the disturbances in the process.

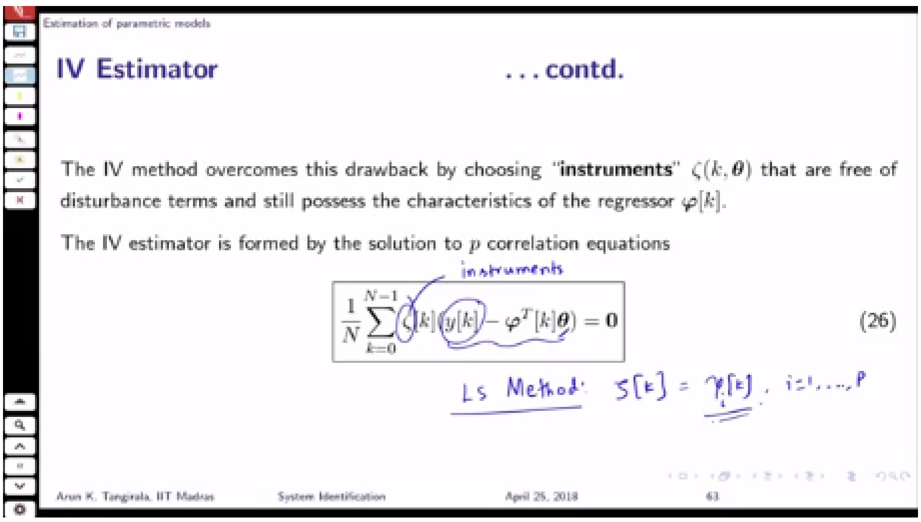
(Refer Slide Time: 05: 29)



That means, or you can say it's uncorrelated with the residuals. But remember, the stochasticity is in y k. That is also stochasticity in Psi transpose, but the main issue we know in the IV method is that the disturbance in y k is colored with the regressors, so we want essentially, zetas to be uncorrelated with the disturbance in y k. All right. Now comparing with the least squares the instruments, the least squares method can also be thought of as an IV method, but in the least squares method, what are the instruments?

Least squares method also solves p set of equations, you remember, right? But in the covariance domain, where the instruments are nothing but-- sure? The regressor itself. So you're using the regressive itself to kill the disturbance. But when the regressor is correlative with instruments, it cannot kill it. Without cross checking, if you blindly go ahead and implement least squares method or without knowing, then, depending on your luck, you may get biased or unbiased estimates. So in the least squares method can also be thought of as IV method where the instruments are the regressors themselves. See, there are many ways of looking at the least squares method as I have explained in the least squares lecture, you can think of it as an optimization problem or you can think of it as correlating out the noise by two by correlating into the regressors and so on.

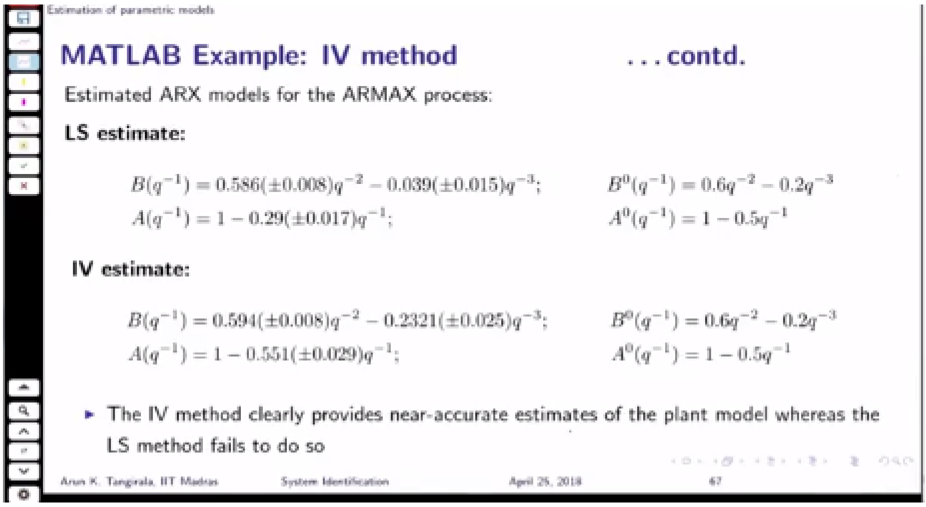
(Refer Slide Time: 07: 18)

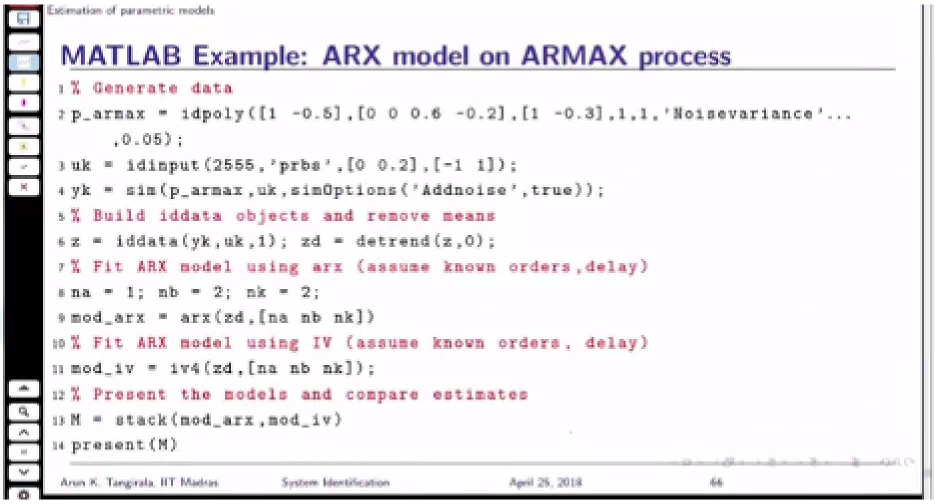


So, you can look at least squares method in different ways. So, one choice of the zeta that is instruments is, first estimate an ARX model using standard linear least squares. What should be the requirement? Zeta essentially should is closely related to regressors. What does the regressors consist of? Past outputs and past inputs in SYSID, right? Which means Eta should contain effects of inputs for sure, but not eta sorry, the zeta should contain effects of inputs.

Therefore, one choice of instruments is to first obtain some g, it may be biased; it doesn't matter. Then simulate it. So these are nothing but simulated outputs, where you don't add any noise. What is the net result? Zetas will be strongly correlated with the regressors but they'll be devoid of any disturbances, right? Of course this is a preliminary estimate of instruments. You can improve the instruments if you want. Or you can pass u k through some suitable pre-filters, you've chosen pre-filters, whatever pre-filters you want, you don't have to necessarily estimate an ARX model and do that. But the basic fact remains that the zeta, the instruments would contain strong components of inputs and zero components of the measurement error or the observation error. That is the idea.

(Refer Slide Time: 09: 07)

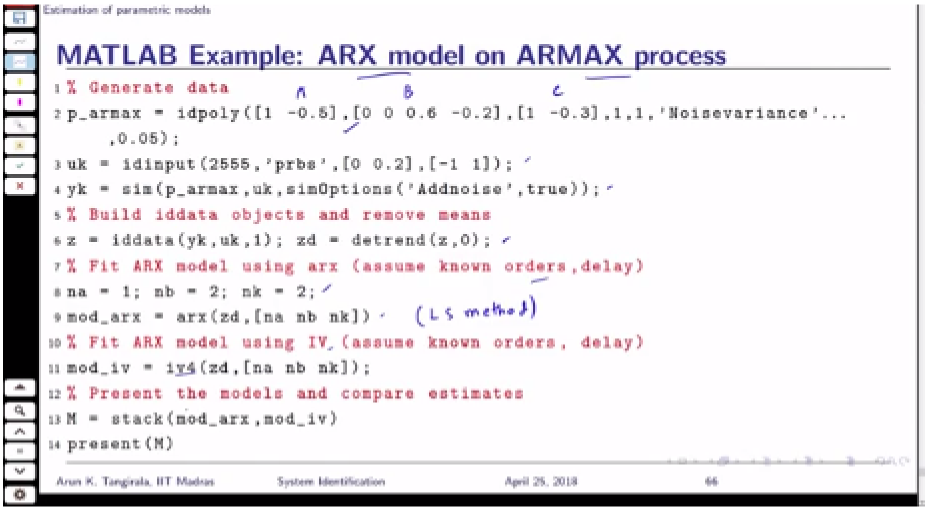




So you can refer to these books by sort of [8:59 inaudible], for a more generalized discussion. Let's actually, look at an example here, where I'm estimating an ARX model on data generated from an ARMAX process. So the first line here creates the ARMAX object, you can see that this is A, B, C and D and F are 1, so it's an ARMAX process. And I'm using a band limited PRBS, simulating it with the noise. I have set the noise variance to 0.5 already. Then you fuse the data. Assume that delays known, order is known because the purpose of this discussion is something else. So you specify the orders and now you estimate the ARX model.

We know from theory that the estimates of this ARX model-- what does ARX use? ARX uses least squares. So the estimates can be expected to be biased. Now, the same model but using IV, right? No noise and delay, and you can use the IV for routing for this. I'll tell you what is IV for shortly. IV will also has another expression in the medical field. But I hope it is not as bad-- the method is not as bad, the method is much more simple than taking a IV. Anyway, so that I before routine implements this and you obtain the estimate of the ARX model through a IV method.

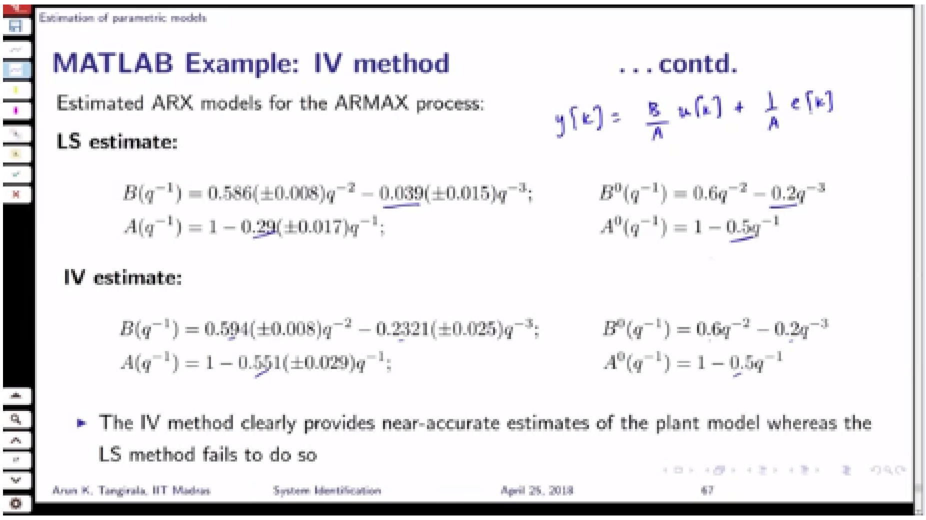
(Refer Slide Time: 10: 39)



So let's compare these two models. Sorry, these two estimates of the same model. You can clearly see the bias in the estimate of this coefficient here, and more importantly significant bias here. That means we know that point estimate will be different from the truth. But here the point estimates are different from the truth, not by chance. There is a systematic error and that's what we mean by bias. Whereas here with the IV method, pretty close. If you increase the length and do amount of the data and do a Monte-Carlo simulations in average, you can show that the average will coincide with the truth.

Okay, so the IV method in this case clearly provides near accurate estimates of the plan model. So always for ARX estimation, if you want to estimate the ARX model for any data, it's better to use IV. Unless you're sure that the underlying-- the data generating process is ARX. Because you're guaranteed unbiased estimates of g, remember the focus here is on g because ARX anyway you know, the model structure is B over A plus 1 over A. So the anyway you can't do anything much about the noise model. So if you want to obtain unbiased estimates of G, regardless of the data generating process, then it's better to use IV. So it's like also looking at OE way of estimating. So IV provides an alternative to OE. But the estimates may not be as optimal as you obtain with a quadratic criterion PEM for OE.

(Refer Slide Time: 12:32)



That is why about on the other hand, computationally, IV methods are very simple. You just have to generate instruments through some pre-filtering and solve a couple of linear equations is, after all, you're solving p linear equations. That is the reason why for B-J models in the system identification toolbox, the estimation is initialized through IV rather than through OE. Because OE requires a laborious nonlinear least squares estimation.

Any questions? Okay. So very quickly let me conclude with the multi-stage IV 4 algorithm. IV 4 is nothing but a four stage IV there, prepared the data estimate delay, that's fine. All of this is okay. Specify the orders of the ARX model. Now estimate an ARX model the desired order using the least squares method as usual. From there construct the instruments. So up to state 4 we know and up to stage 5 it is IV. Then you build a time series models for the residuals and pre-filter the input-output data the inverse of the noise model. Okay? Once you do that then you can recomputed the IV estimate of the ARX model parameters using the filter data, and the same instrument filters as you're used in step 2. The step 6 is to take care of the fact that the underlying process may not be ARX, could be ARMAX, could be OE, BJ and so on. Just to account for that, you include this. And this is what is implemented by IV 4 and so on. IV doesn't implement all of this, IV goes up to stay a step 5. Okay?

So these are the steps involved in IV 4. Any questions on instrumental variable method? Fairly simple. The basic concept that you remember is that you use instruments to kill the effects of noise. Everywhere, that is the effort. Yes.

Student 1: Why is IV method sub optimal compared to OE?

Well, basically, also that you're not solving I think it is because you could not necessarily generate instruments in the most optimal way. And also you're not explicitly optimizing any minimizing any objective function here.

Student 1: If there would still at optimization objective [15:12 inaudible] quadratic optimization--

Where?

Student 1: y minus y, I mean, least squares y hat was constructed using the regressors being--

(Refer Slide Time: 15:20)

Multistage IV4

1. Prepare data and estimate I/O delay n_k (samples).
2. Specify the orders n_a and n_b of the ARX polynomials to be estimated.
3. Estimate an ARX model of the desired order using the LS method to obtain $\hat{A}^{(1)}(q^{-1})$, $\hat{B}^{(1)}(q^{-1})$.
4. Construct instruments from the fit ARX model: $\zeta^{(1)}[k] = \dfrac{\hat{B}^{(1)}(q^{-1})}{\hat{A}^{(1)}(q^{-1})}u[k]$
5. Compute the IV estimate using the instruments $\zeta^{(1)}[k]$ and the residuals from the resulting model.
6. Build a time-series model for the residuals and pre-filter the input output data with the inverse of the noise model.
7. Re-compute the IV estimate of the ARX model parameters using the filtered data and the same instrument filters as in Step 2.

MATLAB: `iv4, iv`

MATLAB Example: IV method ...contd.

Estimated ARX models for the ARMAX process:

$y[k] = \frac{B}{A}u[k] + \frac{1}{A}e[k]$

LS estimate:

$$B(q^{-1}) = 0.586(\pm 0.008)q^{-2} - 0.039(\pm 0.015)q^{-3}; \qquad B^0(q^{-1}) = 0.6q^{-2} - 0.2q^{-3}$$
$$A(q^{-1}) = 1 - 0.29(\pm 0.017)q^{-1}; \qquad A^0(q^{-1}) = 1 - 0.5q^{-1}$$

IV estimate:

$$B(q^{-1}) = 0.594(\pm 0.008)q^{-2} - 0.2321(\pm 0.025)q^{-3}; \qquad B^0(q^{-1}) = 0.6q^{-2} - 0.2q^{-3}$$
$$A(q^{-1}) = 1 - 0.551(\pm 0.029)q^{-1}; \qquad A^0(q^{-1}) = 1 - 0.5q^{-1}$$

- The IV method clearly provides near-accurate estimates of the plant model whereas the LS method fails to do so

Estimation of parametric models

MATLAB Example: IV method ...contd.

Estimated ARX models for the ARMAX process:

$y[k] = \frac{B}{A}u[k] + \frac{1}{A}e[k]$

LS estimate:

$$B(q^{-1}) = 0.586(\pm 0.008)q^{-2} - 0.039(\pm 0.015)q^{-3}; \qquad B^0(q^{-1}) = 0.6q^{-2} - 0.2q^{-3}$$
$$A(q^{-1}) = 1 - 0.29(\pm 0.017)q^{-1}; \qquad A^0(q^{-1}) = 1 - 0.5q^{-1}$$

IV estimate:

$$B(q^{-1}) = 0.594(\pm 0.008)q^{-2} - 0.2321(\pm 0.025)q^{-3}; \qquad B^0(q^{-1}) = 0.6q^{-2} - 0.2q^{-3}$$
$$A(q^{-1}) = 1 - 0.551(\pm 0.029)q^{-1}; \qquad A^0(q^{-1}) = 1 - 0.5q^{-1}$$

- The IV method clearly provides near-accurate estimates of the plant model whereas the LS method fails to do so

Arun K. Tangirala, IIT Madras System Identification April 25, 2018 67

MATLAB Example: ARX model on ARMAX process

```
% Generate data
p_armax = idpoly([1 -0.5],[0 0 0.6 -0.2],[1 -0.3],1,1,'Noisevariance'...
    ,0.05);
uk = idinput(2555,'prbs',[0 0.2],[-1 1]);
yk = sim(p_armax,uk,simOptions('Addnoise',true));
% Build iddata objects and remove means
z = iddata(yk,uk,1); zd = detrend(z,0);
% Fit ARX model using arx (assume known orders,delay)
na = 1; nb = 2; nk = 2;
mod_arx = arx(zd,[na nb nk])
% Fit ARX model using IV (assume known orders, delay)
mod_iv = iv4(zd,[na nb nk]);
% Present the models and compare estimates
M = stack(mod_arx,mod_iv)
present(M)
```


Choice of instruments

Certain natural choices of instruments are

1. **Noise-free simulated outputs:** Estimate an ARX model of the necessary order and simulate it in noise-free conditions

$$\zeta(k,\boldsymbol{\theta}) \equiv x[k]: \quad x[k] = \frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})}u[k] \tag{27}$$

   These instruments can be expected to satisfy both requirements under open-loop conditions.

2. **Filtered inputs:** Pass $u[k]$ through suitable pre-filters

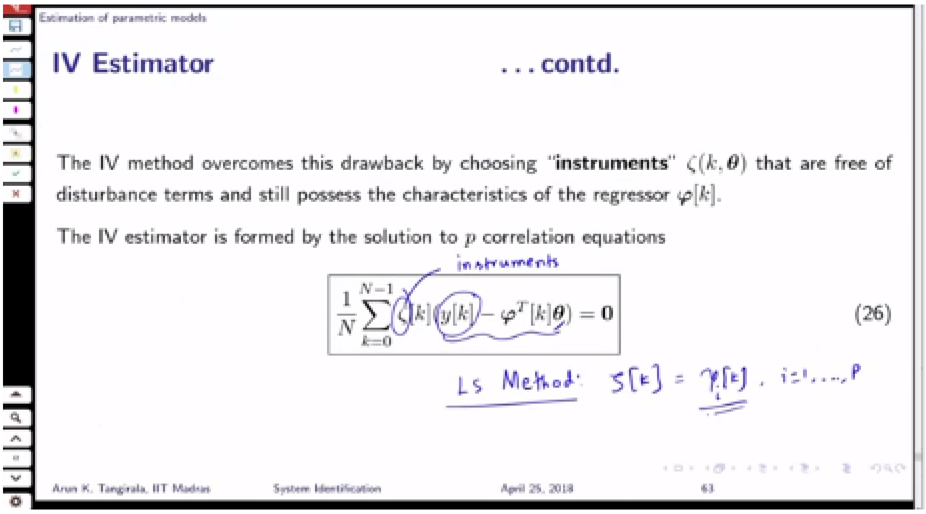
$$\zeta(k,\boldsymbol{\theta}) = \mathbf{K}_u(q^{-1},\boldsymbol{\theta})u[k] \tag{28}$$

For an in-depth treatment of this topic and a generalization of the IV method see Soderstrom and Stoica, 1994 and Ljung, 1999.

IV Estimator ...contd.

The key to the success of IV method is the choice of instruments, which have to satisfy two conditions.

1. $\displaystyle\frac{1}{N}\sum_{k=0}^{N-1}\zeta[k]\boldsymbol{\varphi}^T[k]$ should be **non-singular**. This requirement is to ensure uniqueness of estimates.
2. $\displaystyle\frac{1}{N}\sum_{k=0}^{N-1}\zeta[k]v_0[k] = 0$ (**uncorrelated with disturbances in the process**).
   This is the key, but a difficult requirement to strictly fulfill since one never knows the true correlation structure of the disturbance. Usually this is addressed by ensuring that $\zeta[k]$ is noise-free.



Correct. But in a way, where is an objective function?

STUDENT 1: But we showed that it is equivalent having that objective--

To solving this p equations, correct. That's for linear least square only. For the nonlinear least squares problem. There's no equivalent projection theorem. Okay? So that that is one of the issues and also the reason for IV 4 is remember, your IV method generates instruments using this ARX model, which itself is biased. So its instruments may not be so, so good. But what we mean by so good is, it may not be-- it may still be satisfying these two criteria that we set out. But ideally we want this to be as zero as possible. Numerically it won't be zero. Because there is a bias in the parameter estimate that you have generated in.

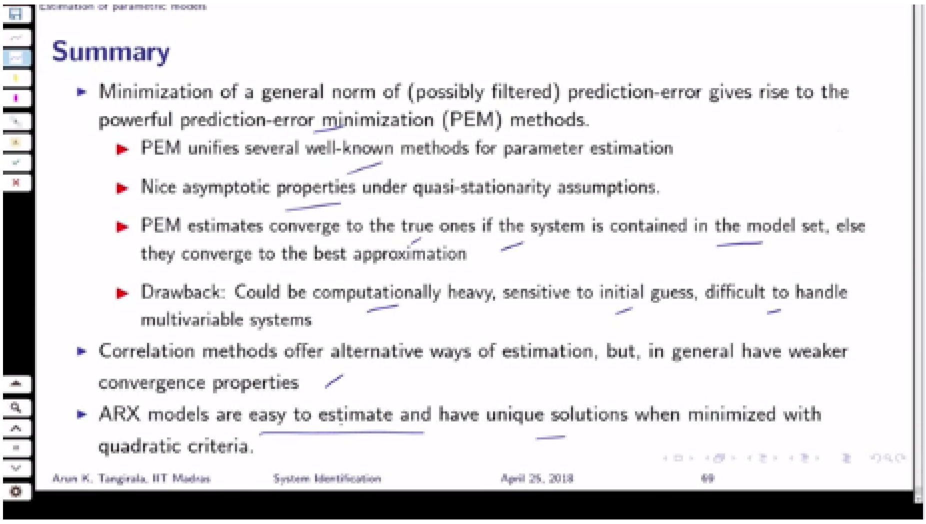
And maybe this is not rich enough, the covariance is-- see also what does the is more importantly, what does the variance of the parameter estimates depend on from at least squares viewpoint? You want Phi transpose Phi to be very high. Here, we are only saying should be non-singular, but actually we should impose other conditions that the covariance between the instruments and regressors should be very high, as high as possible, right? But that as high as possible is not necessarily guaranteed by IV but better by IV 4. And that's why IV 4 is better than IV. But in general, you can't expect the same level of optimality as but having said that, I don't know in the sense that-- see practically you will be stuck with some local minimum, how good that local minimum is with respect to the IV method that you obtained, there is no analysis whatsoever. It's all asymptotical expressions.

So, practically IV 4 is very good, if you want to get only good estimates of g, right, because the instrumental variable method is focused on getting you a biased estimator of the plant model. Whereas it doesn't really, sorry, unbiased estimator or the plant model as against the noise model. So, anyway, so, let that brings us to the close on estimation of parametric models. To summarize PEM unifies a lot of several well-known methods and is based on minimization of prediction error or function of prediction error. And there are some additional embellishments that you can make. Then, it has some nice asymptotic properties. Gaussian distributed estimates, consistency and so on.

And we have studied the consistency properties of them. The PEM estimator converges to the true ones. If the system is contained in the model, other ways it may converge to the best approximation. And the best approximation is determined by the input, lastly for a given model structure. Drawbacks is it could be computationally heavy and sensitive to initial guests, difficult to handle multi variable systems and so on. Because the moment you go to multi variable let us say some 3 by 3 or 4 by 4. Then you have 4 by 4you have to estimate, 16 transfer functions.

So you've to set up PEM on those, for the 16 transfer functions and that can be demanding computationally. That's where the state-space modeling comes in. And that is this juncture actually offers a very nice transition to switch over to state-space identification. We have also studied correlation methods and in particular the IV method. In general, we know that the ARX models are easy to estimate and have unique solutions. And you can estimate them either using least squares or IV, IV is preferred. Any questions on PEM? Or anything that we have discussed this far. Alright. So, that kind of brings us to a close on the parametric estimation part.

(Refer Slide Time: 19: 30)



I just wanted to spend a couple of minutes on state-space identification, just get started. And then come to the full details in the next lecture. Okay, so we are moving from the input-output world to the state-space world, for various good reasons. We know some of the reasons already. Can you name-- recall one advantage of working state-space models as against input-output models. Long ago and we discussed state-space descriptions; I mentioned at least two advantages.

STUDENT 2: We talk about the hidden variables.

Correct. So what about the story of hidden variables? Good. So what happens when you have hidden variables? Which are nothing but what we call as states. So what advantages that state-space model offer, when you have hidden variables? Sorry. Dynamics of the--

STUDENT: [21:00 inaudible]

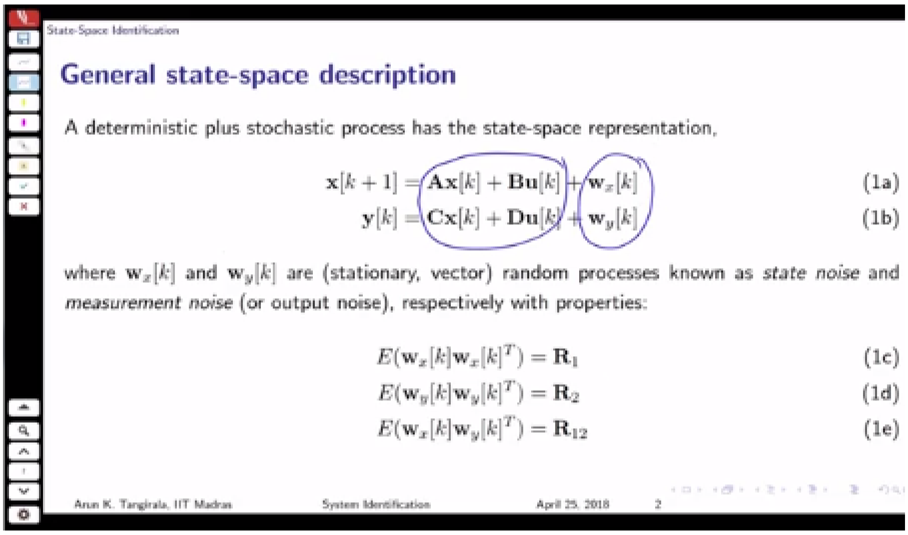
Okay. I want a better articulation of the advantage. One of the prime reasons for turning to state-space identification is joint estimation of hidden variables and identifying the model, when you have to do both. That is not only reason, but that is one of the prime reasons. So, you have a situation where you have to build a model, as well as estimate the directly observed variables that is a hidden variables. So, you want to estimate the internal dynamics, so, to speak of the hidden variables. Like we have given several examples to that, but essentially, in a state-space model you have states, inputs and outputs, and when you want to build a model between the states and the inputs, where the states are not observed, but directly, rather than that measurements of the states are obtained.

So from the measurements of the states we are supposed to do two things. One is estimating the hidden variables, as well as building a model for the dynamics of those hidden variables. Which is not possible at all in the transfer function domain. And the second advantage of working with state-space models is multi variable systems, right? We have just discussed for a 4 by 4 system, I need to estimate 16 times of functions, whereas with the state-space model is only one state-space model. What changes is only the dimensionality of the matrices, but that doesn't really make much of a difference at all. The algorithm is the same, it doesn't matter to it, whether you're A, B, C, D, or matrices of one size or the other size, it does the same job.

Computationally bit demanding, but not as painful as estimating input-output models. Imagine with 16 transfer function models I have to get the orders of each of those channels and that become a hassle, right?

So, I'll just leave you with this general state-space description for this lecture. We have seen this part of the model before, right? When we talked about state-space models for deterministic models, we had but deterministic processes. We had only A,B,C, D. But now we are dealing with a deterministic plus stochastic system and the stochastic terms appear both in the state equation and the output equation. What is the meaning of having these two terms separately?

(Refer Slide Time: 23: 29)

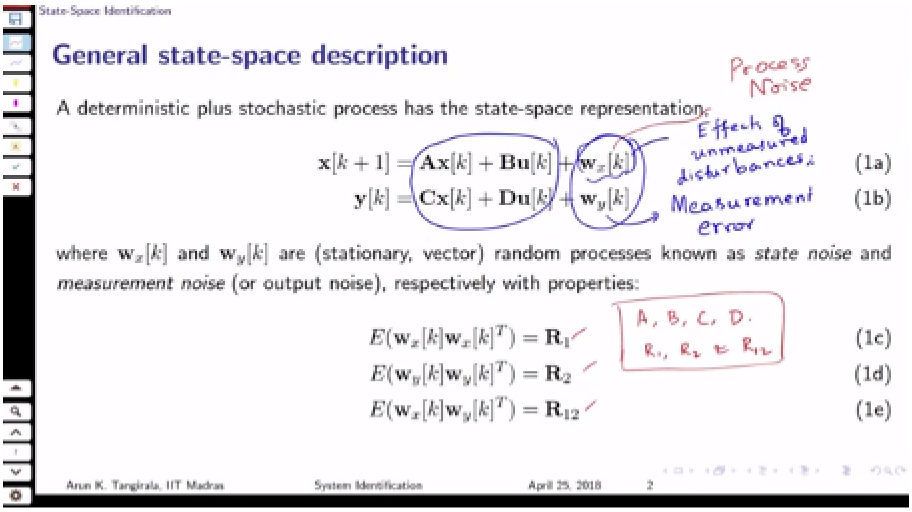


Earlier in the transfer function description, we had only one stochastic term v k, which is a lumped effect, right? It contains the effects of sensor-- a measurement error as well as unmeasured disturbances. But now we distinguish, w y is going to help me describe measurement error. Because y is now treated purely as measurements of states x and w x will carry now effects of unmeasured disturbances, usually called process noise. It's not only characterizes effects of unmeasured disturbances, but also any other stochastic nature of the process itself. Okay?

So, your underlying process itself could be stochastic, remember. So in effect this is called the process noise. So this term here is called process noise. Both are random signals, and they have their own covariance matrices R 1, R 2 and they could be cross correlated to-- your measurement error could be cross correlated with the process noise. So the full state-space description for deterministic plus stochastic process involves specification of A, B C D, R 1, R 2 , R 1 2. This is the full state-space description.

And you can convert--

(Refer Slide Time: 25: 18)

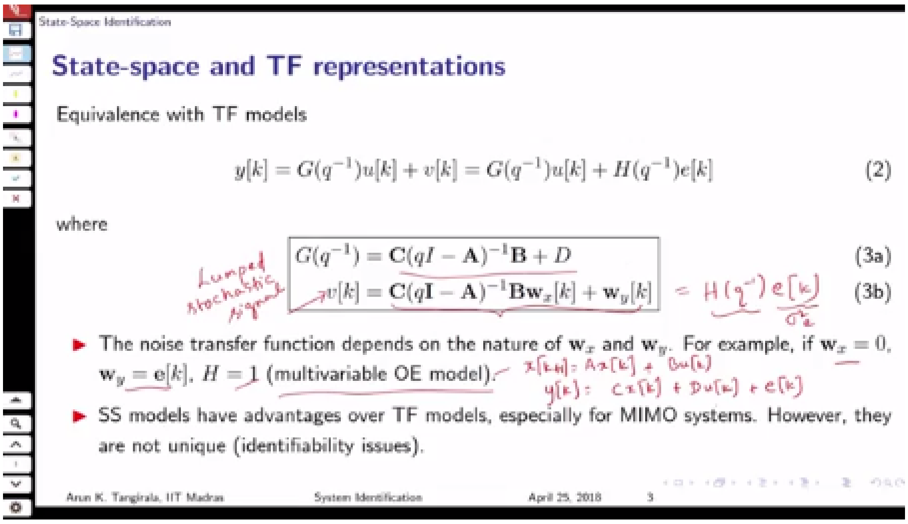


I leave you with this slide. You can convert the state-space description to a transfer function representation. So that I know what a G and H now in terms of A, B, C and D. G anyway is what we have seen before. I have kept D are not bold faced because at the moment have dealing with [25:41 inaudible] systems, but you can have y and u multi-output and multi-input also. V k is this, I'm not given expression for H. Why is that? I instead of given only an expression for v k. Now, H depends on the noise model that I fit to this, right?

But as far as I can see, now, clearly v k is a lumped effect. So, this one is a lumped stochastic variable. Now, I see why it is called lumped, right. So, v k contains effects of unmeasured disturbances, measurement noise everything and when you fit a time series model to this, that is when you fit H of q inverse e k, then that is where you get your H and sigma square e. There is no one on one relation between H and C and B and so on, necessarily. Of course you can have under some special circumstances. Suppose w x is 0, then what is v k? There's no process noise let us say. Then v k is purely measurement error. And on top of it if w y is white, then what is v k? That's an OE model. So that's your multi variable OE model.

So, if you have w x0, w y e k then H is1. So you know now the state-space description of multi variable OE process. Essentially for multi variable OE process you have A x k plus B u k, this is your state equation x k plus 1 equals A x k plus B u k and y K is simply C x k plus of course, you can have D u k if you wish, plus e k. So, you can see why it is called output error now. The measurement error is only in the output.

(Refer Slide Time: 27: 50)



So, the next lecture we will ask, what are the methods available for identification of the state-space models? What is the problem statement to begin with? Then what are the method available in general, the method for state-space identification are mathematically bit more involved than what we have seen until now for transfer function models. And will study two classes of models: the black box state-space models, which are freely parameterized and structured state-space models. And for black box state-space models that exist this nice linear algebra based algorithms known as the subspace identification methods, whereas was structured state-space models unfortunately you have come back to PEM. You have to solve them as optimization problems. Okay, so we'll meet in the next lecture. Thanks.