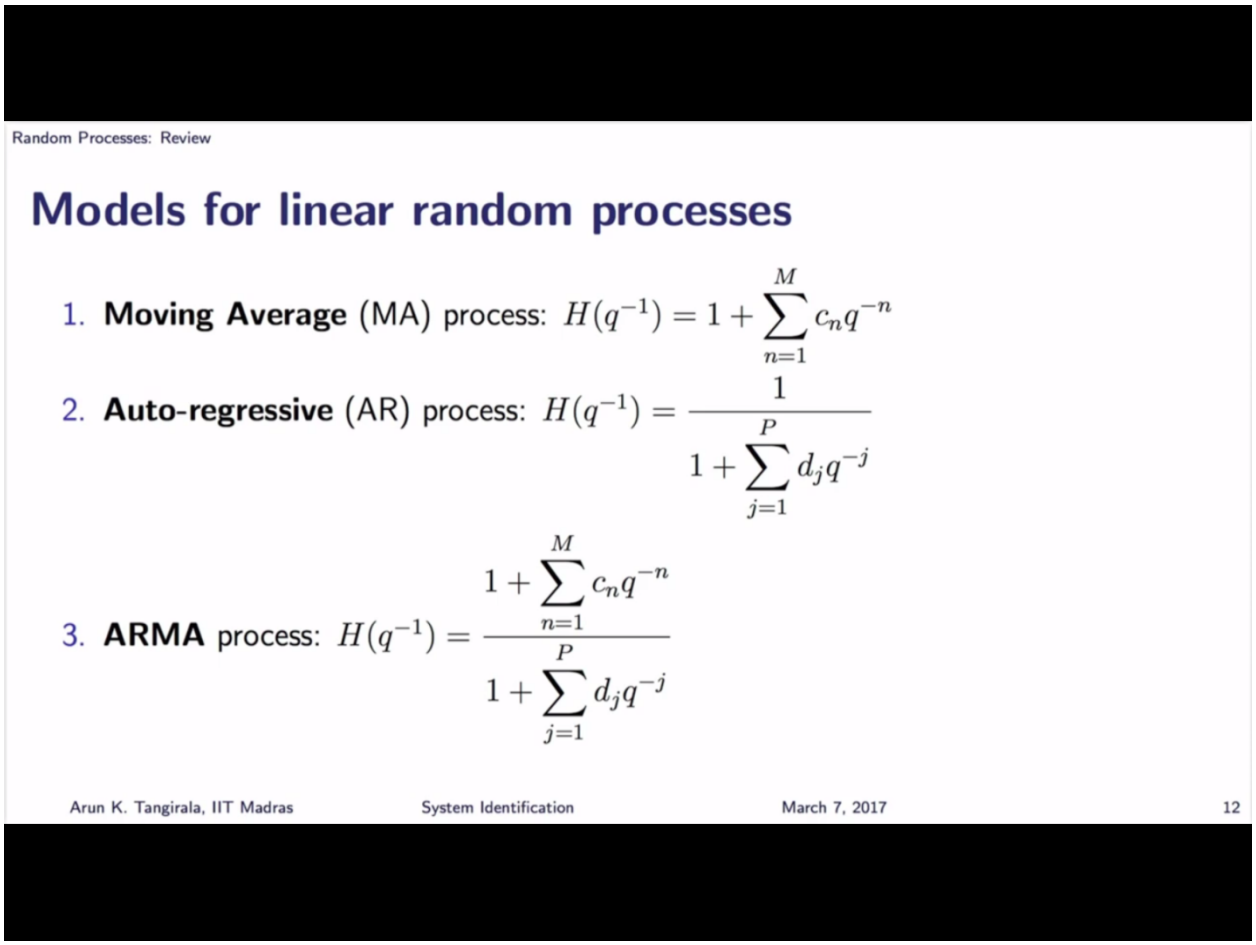
Alright, so now let's move on and look at what are the different models. Now the entire linear systems theory that we have can be brought. Whatever, we have learnt under LTI systems can be brought, only keeping in mind two things that the input to this stochastic process is white, it's a stochastic signal and two it is endogenous, but otherwise, the entire linear systems theory now can be brought in. In the linear system theory for deterministic processes, we learnt two models, right, one is an FIR model and other is a difference equation model, and other is a difference equation model, exactly the same apply here, but the terminology is different, that's all. People working in random processes perhaps or didn't talk to people in linear systems theory, therefore they came up with their own terminologies and that continues to happen even today, in fact the same course called fluid mechanics. You will see in different departments, and you will say introduction to fluid mechanics, foundations of fluid mechanism, basics of fluid mechanism, almost all the same. You will still see the same continuity equation, same Navier-Stokes equation haunting you forever, right. I find them to be demons, okay, particularly the Navier-Stokes equation. There are people who love and they eat Navier-Stokes equation for breakfast, but I don't... I mean, I consider that as a very-very bitter pill. Okay, so here the terminologies are different, but if you see the parallels there, you look at moving average form for example, you can see straightaway an FIR form. What class of processes have this representation. Well all those random processes, whose impulse response coefficient, this HS are now the impulse response coefficients, whose impulse response coefficients died on. Alright, in fact we will give a different definition for moving average process in terms of ACF signature, so we will again comeback to ACF, but let's finish the discussion here. Autoregressive processes are your difference equation forms essentially, I am giving you the transfer function operator, but you can straight away see, it is a difference equation form, right. And for what class of processes does AR process, this AR model apply, well all those processes, which have DK infinitely long impulse response coefficients, and whose impulse response coefficients can be parameterized, same story.

(Refer Slide Time: 2:53)



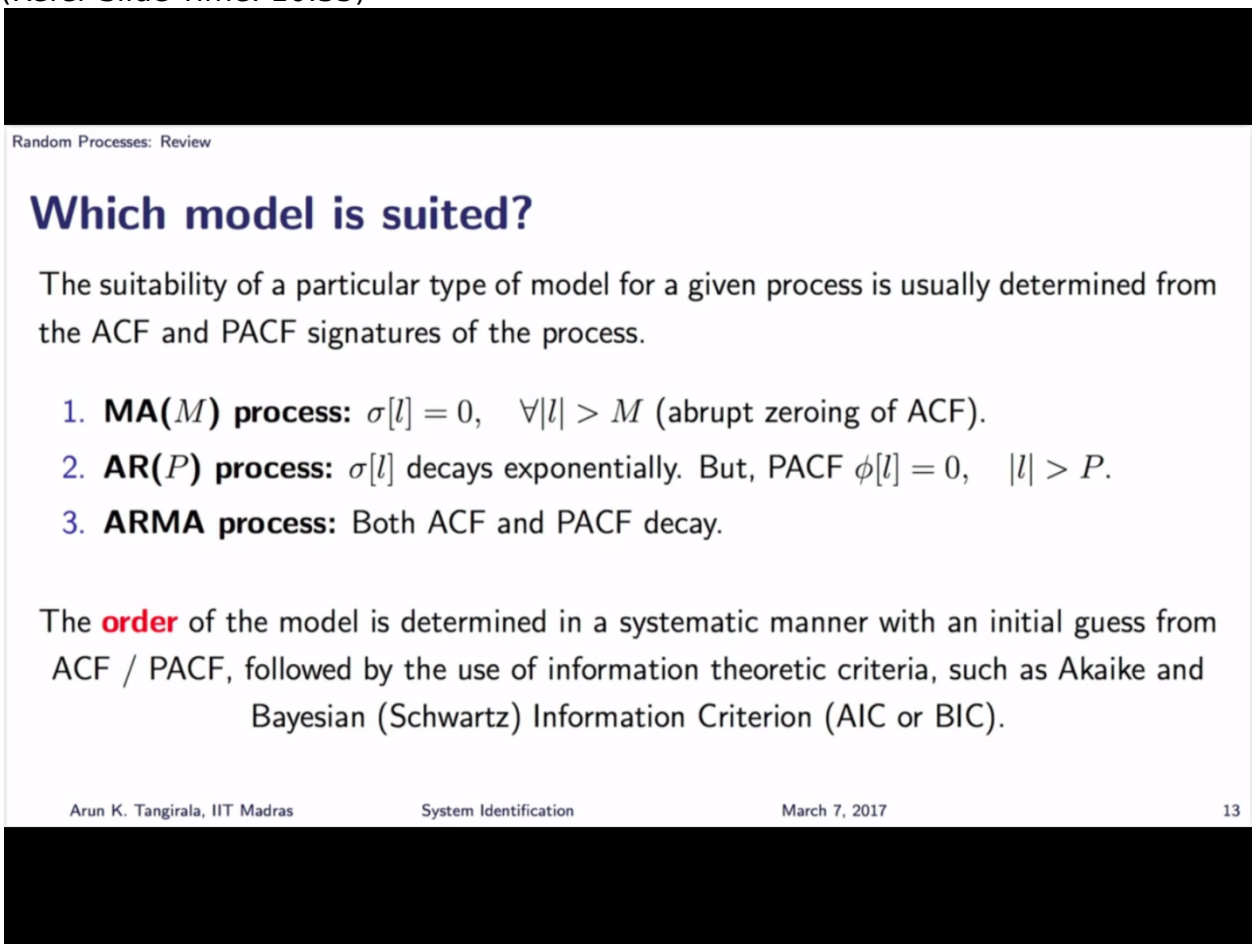
And just like we said, a general transfer function G of Q inverse can have a numerator and denominator, H call also can have a numerator and denominator. Such a process we call as ARMA, it is just falling short of Verma, okay. There is actually a process called VARMA process, but you can also call it as Verma. Verma stands for… VARMA stands for Vector Autoregressive Moving Average, where you have multiple random processes, being modeled at the same time, this is univariate ARMA process, okay. There are variations built on to this. Here also you can think of H of Q inverse containing an integrator. We talked about the presence of an integrator in G of Q inverse, right here also I can think of H of Q inverse continuing an integrator. There we said if the process has an integrator, it is marginally stable, it is not stable I a strict sense, same applies here, but we don't use a term stability, instead we use the term stationarity. So we say if H of Q inverse has an integrator, that means has a pole on the unit circle, it is not stationary, but it is a special kind of non-stationary process and the name given to such processes are random walk processes, Brownian motion, special class… Brownian motion. We will talk about that and then if you go to the time series literature, you have many other names SARIMA, GARIMA and so on. Very beautiful names, right. Alright, so these are the three different model structures that are available and they are no different from what we have seen for G, which is good news. I don't have to remember separately. The form looks exactly identical. The coefficients are different and there is one more difference, what is that? If you look at ARMA, compare to general G of Q inverse, I have spoken about this before… earlier, what is the difference? When you

write general G of Q inverse with numerator and denominator, versus a general H of Q inverse with numerator and denominator, and leading coefficients, that in numerator and denominator for H is 1, again that comes, you have to ask, keep asking why, because we have fixed H of 0 to be 1, the first impulse response coefficient to be 1. We have not done anything of that sort in G, the first impulse response coefficient in fact can be 0, and the other thing that you should note is… so while that is the important, one important difference, a consequence or a corollary of that is that there is no provision for any delay here. Whereas in G of Q inverse, you can say the first few impulse response coefficients corresponding to the delay period, they are going to be 0, that means your numerator, the numerator in G can begin with Q inverse – D, sorry Q inverse raise to D or Q to the min D. Whereas no such provision exist in H of Q inverse, why is that? As I have said EK itself is fictitious. In a fictitious world, what do you mean by delay fiction, that's a informal way of arguing it, spectral factorization result will tell you that it is impossible to resolve the face, whereas here it is possible to resolve the face in G, okay. So there are these striking difference that you should remember, but otherwise, the theory remains the same, I don't have to worry about this.

The question now is, which model is suited for a given series, so I have given you time series and I say, go ahead and build a time series model. What does it mean, you have to guess whether an AR model is suited, an MA model is suited, or an ARMA model is suited. You do not know, you will… you will never know, the process.. the signal is not going to come out and tell you, hello I am AR, no it's not going to tell you that, you are supposed to figure out, right. How do you go ahead and figure out, of course assuming that the given series satisfies the conditions, again what are the conditions under which you can give an LTI representation. First… first requirement, stationarity. So you have to plot the series to begin with, to check for any… visually do you notice any non-stiationarities, what do you mean by visually checking for non-stationarities? See for any trends, right, see for any trends or you see portions of data with completely different fluctuations, variability, that's an indication of non-stationarity and so on. So those are visual ways of checking for non-stationarity, strictly speaking you should perform statistical tests on the given series for stationarity. We will not discuss any of that at the moment. There is one test that you can perform, which is called a unit root test, that means presence of unit, root, or the pole on the unit circle, which we will talk about a bit later. However, at least to begin with, you should plot the series, check for any visually obvious non-stationarities. At the moment, if you don't find anything, assume it to be stationarity, later on I will show you how to do things formally, okay. So that is the first step you have to do for the assignment question. You have to plot the series, examine the series carefully. Now having convinced yourself that the series is stationary there is yet another condition, remember I said… I keep saying there is a condition on the spectral density, which there is no way of verifying, I do not know… theoretically the series… the process is suppose to satisfy a condition on its spectral density, but that is only telling you what are the boundaries of your framework. Can I verify that in practice, unfortunately no. So what if it doesn't meet the spectral density requirement, and I go ahead and fit this time series model here, well I am going to build only an approximation, and that is what is mostly the reality. You will never know whether the given process actually satisfies a condition of the spect… on the spectral density. How will you know that given process is linear for example, even for a deterministic process, suppose you go ahead and fit

an LTI model, may be from physics you can say, but suppose I don't... I don't know enough about what's happening in the process, how can I claim that it is linear, how can I test for time invariance, linearity may be a bit easy, time invariance how am I going to test? I don't. So what happens when I build an LTI model for a given process, for which I do not know, whether it satisfies the conditions of LTI, well I am going to build an approximate form, that's all. So that is something to remember. We seek to build an optimally approximate model, but that is all system identification is about. You only build optimal approximate models. You may think it is obvious today, but apparently until 60s 70s, people were seeking to build exact models, now you will say, are you crazy, you are going to build an exact model, how is it possible to build an exact model. Any process for that matter is nonlinear. Even if you are to build a nonlinear model, how do you know it's exact. So then it will say, forget it, we are going to build only approximate working models that do a good job. Unless of course, you are doing grey box modeling, different story. Okay, so now.... Given time series, whether I should build... ARMA, here is where we turn to some signatures that this process have.

(Refer Slide Time: 10:35)



In the deterministic world, did we ask this question, how do I... if I have to build an FIR model, how do I know if I have to build an IIR model, did we asked these questions, yes or no, what do you think? Did we asked this question? What kind of a G should I choose, given Y star and U, did we ask this question? In the liquid level system case study, did we asked this question? Think about it. So what do you

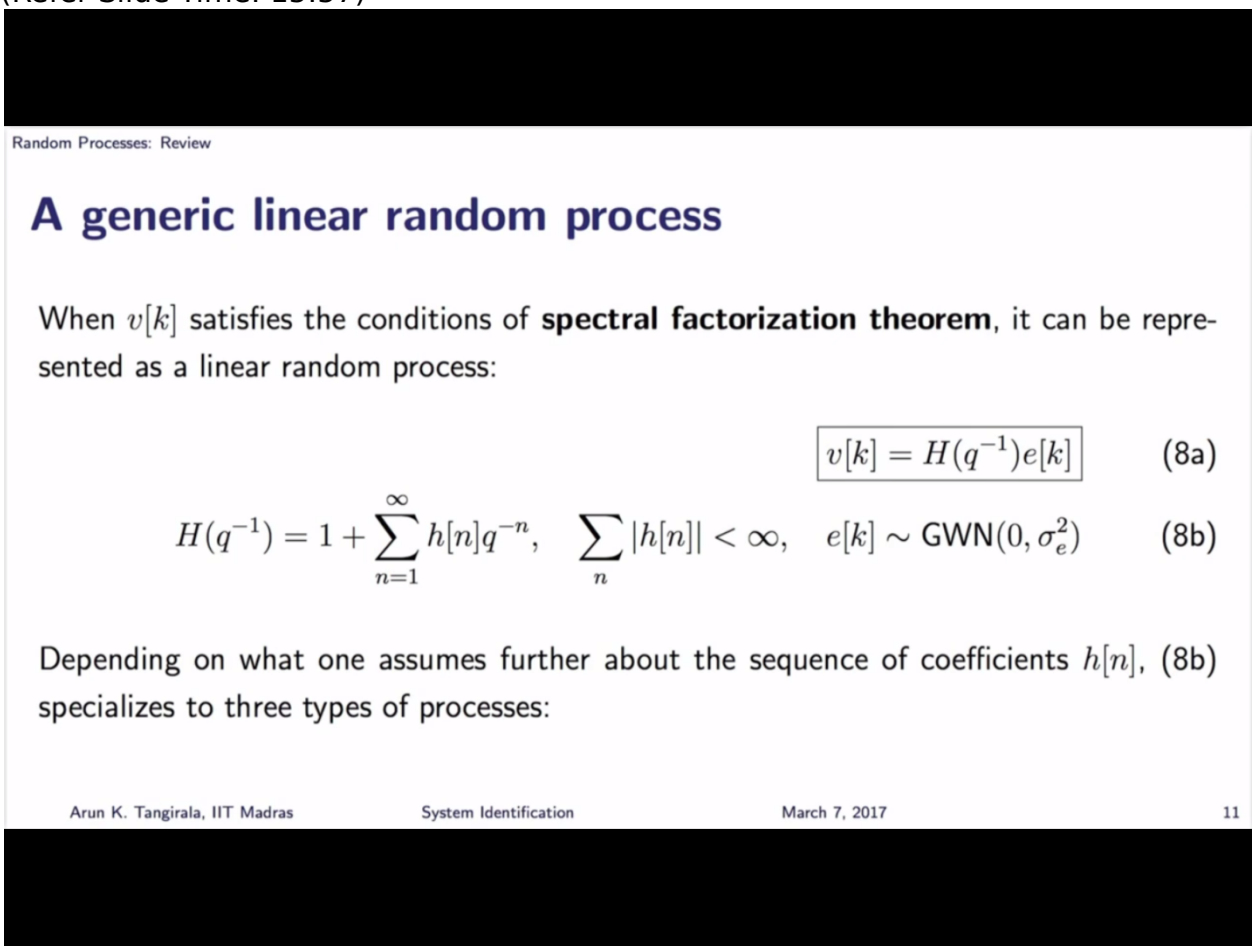
think, did we asked this question for the liquid level case study? Should I build an FIR model, should I be happy with an FIR model, or should I go ahead and fit an IIR model? That means a transfer function or difference equation type. Did we asked this for the liquid level case study or not? We never asked, you think so… What do you think, you recall the discussion we had on liquid level case study, what is the first thing that we did after we were given the data, we estimated a non-parametric model, we estimated impulse response coefficients, right, we plotted them. At that point what did we realized, that means the impulse response coefficients, there are too many I have to estimate and it is decaying, and that is when we decided, oh this is just too many unknown and that there is some pattern to this and I am going to fit a parametric model. So indirectly we did ask that question, although we didn't formally ask that question, we did make an attempt to figure out, if an FIR model was suited. Supposed it turned out that the impulse response coefficients, only three of them were non zero, it decayed very quickly. Then fine, right, then I can live with an FIR model, I don't have to worry about fitting any other model, because this three unknowns is good. There we worked with impulse response coefficients, right. How are we able to estimate impulse response coefficients. Why am I saying that now we have to turn to ACF here, that's a very beautiful point there. Again it takes us back to the same reason, but if I were to ask you this question, why am I not saying that look at the impulse response coefficients and figure out if an AR, or an MA, or an ARMA model is suited, why am I saying turn to ACF? There we specified everything in terms of impulse response coefficients, right, our estimates directly were impulse response coefficients and so on, why am I not giving such a recommendation here, my… why my prescription is not based on impulse response coefficients, but rather ACFs.

(inaudible)

Correct, I do not known the input, so I cannot es… estimate H. I have to take an alternative root. So if you look at this model that we had, right, if I look at the model, this model, yes relates V to E, good, can I use this to estimate H, I cannot, why? because I do not know E, what do I know about E? about its correlation structure. I do not know the values of E, whereas in the deterministic world, I knew the input, so straightaway I can estimate G, and then I can look at the impulse response coefficients and figure out if an FIR model or a difference equation model, that is an IIR model is suited. I do not have the luxury here. Look at how many differences exist or arise the moment you do not know the input, right. So in life when you do not know the causes, it's not so easy. We simply try to find the cause and keep blaming it, but that is also may be fiction, the actual cause may be internal, endogenous. So when you do not know what is causing some phenomenon, it's a big challenge, in fact sometimes it's called blind identification. So you can think of time series as being blind identification. Therefore we turn to auto correlation functions for telling us whether an MA model or an AR model or an ARMA model is suited, as I said, because I do not know the values of the white-noise input. I know its correlation structure, so I am going to re-write this equation in some sense, which relates V to E in terms of auto corre. In other words, I am going to ask if it was an MA model, how would the ACF of V look like, if it was an AR model, how would the ACF of V look like, or if it is an ARMA, how would the ACF of V look like, why because in practice I estimate the ACF, I am not given the expression for estimating ACF, but theoretical definition has been given, we will learn the

expression for estimation a bit later. There are routines in MATLAB that will do for you.

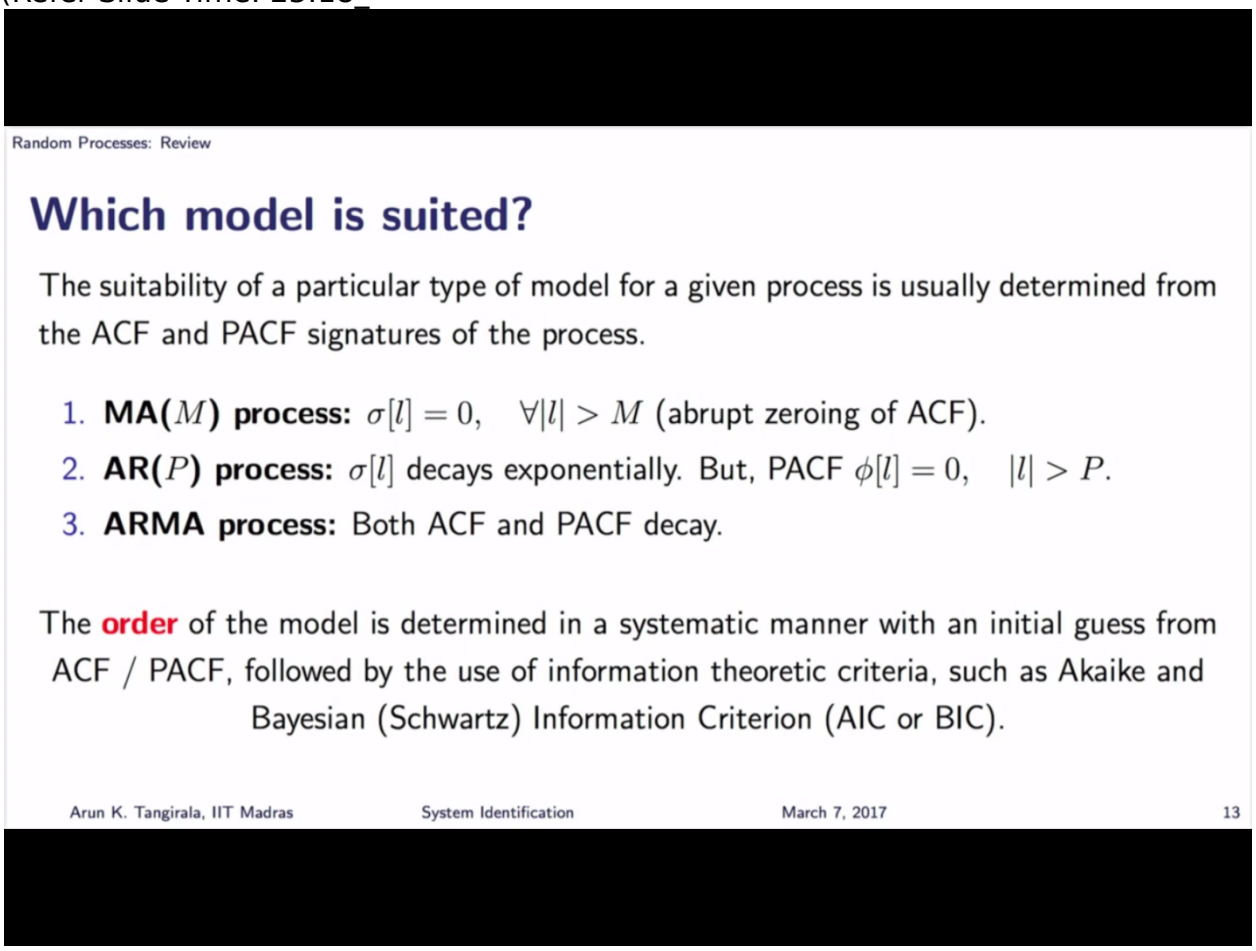
(Refer Slide Time: 15:57)



But the fact is I can estimate ACF, I can plot the ACF of V, and then figure out now whether an MA or an AR or an ARMA is suited, that is the hope. So these are the signatures that ACF leave behind. I will quickly go through the theoretical analysis of ACFs for each of this. The first thing… the first point there is that if the ACF goes abruptly to 0 after a certain finite lags, then it's an indication that the underlying process is MA, okay, so it's more or less your condition or impulse response coefficients. In the deterministic world we said if the impulse response coefficients go to 0 abruptly, after some lags, then it's an indication that it's an FIR model, but there it is impulse response, here it is correlation coefficients. The second point is, which is that for AR processes, if the underlying process is an autoregressive process of order P, by the way those M and P refer to the orders. If the underlying process is AR of order P, then you can show that the ACF decays exponentially, but it doesn't tell me what is the order, whereas in the first condition you see, it tells me both, whether it is an MA process and what is it's order, because the ACF goes to 0 exactly after M lags. Second condition says only ACF dives down exponentially, that's only an indication, it's AR, but what is the guess for P, order, I need to know, right. In the deterministic world we tried this by trial and error, but fortunately here I have even a recommendation for the order, so that's a nice thing. In the deterministic world I don't have any recommendation for the order, I have to figure

out by trial and error, but of course we learned the state space method to figure out the order, here also you can use it. Anyway so coming back to the point, if the underlying process is autoregressive and of order P, as... regardless of the order, ACF will dive down exponentially only, again stationary processes, but there is something called partial auto correlation function. What is this partial ACF, it is build on the notion of partial correlation. What does ACF do? What does it measure? It measures the correlation between any two observations, separated L samples apart or L sampling instants apart. What do you think PACF would do? ACF measures a correlation between two observations that are positioned L instants apart in the signal, what do you think PACF, partial auto correlation function would do? What do you think it will do? What did partial correlation do for us? Condition not normalize. If I have two variables X and Y, I am measuring correlation that standard correlation, but suppose I believe Z is connecting X and Y, right that is when we introduced partial correlation, if Z is influencing both X and Y, then the correlation between X and Y is influenced by that confounding variable, correct, unless I remove the effects of that confounding variable, I wouldn't know, whether X and Y are directly influencing each other. Here in the random signal world, when I am looking at two observations, and I am looking at the correlation, is there a chance that the intermediate observations are also contributing to the correlation? Right, if you take AR2 for example or even AR1, look at this, suppose... you just take this AR2 model, yes there... VK-1 as per this VK-1 and VK-2 are influencing VK directly, but do you think VK and VK-3 will be correlated? How, indirectly, because VK-1 is being influenced by VK-3, right. VK-2 is also being influenced by VK-3, and VK-4 and so on. So in all autoregressive processes, this is the classic nature that although the observations beyond the order are not directly influencing VK, they indirectly influence VK and that is why ACF is unable to tell me what is the order. Partial auto correlation function accounts for this, it says do not evaluate a plain vanilla correlation. When you are evaluating auto correlation at lag 2, take into the possibility that VK-1 is influence... is connecting them to. So your confounding variable is VK-1, when you are looking at coronary artery disease between VK and VK-2. So you calculate that, you calculate now the conditioned auto correlation, which is the partial auto correlation. When you move to lag 3, what you have to do, I have to worry about the possibility of VK-1 and VK-2 linking, VK and VK-3, those become the confounding variables, as I keep moving at higher and higher likes, I have to keep including more and more confounding variables, but once you account for this confounding variables and compute the correlation, what you are computing is partial auto correlation, that is what is PACF. So here suppose I removed, I accounted, let us say for effects of VK-1 and VK-2 and then worked out the correlation between VK and VK-3, I should theoretically expect the correlation between the conditioned VK and VK-3 to be 0, because I have already taken into account the fact that there is a link. How I am doing it mathematically and so on, we will not necessarily discuss here, but I will give you a procedure to do that shortly. But that is what the beauty of PACF is. The PACF is ideally suited for analyzing autoregressive processes, because just like the ACF goes to 0, after finite lags for MA processes, PACF goes to 0 exactly after the order, that is what the condition says here. We denote the PACF with the FI. FI at lat L goes to 0, which is... FI is the PACF, exactly after the order, correct order. What is FI measuring? Direct correlation. So directly VK-3 won't effect, if it's an AR2, not only VK... VK-3, but all the other observations. So by plotting the PACF, I will show you the plots

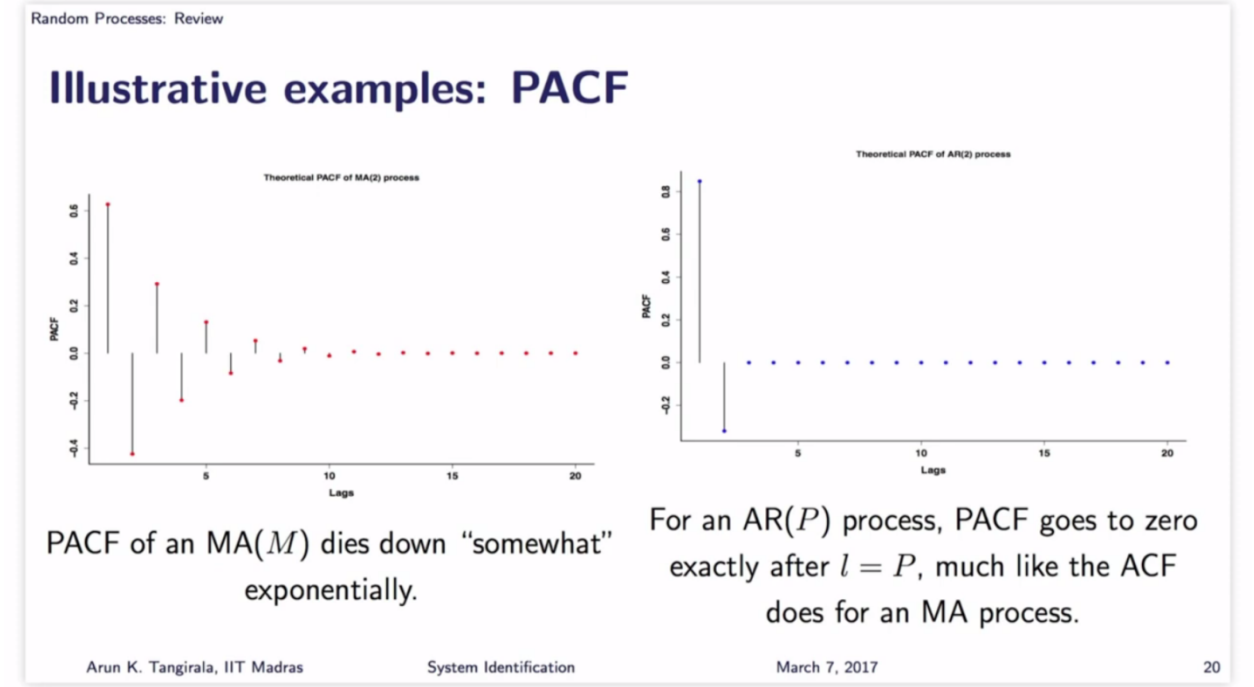
very quickly, you can figure out whether the process is an AR and if it is, what is the order?
(Refer Slide Time: 23:18_



A decent guess can be obtained, don't expect gold mines here. In practice you are going to have finite sample effects and so on, but this is all theory. Now when it comes to ARMA processes, unfortunately there exists no clear cut signature, all you know is ACF and PACF decay. What are the orders of the autoregressive and moving average parts, that you have to go by trial and error, okay. And then you use criteria like Akaike information criterion or BIC and so on to figure out what is the order. So that has to be done in a somewhat trial and error manner. People have attempted to come with, come-up with measures, some kind of signatures and so on, but they have all failed, failed in the sense, theoretically they looked okay, but when it come to using it in practice, they went out of the stadium, okay, so they did not really workout well. Let me first show you the figures and then I will come back to talking about how the ACFs are calculated. So here is an ACF of, on the left hand side you will see the ACF of an MA... M process, what do you think is the order? 2, right. Because it exactly goes to, these are theoretical ACF I am plotting, it's not an estimated one. In the case of estimates, they won't go to 0 identically, there will be some small values. And you have to conduct a significance test to figure out beyond what lag things are small. On the right hand side you will see the ACF of an AR process, where the ACF is decaying exponentially, that's all you can say, from this can you say, ah! I can see this straight away that it is suppose to be AR2, I can't

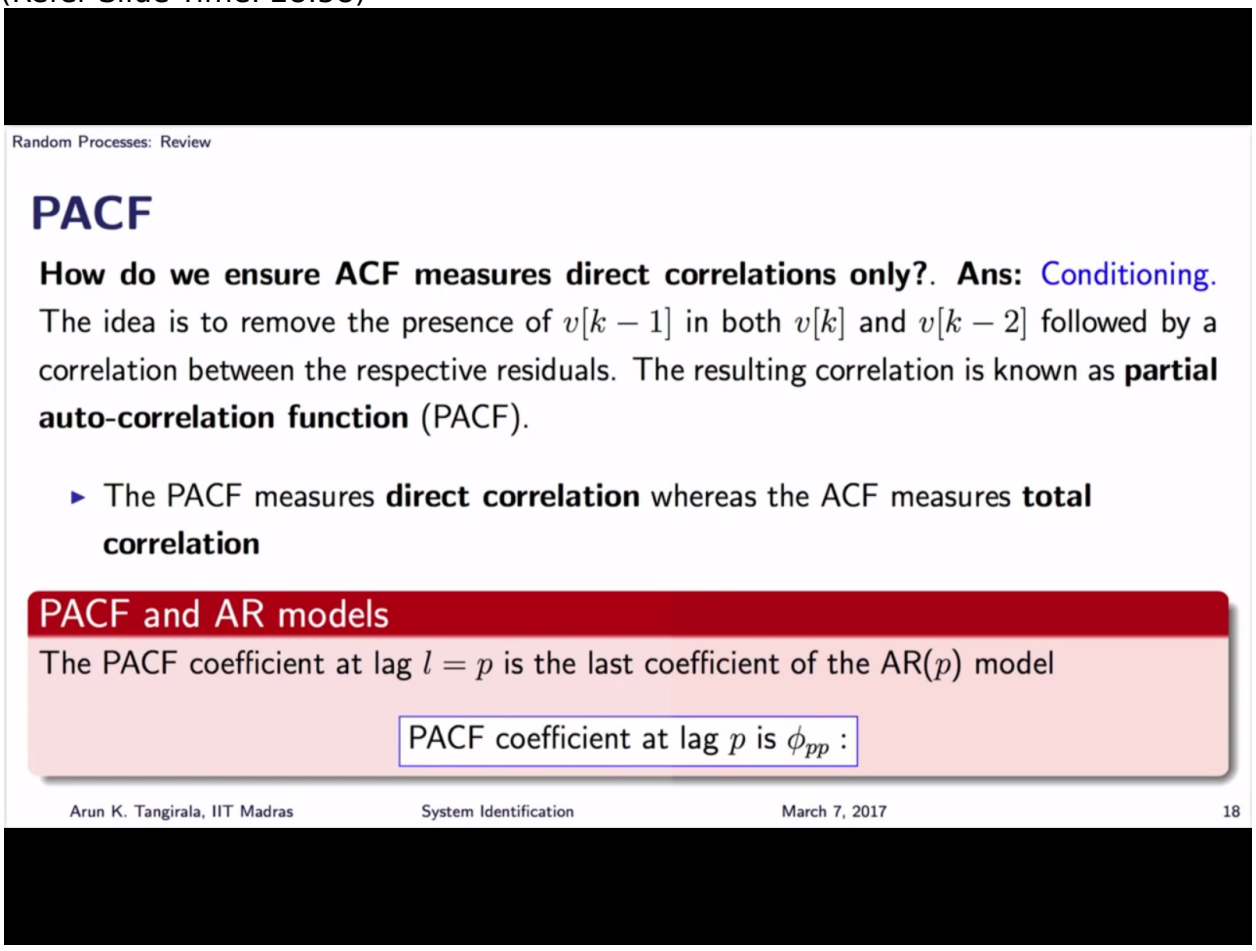
say that, okay. On the other hand you look at PACF, come back to the definition later. The PACF now for an MA process versus a PACF of an AR process. For an MA process PACF goes down exponentially. So it's not so well suited, I am not worried, you know, I am not any way going to rely on PACF. How and why PACF diced on exponentially for MA processes, we may not discuss, but you can think of the duality, what ACF... how ACF kind of behaves for AR, PACF behaves not in an exact way, but in an analogous way for MA processes, for an AR process, what do you think, so it is an ARP process, what is the order, read carefully, look at the lags, I will zoom in for you, I am not so sure, how well you can see, I am... the starting point is not 0, unlike ACF, okay, the starting lag here is 1. So what do you think is the order? 2, because after lag 2 the PACF goes to 0 theoretically, its... so the underlying process is an AR2 structure. I am not showing you how things look like for ARMA, but as I told you it will decay. For an ARMA, so if you were to figure out now what kind of model is suited, you have to look at the ACF and PACF.

(Refer Slide Time: 26:32)



The assignment is self equipped, that means you are going to write a code for PACF. ACF, you can compute, I can post the code for ACF or you can write your own. There is a function called XCOV, I will talk about the MATLAB commands may be in tomorrow's class, but there is, if you are in a rush to solve the assignment, then there is a routine called XCOV, in fact there are two routines XCOV and XCORR, you have to understand the difference in MATLAB, XCOV computes the cross covariance between any two signals. You can use it for auto covariance as well and it follows

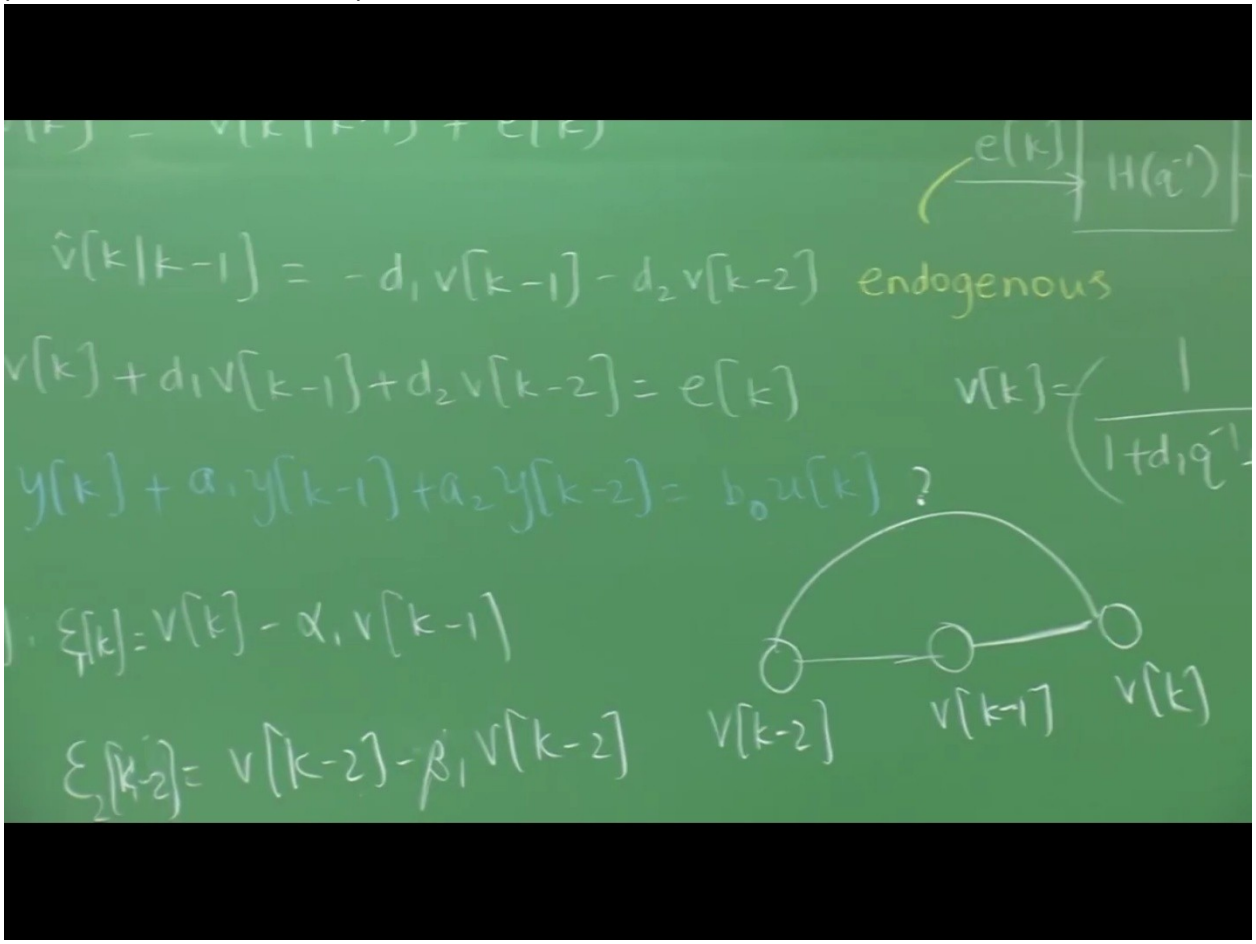
the definition that we follow in statistics, that is exactly what we are using. That means it removes the means and then computes a covariance. There exist another definition of covariance in statistical signal processing, which unfortunately does not remove the mean. Of course signal processing may say that I am unfortunate to work with a statistical definition, but it's okay, it's mutual. So it's relative. The definition of... sorry... covariance in the statistical signal processing world differs from the one in statistics, in that, it doesn't mean center, when it compute the covariance. Remember we have expectation of X-Mu times Y-Mu Y. There it straight away, expectation of XY. So it doesn't take into account expectation of X and exaction of XCORR works with that definition, so you don't want to use that, use XCOV and figure out how to do the ACF, but if you have difficulty, I can always post the ACF code, it's a very, may be three or four lines. PACF, there is an algorithm for computing PACF and that is given by the Durbin Levinson's algorithm. I will just quickly talk about PACF and it just takes a minute. We have just now said PACF measures direct correlation. How does it do that? Theoretically what it does is, it conditions, meaning it removes the effects. So here suppose I am looking at PACF at lag 2 so FI at lag 2, theoretically what it does is, it removes the effects of VK from VK-1, so it constructs... constructs a residual, let's say... call this is Epsilon K.
(Refer Slide Time: 28:58)



Epsilon 1 and then there is an Epsilon 2, which removes the effects of VK-1 from VK-2, so what you are doing here is you are constructing two new series, one... one series, which actually removes the effects of VK-1 form VK and then another series,

which consists of removing the effects of VK-2, you can say Epsillon K-2, it doesn't matter, anyway dummy variables. So I have here Epsilon 2 K-2, which removes the effects of VK-1 form VK-2, so what is happening here is you have VK-1, VK-2 positioned here, VK-1 positioned here, and VK positioned here, we don't know how they are connected, whether this also exists is what you are asking, this is what PACF is asking, whether there exists a direct pathway, okay. So what it does is, it removes essentially, symbolically these two connections, it is severing those connections and then looking at the correlation.

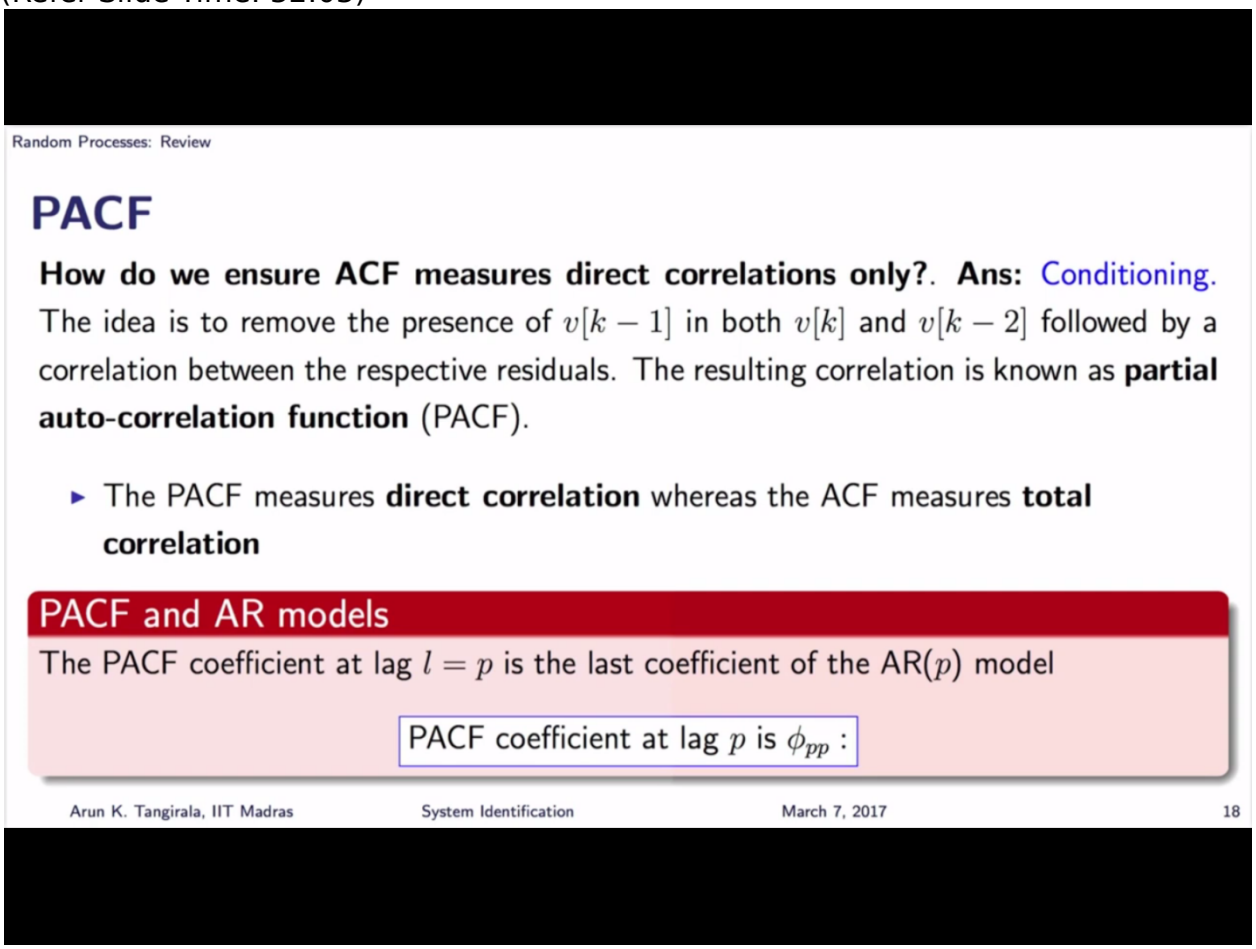
(Refer Slide Time: 30:12)



Now the practical way of computing PACF relies on a… another theoretical result, which is that build AR models of successive orders, that is you go ahead and build an autoregressive model of order 1, of order 2, order 3, and so on, continue in this fashion. Every time you build an AR model, look at the last coefficient. For example you build an AR2 model, you would have estimated D1 and D2, the last coefficient is D2, that is nothing, but the PACF at lag 2, you can show that, I am not proving that, I prove that in a time series course the PACF at any lag L is the last coefficient of the ARL model. In fact since we use this notation here in practice the PACF at lag 2 would be –D2, because I am of the notation that I have used. And PACF as lag 1 is not –D1, that's a big mistake. What would be the PACF at lag 1?

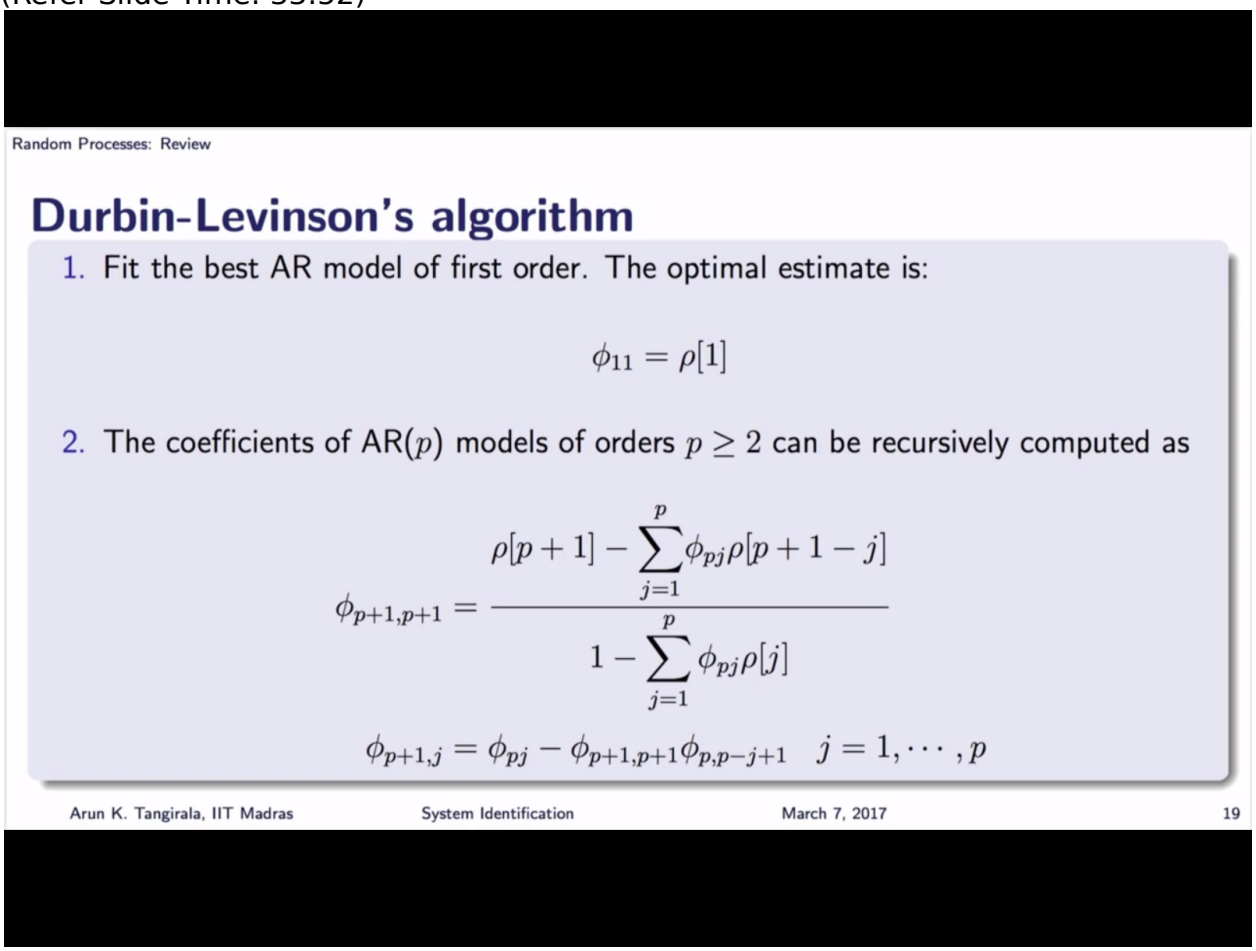
(inaudible)

No, no it won't be at lag 1, there is nothing to confound, PACF at lag 1 is simply the correlation between VK and VK-1. What is PACF, it is also correlation, but after removing the effects of confounding variables. When I am looking at PACF at lag 1, I am looking at PACF between VK and VK-1, there Is nothing confounding VK and VK-1, they are just successive observations. So the PACF at lag 1 is nothing, but the auto correlation at lag 1. In terms of the result here, you fit an AR1 model, the last coefficient, that is only one coefficient, that is the PACF.
(Refer Slide Time: 32:03)



Now Durbin Levinson's algorithm offers a recursive way of computing AR models of successively increasing orders. Instead of you going about all the way, building AR models again and again and again, it says, if you have build an AR1 model, I will tell you how to recursively compute AR2 model, how to recursively compute AR3 model. So it's a very computationally friendly algorithm and the code is given here, as you can see here the PACF at lag 1 is initialize to auto correlation at lag 1. And thereafter the PACFs at lag 2 are given by this FI at P+1 P+1, I am using subscripts here, whereas I have used square brackets on the board and so on, but you should read them. The FI at P+1, P+1 is nothing, but the PACF at lag P+1. In order to compute the PACF at lag 3, you would need the other coefficients in the AR2 model. So to summarize, you kick start your algorithm by computing auto correlation. Take the first coefficient that is PACF at lag 1, PACF at lag 2 is computed using this, which is the last coefficient of the AR2 model. When you want to compute PACF at lag 3, of course you come back to the same algorithm here, same expression, but you

need these coefficients as well. So in step 2 you would compute FI 2, which is minus D2, but when you want to compute FI 3 you will need D1 also and that I why you have the bottom expression, and so on. So you will need the coefficients, other coefficients of the AR model.
(Refer Slide Time: 33:52)



So you just write, it's in fact may be a two or three liner algori... routine that you can write for PACF and that is what your assignment is asking you, that's it, so this how we compute PACF and this is how of course ACF, I have told you, use XCOV. Remember XCOV there is an option to ask for auto correlation or auto covariance. Lookup what option the... what is the correct option you have to give, okay. You can... you can ask for covariance as well as correlation, that's it, so apply that and... and first test out your algorithm, sorry... your routine on sample processes and may be if I get an opportunity, I will show you how to work out on a sample process. I have written my own code for PACF, I will show you how it... how to simulate an AR process or an MA process in MATLAB and then you can get going. So tomorrow we will talk of cross covariance functions, what is the role in society and then talk of spectral densities. And that hopefully we will draw curtains on the review of random processes.