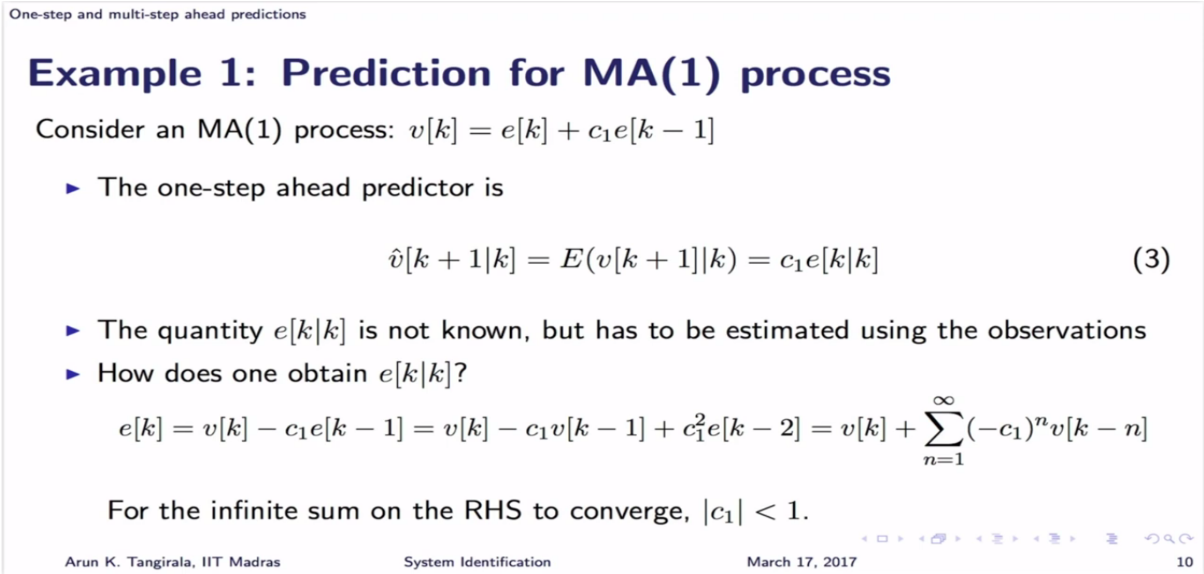
CH5230: System Identification

One step and multi-step ahead prediction 2

So let's take a moving average process. All right. So I have a moving average process and given now v evolves in particular way.So we are jumping from the world of random variables to the world of random signals. You should observe that now given that this is an MA1 process. And I want to construct a one-step ahead prediction.What is the one-step ahead prediction?Given all the information up to k I want to estimate, what happens at k plus 1? I want to guess what happens at k plus 1? That is one-step ahead prediction.So given that vk, sorry, evolves this way. What would be the one-step ahead prediction? Well, v hat of k plus 1 given k.Given k meaning not just that k given information up to k. Is conditional expectation itself. What do you mean information here?How do you understand my information?

(Refer Slide Time: 01:13)



Simply observations of v, there is no mystery about anything. All of this is pretty straightforward when I say information, I'm given observations of v. I am given v values up to k. Of course, it's a single realisation, but yes, I'm given up to k. So which means I'm given vk, vk minus 1, vk minus 2, each of this is a random variable. We study the result between, for univariate case that is yis univariate x is single random variable. Now,vk plus 1 you can think of as y. If you apply the previous result what would be x? With respect to whateverwe are study until now, I can say vk plus 1 is y, I want to predict vk plus 1. What is x? Simply vk.We just know said we are giving the information up to k. So why are you again going back to only k?All the random variablesx is now not a single random variable. x is now going to be from minus infinity, whatever observations of v up to k,which means vk, vk minus 1, vk minus 2, up to v8 minus infinity.But the result still holds. So we don't apply that conditional expectation and say that the best prediction Of v at k plus 1 in the minimum meansquare errors sense will not keep saying that ek is conditional expectation of vk plus 1 given k.

Now, you apply that to the given model, right. So you have vk equals ek plus c1ek minus 1. So when I take the conditional expectation of the right hand side, because I want to construct this optimal prediction. I have expectation of ek plus c1 ek minus 1 given k, information up to k. What would this, sorry. What should we hear? Is that a confusion? Some of you are very silent today. It's a conditional expectation of vk plus 1. What is vk plus 1 as per the model?That's all. So ek plus 1 plus c1 ek. Now what can you say about these two terms here. I have expectation of ek plus 1 given k. What is that? That zero, very good. Because there is nothing in k or k minus 1 or anything in the past, which will improve the prediction of ek plus 1 beyond its average.What about a second term? Now there was a question about this term about this expression here, after the end of the class yesterday, y is not expectation of ek given k 0. First of all do you believe it should be 0?What do you think? Do you think it should be 0?Nithya, what do you think? What do you think sir?

[04:54 inaudible]

Sorry.

We know information about [04:58 inaudible].

The information is it in the form of ek or vk? It's a good point.Do we havevk or ek?

Vk.

Vk. So we do have the information but in an indirect fashion, correct. So that's the key there that I do have the information about ek now. Whereas in the first term I don't have anything in the past forek plus 1, whereas expectation of ek given k is, forget about vk minus 1, vk minus 2, because that won't have any effects of ek, vkwill have something about ek, because after all as for the modelvk contains partly the ek.Sowhich means expectation of ek given k, what is it now, is it a prediction? We use some. We introduce some terminology earlier, right? What is it? It's a filtered; it's a filtered version of ek.Now we are looking at filtering.We are not looking at prediction.Prediction always involves at least one-step in future. So the second component is a filtered version.So from vk somehow I have to apply. Like your coffee filter or whatever. I put it through a filter only ekshould come out. See the simple example that I give is ek is like a shockwave, like an earthquake, ek earthquake you can remember.

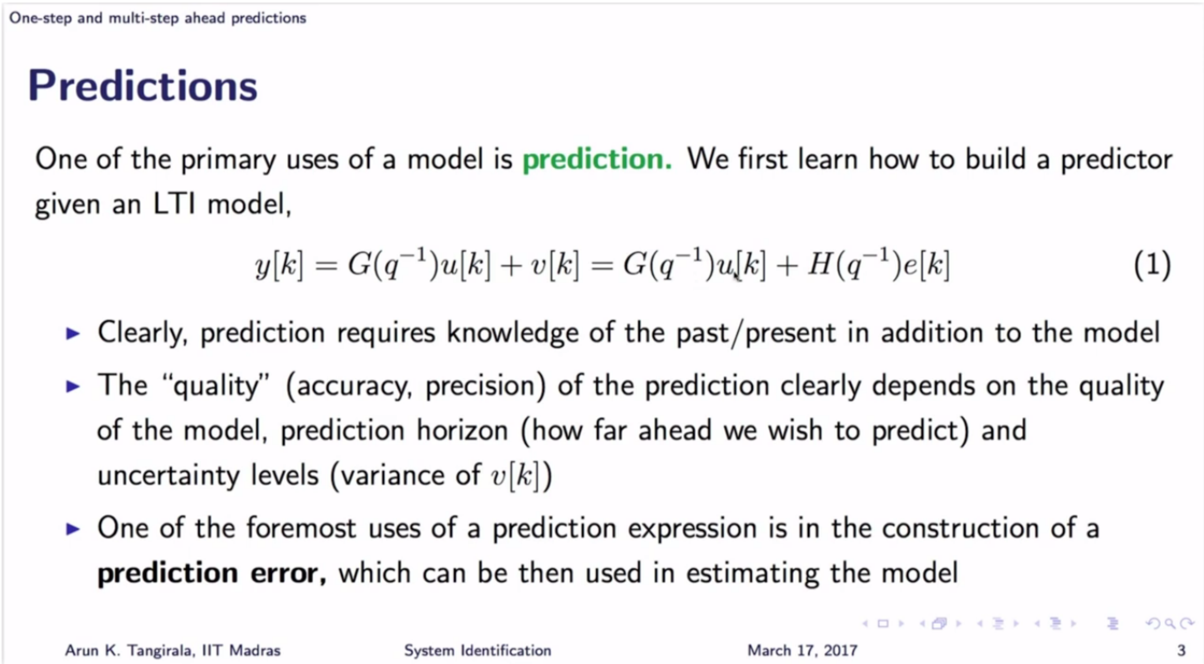
Now,I can't predict an earthquake. How many of earthquakes have occurred in history until now, still my prediction of earthquake is kind of mean whatever average. Now if I want to assess how intense that earthquake was. That is filtering given the effect of the earthquake. After the earthquake has occurred, I want to estimate or at the time the earthquake is occurring, I can guess, but I can predict, correct. So what you are observing is vk. You're shaking of the building or whatever structure that you are looking at that is your vk. That is what you are observing. You are not observing the earthquake. You are observing the effect of the earthquake that is vk. Based on that you can assess how intenseek would be. But you couldn't predict the earthquake. That is how white noise is, right. So you can't say since I couldn't predict whatever I'm observing my prediction of earth, my estimate of earthquake is 0. That doesn't make sense. The effect of earthquake is visible right in front of you. V is for visible. You can think of it that way. So,e expectation of ek given k is exactly that.

It cannot be 0.Now, how do they recover this that's extremely important? Another way to recover this is go back to the generating equation. What do we mean by recovering? I want estimate ek given vk's. So what you do is you rewrite ek in terms of vk's. You see that all I have done is recursive substitution for ek's. So what does it this equation tell me, if I want to recover ek accurately what do I need? I need observations all the way up to minus infinitynot just vk alone. Okay. I thought I could just use vk,but it turns out to correctly recover ek I need all the observations in the past. Now that sounds a bit awkward, because the observations in the past actually did not have any information about ek as such however, if you look at the equation ek says it is vk minus c1 ek minus 1. What is c1 ek minus 1? It is the best prediction of vk, if you are standing at k minus 1, if you are given information up to k minus 1. So you can think of it this way you can say ek, instead ofasking in this way you say ek isnothing but vk minus whatever prediction you would have made of v.Standing at k minus 1, whatever you would have predicted of vk, correct. This is the trick that is used. Otherwise how will you ever make a prediction at all for this?Apart from that the most important thing that you should observe is, now ek has been obtained, is a non-linear function of your model parameters c1. Do you see that? ek is a non-linear function of the model parameters c1.Which means my v hat of k plus 1 given k is going to be a non-linear function of the parameters.And that's always the case with moving average. So v hat of k plus 1 is c1 times an estimate of ek obtained from information up to k, but that estimate in turn requires the model itself. And in the net effect is that the predictor isn't non-linear function of the parameters c1. And that is way estimating MA models is not so easy. Compared to auto regressive models, which will give you linear predictors. Now another point that you should observe is if you go back to this ek here, expression for ek. If I want an estimate of ek and it tells me that there is an infinite summation here that is involved. I know ekis stationary which means it cannot the amplitude should be bounded. That's the foremost thing. Now, what you are actually implying here is this submission should converge.When it will converge? When modc1 is less than 1. Now that's a new restriction I'm seeing on moving average model, right.Earlier we never restricted. We never imposed any restrictions on moving average model at all. We simply said that moving average models are stationary as long as the coefficients are bounded. AR models on the other hand have notnecessarily stationary unless the polls are within the unit circle.But now we are saying that there is an additional restriction on the moving average model, which is that it should be invertible. So what you're essentially doing is, in order to complete the prediction here. We are saying there is an e that is generating b. There is an e that is generating b. And now in order to predict v, I need to recover e. So therefore I need some other filterwhich will take in all the information of vk.The vk, vk minus 1 everything and it'll produce this ek. So it is as if you are doing an inversion here. You're the forward problem is ek generating vk.The inverse problem is given v recovery ek. This inverse now in some sense should be stable. If this inverse is not stable, my predictions of ek will go haywire. Ek as such as white noise, but my estimates of ek are going to go haywire. Therefore moving average models are in general any time series model, if you want to use it in prediction, the inverse of the model should be stable. In some sense you can say here h of q inverse. For this MA1 is 1 plus c1 q inverse. What is the inverse although there is a lot of formalization that is required I'm cutting short all that formalization and I'm straight awaywriting the inverse. You should not think it is so obvious that the inverse is simply 1 over 1 plus c1 q inverse. Because what is this rule of inverse?The rule of inverse is to connect v to e. And that requires some formalization to prove that the inverse is 1 over 1 plus c1 q inverse. Anyway, so cutting short that formalization simply h inverse is the inverse of h. What are we saying now? To obtain a stable estimate of e, it is like any other random process now, v issome random process generating e, correct. And what kind of a transfer function is this? This is MA. This is?

AR.

AR. We know already if AR transfer functions have to produce stationary signals. Then the pole should be within the unit circle. Which means we are saying that the poles of h inverse should be within the unit circle, which means the zeroes of h should be within the units circle?And we say any noise model is invertible, if the zeros are within the unit circle.Any noise model is stationary. If the poles are within the unit circle, which means now we have an additional requirement that the noise model should not only be stationary but also invertible.Inversion is becoming necessary because of one prime reason, what is that?Because I do not know the signal input that is generating v. If I know the input I don't have to worry at all, right. I can straightaway use this equation.I can straightaway use ek, I don't even have to say ek given k, I would know ek. It is this white noise that is not known to me that is endogenous to v that I have to estimate is forcing me to seek this inverse model.And that is why we have an additional equipment.On the G there is no equipment.By the way, why are we studying predictions of random processes, because that's more challenging, if you go back to the expression that we had here. If I were to predict y using u, I think I'll be given G, I'll be given H and I'm given u.And the measurements up to k minus 1or up to k depending on what I have. What is a big

challenge in this? The big challenge is not in the first term, because the first time is deterministic. It is the second term that is going to present challenges and that's why we're discussing that. Okay.
(Refer Slide Time: 15:39)



So now let's formalize this very quickly. Whatever we have observed for AR models it's pretty straightforward.AR models say there is nothing to worry about as long as your model is stationary. See the AR model also there is an inverse involved, but we have already said that AR model is stationary. You have already required that.If it is stationary, it's going to be invertible as well. That means the predictions will be stable. So look at this, I have an AR1 process and I have this prediction written up very easy, because the AR model is purely in terms of observations, it makes it. It's a pretty sweet predictor.No, it'slike very, it's like a laddu. Do you have vk equals minus d1 vk minus 1 plus ek. All you have to do is apply the conditional expectation. The first term is given to me. The second term what is a prediction? Zero. So the predictor is simply minus d1 vk. I don't have to recover anything vk is given to me.And since d1 is going to guarantee that I'll get stable predictions. There is no issue at all. So that is AR models are preferred. That is why you have a different routine for AR model estimation, because you will minimize, you will write quotes or algorithms that minimize the prediction error. And if we use Least Squares Algorithm the predictor is linear in unknowns. Here we are assuming model is given in making a prediction, but in the estimation problem, model estimation problem. I'm given data and I'm supposed to estimate d1. So this is always the case with AR models, if the predictor if you can imagine if it was AR2 then you would have minus d1 vk minus 1. Sorry, minus d1 vk minus d2 vk minus 1.

(Refer Slide Time: 17:32)

One-step and multi-step ahead predictions

Example 2: Prediction of an AR(1) process

Consider now an AR(1) process: $v[k] = -d_1 v[k-1] + e[k]$

- The one-step ahead prediction of $v[k]$ is then,

$$\hat{v}[k+1|k] = E(v[k+1]|k) = -d_1 v[k] + E(e[k+1]|k) = -d_1 v[k]$$

- Notice that the **predictor is linear in the parameter**!
  - This is always the case with auto-regressive models.
  - Minimization of squared prediction errors with linear predictors provide unique solutions
  - Thus, AR models hold an edge over MA models in identification.

Arun K. Tangirala, IIT Madras System Identification March 17, 2017 12

Okay. So let's very quickly formalize this, before we conclude the class.So vk in general for a linear random process we know h of 0 is 1, this is a standard linear random process. Let us write this vk in two parts. Very simple now we want to graduate from AR1and MA1 to ageneralARMA process or ageneral linear random process. Why are we doing this? Because when I want to construct the one step ahead prediction, let us say I am given the formation up to k minus 1 and I want to predict k, I can straightaway see that the first term here won't participate, because expectation of ek given information up to k minus 1 is going to be 0. So I have separated out the unpredictable component of vk, whatever is predictable is the summation. But of course, I have to applyan inverse noise model to recovery e, correct. That we have already seen.

(Refer Slide Time: 18:33)

One-step and multi-step ahead predictions

General expression for prediction

Consider the causal impulse response model for $v[k]$

$$v[k] = \sum_{n=0}^{\infty} h[n]e[k-n] \quad \text{with} \quad h[0] = 1, \qquad H(q^{-1}) = \sum_{n=0}^{\infty} h[n]q^{-n} \qquad (4)$$

- Then, for one-step ahead prediction, we re-write the previous expression,

$$v[k] = e[k] + \sum_{n=1}^{\infty} h[n]e[k-n]$$

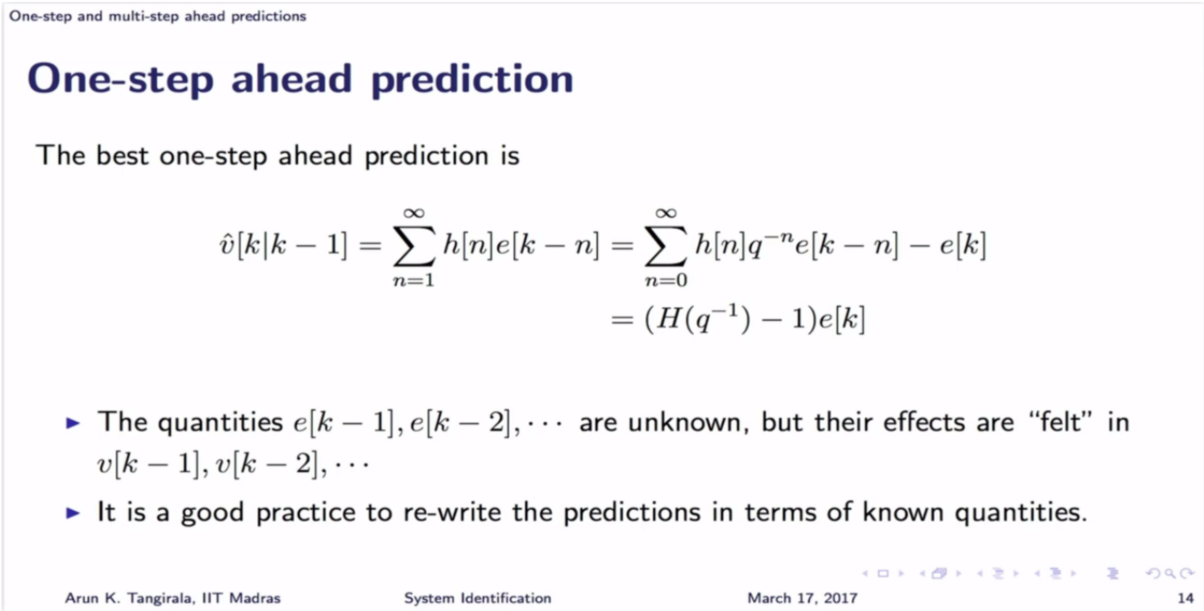
so that the second term on the RHS is (theoretically) known at $k-1$

  - The best prediction of the first term is zero given any amount of past

Arun K. Tangirala, IIT Madras System Identification March 17, 2017 13

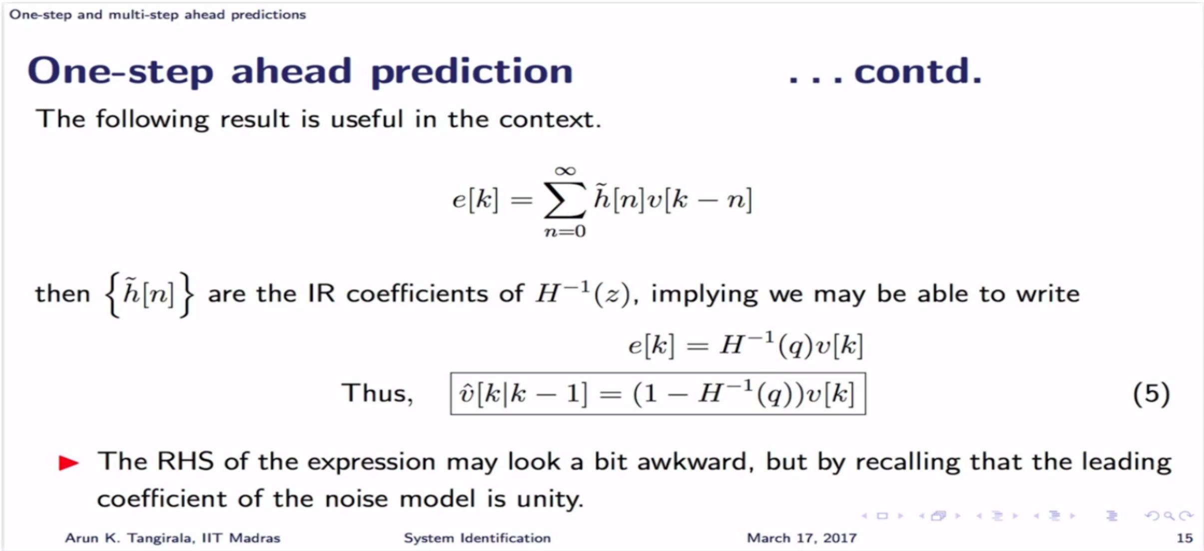
So let's write this now the second term is the prediction v hat of k given k minus 1, I have switched over from v hat of k plus 1 given k to v hat of k given k minus 1 that doesn't matter. It is a dummy variables here, k is just a dummy variable. So the prediction of k given k minus 1 is simply the second term and notice that the second term is nothing, I can rewrite the summation by adding and subtracting ek.All I have done is into this term, this submission begins from 1, correct.By adding ek and I know that already h of 0 is 1. I recognize what is this first term here. The first term is simply h of q inverse ek, correct. So I can write v hat of k given k minus 1 as h of q inverse minus 1 operating on ek. Is it clear? Now what is the ek?Do I know ek? I don't know ek. But ek's h inverse v.Look at this blog diagram.

(Refer Slide Time: 19:43)



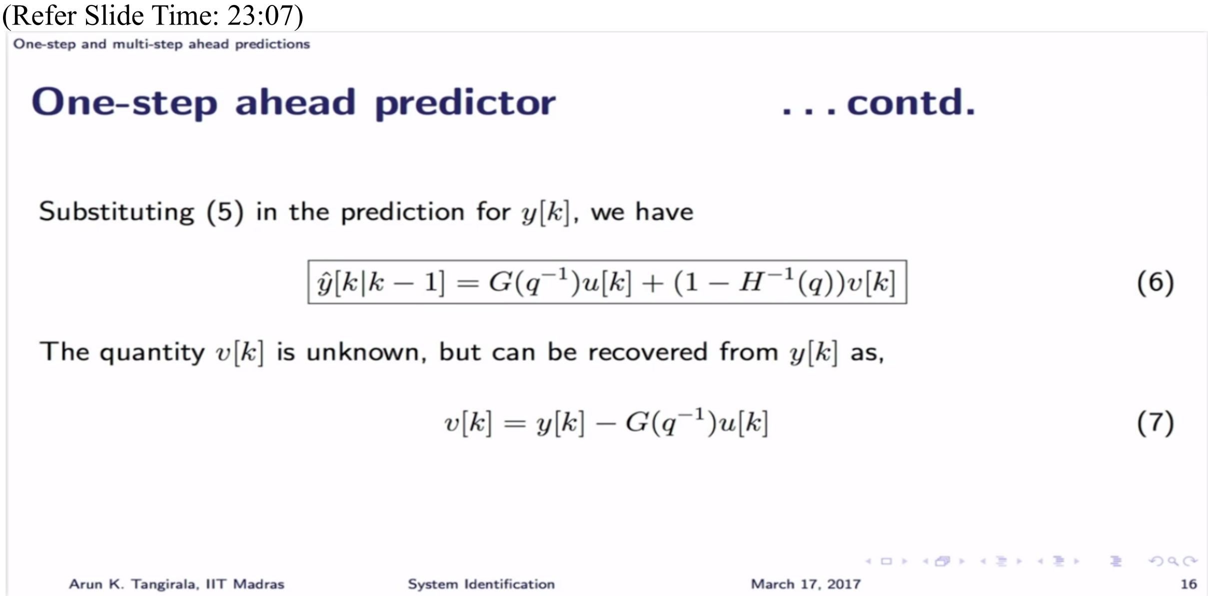
Ultimately my prediction or prediction expression should be in terms of known quantities. I do not know ek. So somehow I have to rewrite this in terms of vk's that is a goal. So I'm going to replace ek with h inverse v.As I said there is a lot of formalization let's not worry about it. So ek is h inverse v, when a substitute for ekas h inverse v, what do I get? How is h inverse defined?H inverse is defined such that when h operates on h inverse or h inverse operates on h you will get 1,like your matrix inverse. So h of q what we have is v hat of k given k minus 1, is h of q inverse minus 1, operating on ek, which is h inverse vk. So what is h operating on h inverse? That is 1. So I have here 1 minus h inverse q operating on vk.Now this is more like it because nowon the right hand side I have known quantitiesvk.However doesn't it sound weird that I'm predicting at k and on the right hand side I have vk, it looks a bit awkward. What I'm intending to do is I'm intending to predict vk. But all you're saying on the right hand side there is vk already. But truly is vk required on the right hand side, what do you think?

(Refer Slide Time: 21:23)



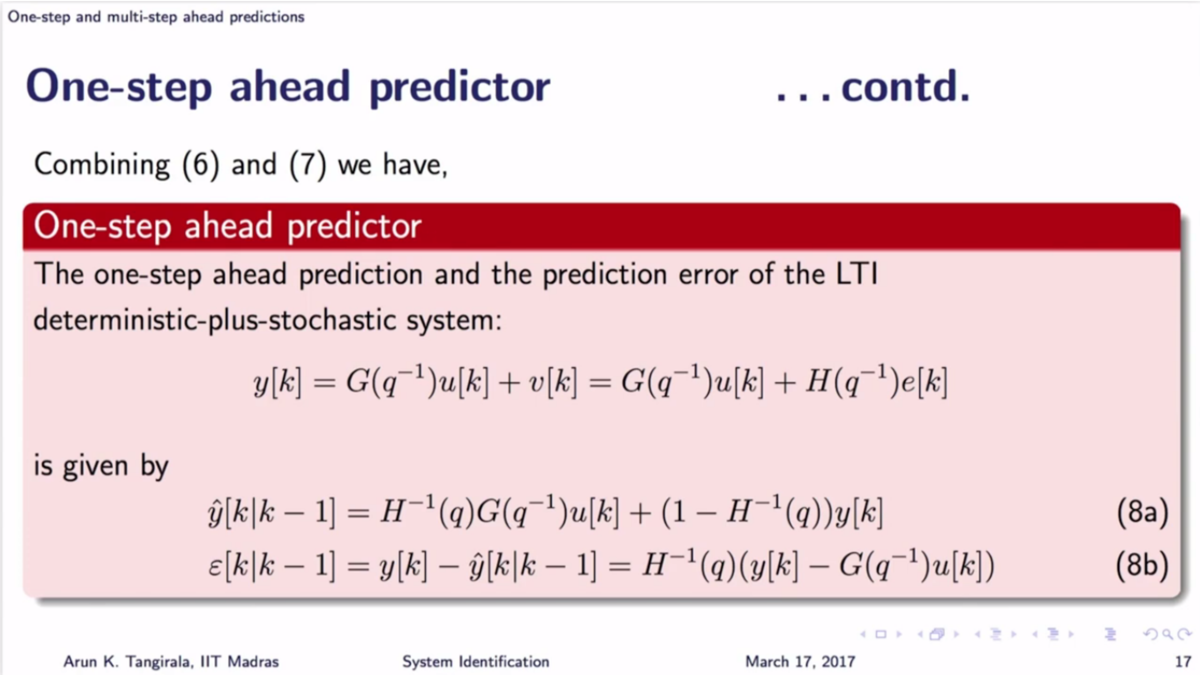
Remember the first coefficient of H is one.What will be the first quotient of h inverse?When I do a long division one, which means when it expanded h inverse in the polynomial form, infinitely long polynomial are finitely one. The first quotient is going to be one,which means that's going to be cancelled out. What about the remaining terms? They'll begin with q inverse, q inverse square and so on. So which means truly although I have written symbolically that vk is required here. Strictly speaking, I do not need vk. So when I expand this term, I will have only vk minus 1, vk minus 2 and so on. So you can straightaway see for moving average. Let's apply this formula here. If h is 1 plus c1 or let us say, I have AR it's easier to verify. What is the transfer function of AR, AR1, 1 over 1 plus d1 q inverse. What will be h inverse? One plus d1 q inverse, therefore when I work this out for AR1,I am lest with minus d1 q inverse, which means I am left with minus d1 vk minus 1, which is what we derive. So we're going to put this result together and derive the one-step ahead prediction. The one-step ahead prediction, therefore of y, the first time is deterministic. I don't have to worry about it. I'm given everything.G of q inverse uk. It's the second term that we have worked upon and we have gotten this result, 1 minus H inverse vk,but I mean given vk, now in SysID? In time series I am given vk, in SysID do I know v? Yes or no?

(Refer Slide Time: 23:07)



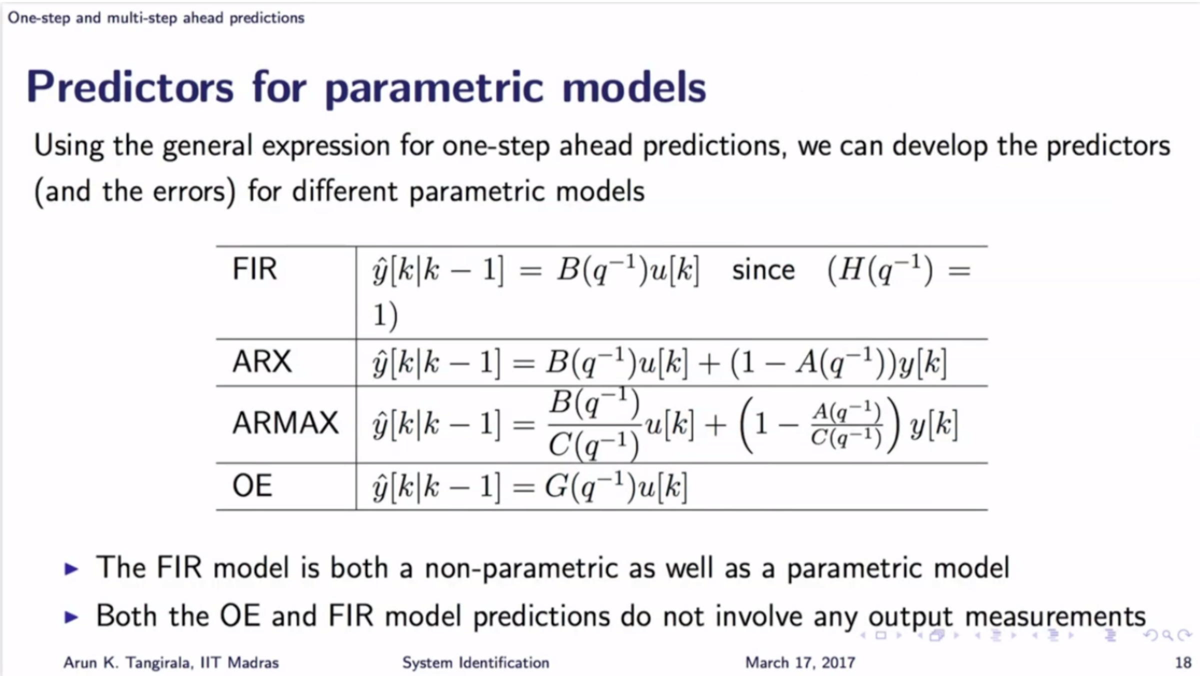
What do I know? I know input and output only I don't know v.What is v? y minus Gu. So I'm going to substitute that.Once I do that I get this beautiful result for one-step ahead predictor.

(Refer Slide Time: 23:23)



So I'm going to substitute 7 in 6.Because what is a goal? I want to write a predictor in terms of known quantities. What are the knowns?The model and the observations,vk is not known to me.So I am going to replace vk with y G and u,because they are known to me.Once I do that I get this result.So in place of vk,I'm going to replace with this expression here in equation 7. And once they complete that I get this expression for y hatin equation 8a.So I have y hat of k given k minus 1 as H inverse G plus 1 minus H inverse one as Etch in words G plus 1 minus H in red light. Again here it may seem awkward that I need wiki but I don't need like it to predict like it because I already know 1 minus H inverse y. Again here it may seem awkward that I need yk, but I don't need yk, to predict yk, because I already know 1 minus H inverse will not have the first coefficient.On the right hand side I will have G u and all the past wise. So let me close by saying here. If I were to apply this result. So what this result tells me is, if I am given G and H I can straightaway write a predictor, which is what I promised yesterday that we will have an expression for one-step ahead prediction for ageneric parametric model. So here I have y hat of k given k minus 1. What is G for FIR? FIR for G is B.And what is H? One. Correct. So when H is 1 H inverse is also 1, which means the second term doesn't participate and in an FIR the prediction is simply B of q inverse uk that means only inputs are involved in prediction of y, naturally because the FIR model expresses why purely as inputs.

(Refer Slide Time: 25:17)



Likewise for output error model also H is 1. But the difference isin the output error model I use G. Where as in FIR model I only use the numerator part of the B, because that's how the FIRis.Whereas for ARX and ARMAX and so on. All the past measurements will come in. So what we'll do is when we meet next class we'll complete this discussion on infinite-step ahead prediction. We are talking about one-step ahead prediction. We'll talk about L-step ahead prediction and infinite-step ahead prediction and we'll notice a very beautiful result, which will tell us again something new about this model structures that we have studied.By the way we have studied IDpoly in MATLAB if you recall. Now you understand what those ABCDF and so on. Okay.So when we come back next week we'll complete this discussion. Thank you.