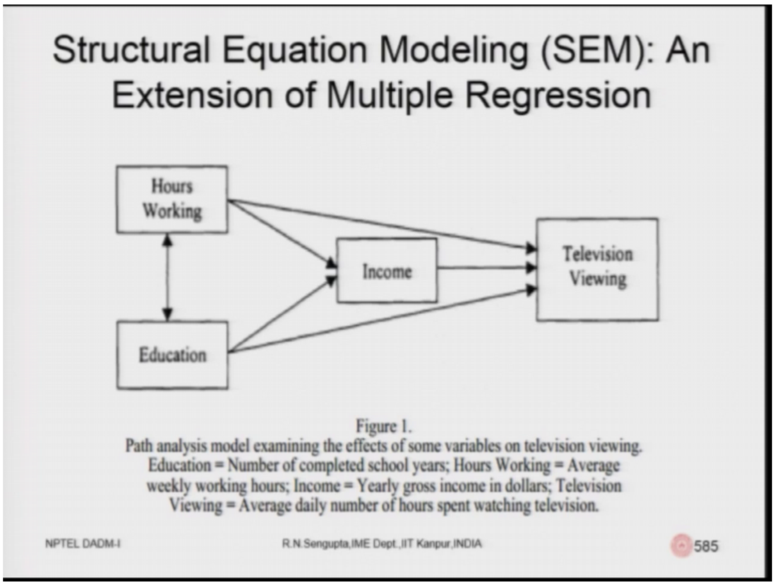
Data Analysis and Decision Making - I
Prof. Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 59
SEM

Welcome back my dear friends and dear students, a very good morning good afternoon and good evening to all of you and this is the last, but one lecture for this NPTEL series of lectures for Data Analysis and Decision Making 1. And as you know this total course it was basically for 12 weeks 30 hours 60 lectures each lecture being for a half an hour and each week we had 5 lectures for half an hour each and I am Raghu Nandan Sengupta from the IME department IIT, Kanpur.

So, if you remember we were discussing about structural equation modeling and how they can be considered in trying to basically combine the latent variables explanatory variables and the relationship can be ongoing in a looping method; that means, the end result can give you the feedback for the initial set of variables, which are the independent variables. So, based on that let us proceed. So, consider the diagram here.

(Refer Slide Time: 01:15).



Structural Equation Modeling (SEM): An Extension of Multiple Regression

Figure 1.
Path analysis model examining the effects of some variables on television viewing. Education = Number of completed school years; Hours Working = Average weekly working hours; Income = Yearly gross income in dollars; Television Viewing = Average daily number of hours spent watching television.

NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 585

So, what we have? We want to basically find out the television viewing, number of hours or the type of programs or the age group who are basically watching the television and we think that the relationship are the main factors which a basically affects television viewing are the hours of working they are the people are working in industry.

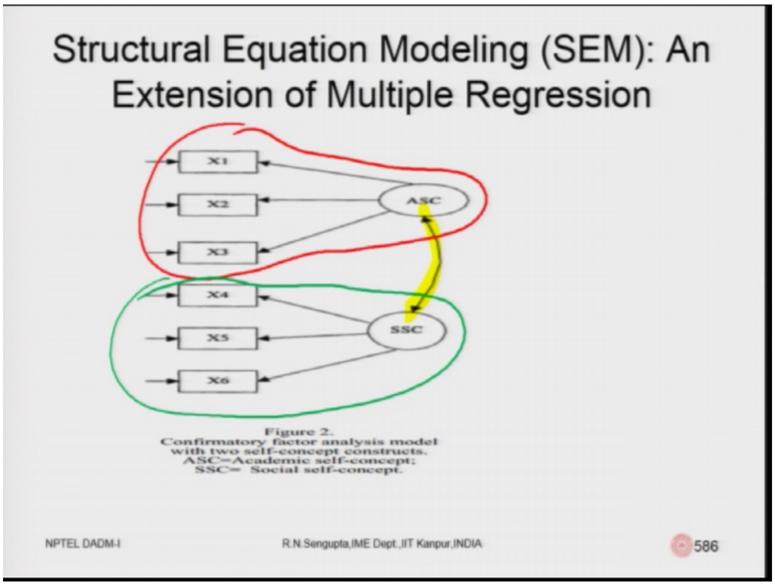
So, then more than number of hours they work less would be the television viewing; it will also depend on what type of program they see. So, that will depend on the education. So, it is more of house wife; they will be more interested in and watching programs relate to the family; if it is more of people who are working in the industry, in the business person, in the finance; they would be more interested toward financial programs. If somebody is from the education sector he may she be may be more interested to watch education programs and income is also a factor. Because more the income or more the number of leisure hours you have more number of hours you want to spend with your family, if it is your you are very specific about type of programs you want to watch with your son and daughter or your family would dictate the type of programs which you see.

So, if you consider this, number of hours working would basically have a direct relationship on the income. So, more hours you work will consider that the p packet or the remuneration switch you are getting is higher. Education would also have a higher implication for them for the income. So, generally we will consider more the person is educated, hired he is he or she is working and more is the disposable or income. Education and hours of working would have a in the relationship which is both way implication; that means, education being more, I would basically work less number of hours in more or more number of hours depending on type of work watch I am.

So, if we I am a I have my education level is less, I have to do more manual hours and work more using my physical labour which means I have to spend more number of hours working and this hours of working as I said would be related to income and hours of working more or less would have an effect on number of time, hours I am willing to spend for the television. Similarly for education would basically give me some relationship of other ideas of what type of programs I am going to watch; so, the path analysis which we have; models examining the effects of some variables on television viewing. So, education is the number of completed school years or number of college years somebody has spent, hours working would mean the average weekly working hours is the person has, for disposable at his or her um under control.

So, that he or she can decide apart from the family time, shopping time, going for a movie outside spending time with relatives, what is the number of hours you want to have to watch a television program? Income would mean yearly gross income in dollars or a number of rupees whatever it is, and the total disposable income which I have. And television viewing would being basically mean the average daily number of hours which I am which I spent or the person spends in watching different type of different type of programs, I am not talking about the quality or the different level of type of programs; it can be so, populars, it can be news, it can be analysis, it can be religious program it can be um natural program by either by history channel or it can be national geographic channel whatever it is.

(Refer Slide Time: 05:01)



Structural Equation Modeling (SEM): An Extension of Multiple Regression

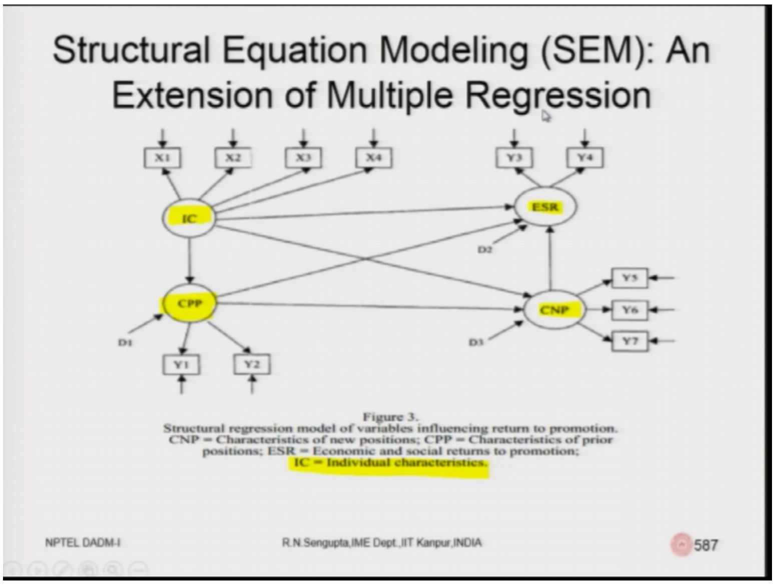
Figure 2.
Confirmatory factor analysis model
with two self-concept constructs.
ASC=Academic self-concept;
SSC= Social self-concept.

NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 586

Now, when you are doing the factor analysis which you have done; so, the confirmatory factor analysis model which self control constructs would basically be the social self concept. And what we mean that depending on the social status which you are and the academics self concept which you have, you will basically decide that what are the factors which will dictate this.

So, consider X 1 to X 6 are the variables it need not be X 6, it can be X 1 X 2 X 3 till X 7 X 8 or it can be less than that, and will consider that all of them are dependent in such a way or the internally I would not use the word dependency, but the relationship between the factors are such that X 1 to X 6 are the random variables, which affects the m the academic self concept and the social self concept. Remember the methodology how you basically decide and what are the 2 different sets. So, what we have is X 1 to X 3 are the academic related concepts and X 4 to X 6 are the social concepts, which we think affect oh I should use a different colour sorry for that. And the relationship which I have between the academic and the self would basically be given by a two way implications; that means, both are affecting each other and I want to find out the relationship. So, this can be one of the way how I can use structural equation modeling.

(Refer Slide Time: 06:40)



Structural Equation Modeling (SEM): An Extension of Multiple Regression

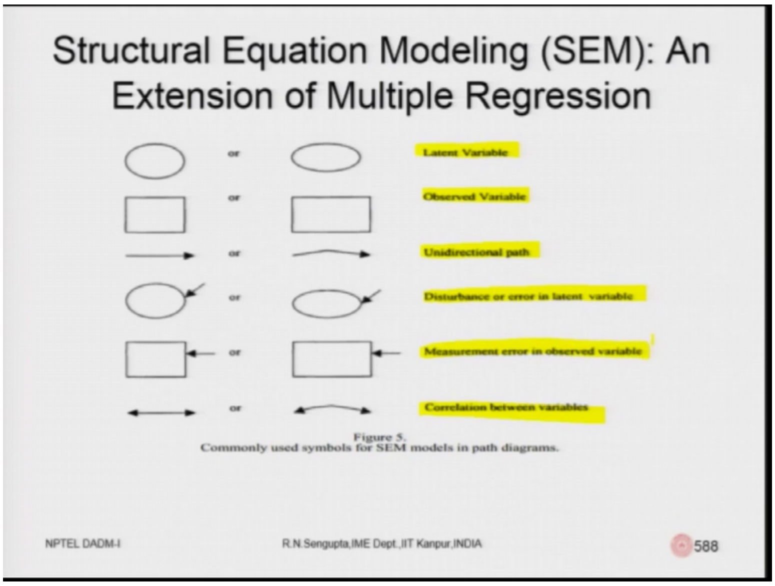
Figure 3.
Structural regression model of variables influencing return to promotion. CNP = Characteristics of new positions; CPP = Characteristics of prior positions; ESR = Economic and social returns to promotion; IC = Individual characteristics.

NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 587

Now, consider the structural equation modeling based on a different idea where you have the effect as trying to analyze that we have individual characteristics, which is basically the type of characteristic a person has what his or her habits of book reading, what is your habits of basically feeling viewing television programs, what is your habits of different relaxation, what type of work they do. You will basically the characteristic the prior position based on which you will be arrived on this results. So, they would be modeled accordingly and we are considering the X 1 X 2 X 3 Y 1 Y 2 Y 3 Y 4 Y 5 are the variables which are affecting in how I analyze the actual variables. So, this X 1 to X 4 or X 1 to X 5 or Y 1 to Y 4, Y 1 to Y 5 are the variables which are external through the system and internally we consider there are variables, which are and I am again repeating individual characteristics they are economic and social returns which are there their characteristic the prior positions which we have an information which have that characteristic the new positions which we want to analyze. These are the factors which would basically be influenced by the external one which will club as X 1 X 2 X 3 or the vector Xs and the vector Ys.

So, here also you see in an example, where the relationship between both the implication one way implications have been modelled such that it gives you the path analysis of the structural diagrams which basically exist between the random variables Xs and Ys and the constructs which we want to analyze.

(Refer Slide Time: 08:38).



Structural Equation Modeling (SEM): An Extension of Multiple Regression

or
Latent Variable
or
Observed Variable
or
Unidirectional path
or
Disturbance or error in latent variable
or
Measurement error in observed variable
or
Correlation between variables

Figure 5.
Commonly used symbols for SEM models in path diagrams.

NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 588

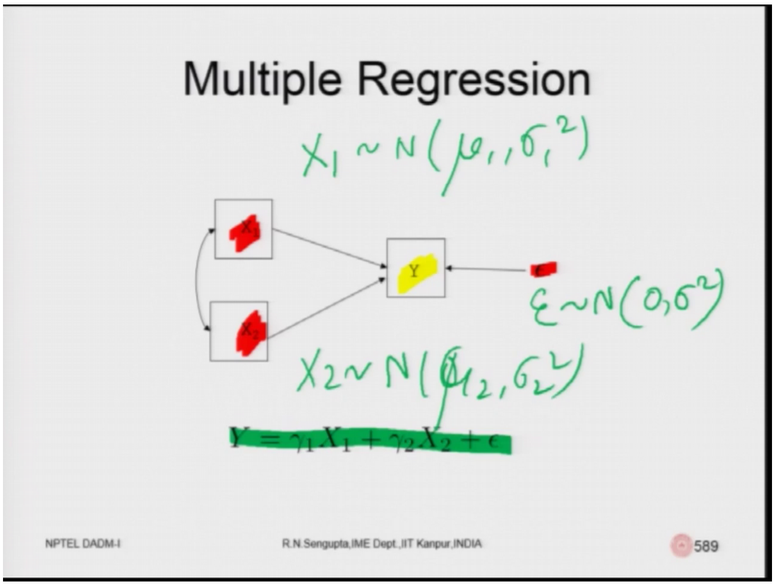
Now, generally when we are doing structural equation modelling, we have different type of variables as I mentioned observed variable, unobserved variable, white noises, latent variables, disturbances or the white noises measurement errors are there, errors which we are not under control what are the correlation effects which is happening between the errors between the random variables, between the explanatory variables between the random the independent variables dependent variables the general relationship is given like this. For the latent variables we use the concept of the oval, the observed variables which you can view are given by the rectangles, the unidirectional paths would give you the one way implications between the dependent and independent variables.

The disturbance would be which we coming outside or the errors in the latent variables would be given by with ellipse being affected by an external source and is one way implication only. Because you your readings would be affected by the external noises and not the external noises not being affected from your side because you are not interested to understand what is the effect on the external noises what your readings would be you are more interested to understand that how would the white noise is affects your latent variables. You will also have some observed variables being affected by the measurement errors.

So, consider the errors would be set of observed variables errors which are under your control which you can which you can take care of like say for example, the temperature vary variability when you are measuring some quality control characteristics, you can control them. So obviously, you will try to consider that variability in the temperatures coming from 2 sources one is under your control and one is basically the white noises which cannot control. So, like say for example, you want to understand the wind speed in trying to find out say for example, the temperature of a place.

So; obviously, if you close the whole space by a room or there is no effect on the winds; obviously, you are trying to decrease the effects of measurement errors to the maximum possible extent, but; obviously, there would be a white noise. And the correlations or the relationship with yes exist both ways would basically be given by a both way implication curve as given here. So, you will basically have latent variables, observed variables, you unidirectional paths one way direction path disturbances affecting the latent variables measurement errors and disturbance affecting the observed variables.

(Refer Slide Time: 11:15)



Multiple Regression

$X_1 \sim N(\mu_1, \sigma_1^2)$
X_1
X_2
Y
ϵ
$\varepsilon \sim N(0, \sigma^2)$
$X_2 \sim N(\mu_2, \sigma_2^2)$

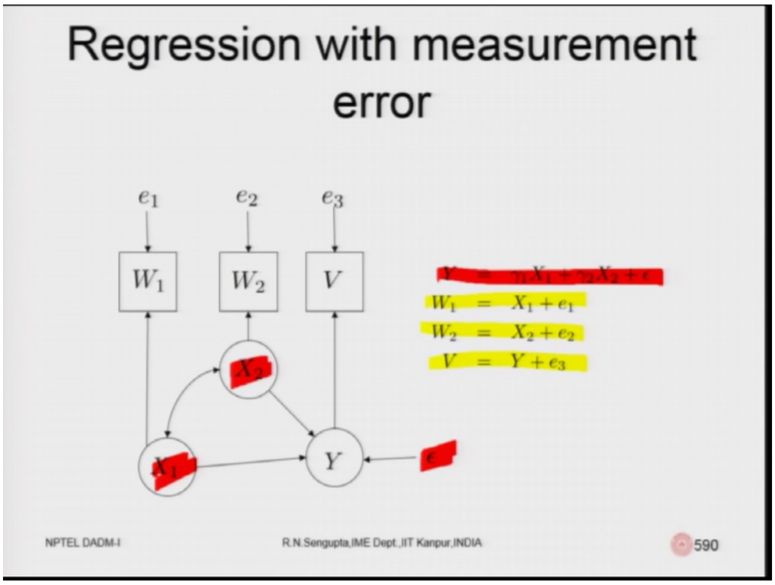
$Y = \gamma_1 X_1 + \gamma_2 X_2 + \epsilon$

NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 589

Now, consider the relationship for the multiple linear regression models what we have. So, in the multiple linear regression model, which generally we have that Y is the observed variable which you want to analyze and there are effects coming from the external sources considered like the regression model multiple linear regression model of temperature being affected by humidity being affected by pressure. So, will consider temperature to be Y and humidity and the pressure being denoted by the red colours which is X 1 and X 2 we had um we had seen that they have would have basically some assumptions; assumptions pertaining to normality, assumptions pertaining to covariance being 0 between X 1 X 2, assumptions pertaining to rank, assumptions pertained to the errors being having a mean value of 0, variance of 1 or variance of sigma square and this is basically the errors which we have.

So, if you consider the equation, the equation is Y is related to beta 1 into X 1 plus beta 2 into X 2 where beta 1 beta 2 have been replaced here in this equation by gamma 1 and gamma 2 and the error term which I have written is normally distributed with 0 mean and sigma square. Similarly X 1 is normally distributed with mu 1 sigma 1 square, X 2 normally distributed with mu 2 and sigma 2 square and; obviously, you can find out the mean value of Y and the variance of Y. Mean value of Y would basically be the mean value of X plus the mean value of X 1 plus mean value of X 2 plus the mean value of epsilon, epsilon mean value being 0. Similarly when you come to the variances of Y, it will basically consist of the variance of X 1, variance of X 2, variance of epsilon and also the covariance is existing between X 1, X 2, X 1 and epsilon X 2 and epsilon. But as per the assumption the co variances values would be 0 because we have assumed it particularly as mentioned time in again in the models so; obviously, you will have a very simplified version of the variances for the Y based on which you can do the calculations.

(Refer Slide Time: 13:41)



Regression with measurement error

$$\begin{aligned} Y &= \gamma_1 X_1 + \gamma_2 X_2 + \epsilon \\ W_1 &= X_1 + e_1 \\ W_2 &= X_2 + e_2 \\ V &= Y + e_3 \end{aligned}$$

NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 590

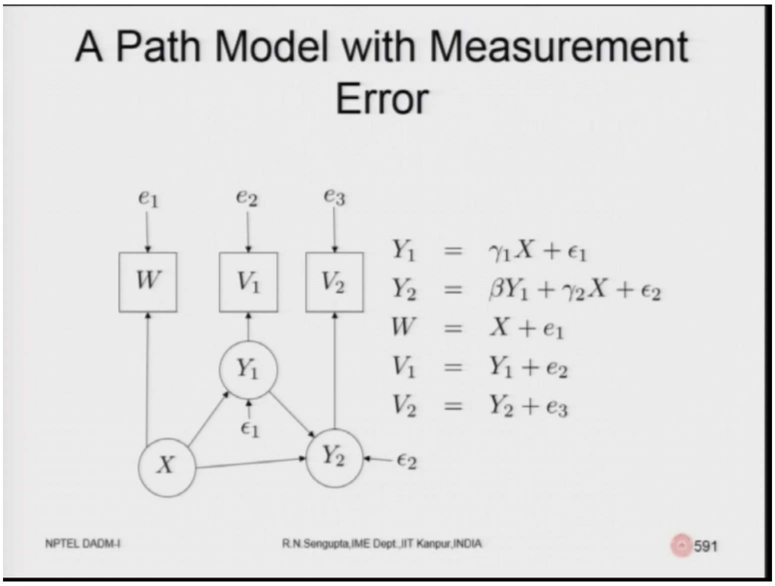
Now, regression with measurement errors; obviously, it would mean that the errors which you are measuring would have some would have measurements based on the fact, that the measurements are randomly changing. So, if the measurement errors are there what you will basically have is e Y which is again the random variable which you want to understand; now apart from X 1 and X 2 which are the independent variables as usual there is an error also.

So; obviously, these would be highlighted using the red colour X 1 is there, X 2 is there and the error is there. Now what you will consider is that there are measurement errors also and the measurement errors are coming by a so, called third set of equations and the third set of equations are I will highlight is the during the yellow colour. And the initial regression model is the red one which has all the assumptions. So, what does W 1? W 1 is basically the affect technically we I assume that X is non-stochastic; in the sense it will have a variance, but non stochasticity would mean that any point of time and any relationship or the effects which are there on X 1 or the effects which are there on X 2 would not be considered in the simple regression model, but in the measurement error models will consider the effects are there and the effects are being assumed like this.

They would be a separate error for X 1 which will given by e 1, they will be separate error for X 2 will be given by e 2. Now the assumptions of distribution for e 1 and e 2 in the simple case can be considered as normal; in the case when it is not simple we will consider a certain distribution existing form of the errors pertaining to measurement of X 1 errors pertaining to measurement of X 2. So, these 3 equations which are yellow in colour would pertain to the fact that there is measurement errors happening for X 1, measurement errors happening for X 2, measurement errors happening for Y also. Apart from the effect which is coming out for on Y from X 1 from X 2 and the error terms which is white noise which is not under your control.

So, technically we are trying to divide the total errors into white noise, which is not under control and some of the measurement errors which you can control, but they would be considered by a certain type of distribution. Now consider a path model with measurement errors.

(Refer Slide Time: 16:17)



A Path Model with Measurement Error

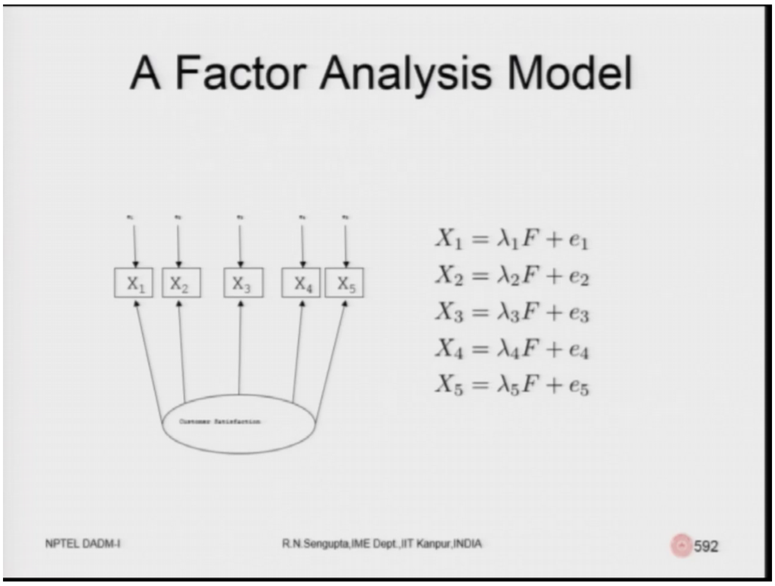
$$
\begin{aligned}
Y_1 &= \gamma_1 X + \epsilon_1 \\
Y_2 &= \beta Y_1 + \gamma_2 X + \epsilon_2 \\
W &= X + e_1 \\
V_1 &= Y_1 + e_2 \\
V_2 &= Y_2 + e_3
\end{aligned}
$$

NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 591

So, in this case you will basically have Y being affected by technically by the X 1 and X 2 which are already there. So, you will basically break down X 1 and X 2 accordingly where you will basically have the errors pertaining to Y. So, Y would basically be affected by X 1 and X 2 but these values which will have we will go and stage by stage; that means, Y would basically have it need not be only one random variable Y would have basically have on one X which is being affected by an error plus a. So, called rate of change or regression value which is gamma 1 and this Y 2 would be the measurement errors which will be coming both by the combination of Y 1 and X 1.

So, now will basically try to break up the overall measurement errors at in stages; in the sense there would be some measurement errors happening for X 1 which is the first random variable, there would be a certain set of measurement errors happening for X 2 which is the second random variable so, on and so, forth plus there would be a measurement error happening for the effect that Y is being affected by an external set of a variables, which is X 1 to Xp whatever the numbers are. So, each would have their measurement data 0.1, the measurement error would be coming from measuring Y also and the third one would be the white noise which is already existing.

(Refer Slide Time: 17:48)



A Factor Analysis Model

$$X_1 = \lambda_1 F + e_1$$
$$X_2 = \lambda_2 F + e_2$$
$$X_3 = \lambda_3 F + e_3$$
$$X_4 = \lambda_4 F + e_4$$
$$X_5 = \lambda_5 F + e_5$$

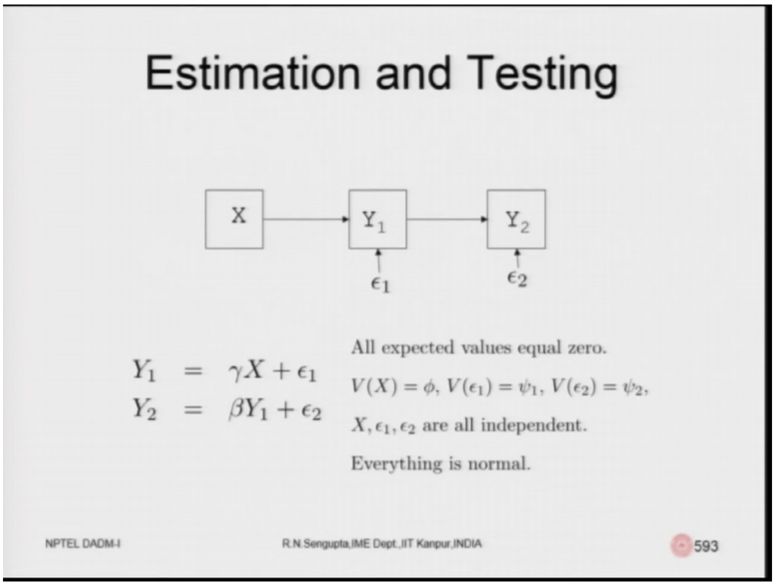
NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 592

Now, in the factor analysis model we consider that all the Xs depending on the assumptions whether their distribution is they would be an error. So, the error would basically be consisting for each or each of the Xs in the very simplistic sense and this is an error which can be controlled measure an error which cannot be controlled.

Now, if you have 3 Xs. So, each of the Xs would have measurement errors plus an error which is a part of the white noise which is being subsumed under X 1 X 2 X 3. So, Y would now have errors pertaining to X, errors pertaining to the errors of measuring X plus the error pertaining to the measurement of Y only. So, if there are 3 X. So, they are technically they would be so, called 6 from Xs and one from the errors which is basically the white noise.

So, you can make it more complicated depending on the internal structure of the relationship which you may have between X 1 X 2 X 2 X 3 and X 1 and X 3 accordingly.

(Refer Slide Time: 18:52)



Estimation and Testing

$$\begin{aligned} Y_1 &= \gamma X + \epsilon_1 \\ Y_2 &= \beta Y_1 + \epsilon_2 \end{aligned}$$

All expected values equal zero.

$V(X) = \phi$, $V(\epsilon_1) = \psi_1$, $V(\epsilon_2) = \psi_2$,

$X, \epsilon_1, \epsilon_2$ are all independent.

Everything is normal.

NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 593

So, in the estimation and the testing model what you do is that, you go step by step you goes to in stage wise where you first basically find out the and the variables pertaining to Y 1 then utilize the variables or the measurement errors pertaining to Y 1 you find out Y 2 go step by step, such that you are able to find out the overall error which can be broken down for all the Xs their measurement of the Xs and also the white noise.

Now, when you measure this type of models, there are different types of errors which we do. So, I will just speak about the type of errors and then again can go into the slide and trying to concentrate on those. If you remember we have considered that generally the errors which we have are basically the estimation errors. So, there is some error problem you have not been able to estimate and you consider different about loss functions. So, loss functions can be either the linear (Refer Time: 19:47) loss functions the loss functions can be quadratic loss functions, loss functions can be linear loss function, but weighted way; that means, over estimation is more penalizing than under estimation under estimation is more penalizing over estimation the graphs which we had done.

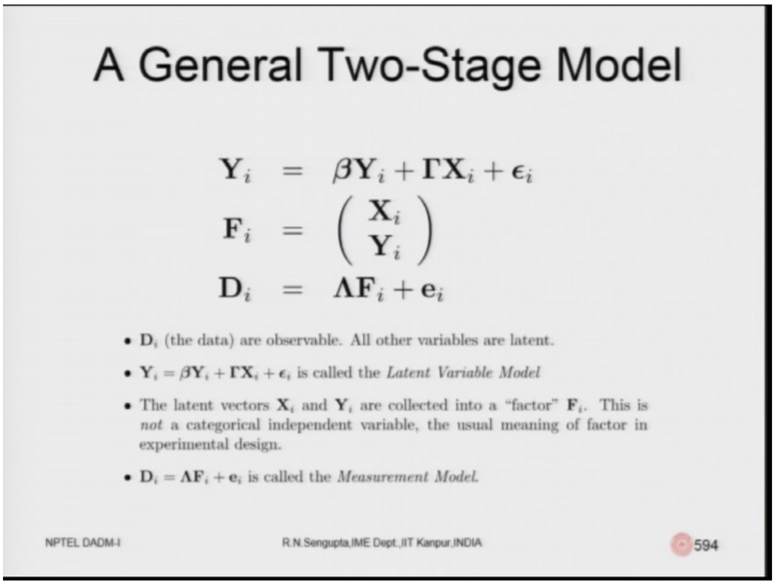
Then we can consider that for the regression model, we have considered the balance loss function, which is nowadays can be utilized for goodness of fit in the precision of estimation and the goodness of fit and the precision estimation can be modelled accordingly. Then we have basically the linex loss and the linear exponential loss function, then you have can have the Taguchi loss function they can be different type of loss function based on which you can model and try to find out that what portions of the errors can be modelled accordingly. Now remember trying to basically model and find out the estimates for one loss function does not mean that the estimates which you find out for the first set of loss function is the best or the worst. It will depend on the type of problems you are going to solve and what is best for that case as you as you analyze.

So, if you consider the simple examples which had given, I am repeating it please, please bear with me. In the case when you when you have considered trying to be is basically build a dam or the case when he had basically consider trying to manufacture or find out the warranty life of the vacuum circuit breakers in a huge electrical circuit. We consider the type of losses would be relevant based on the factor that whether over estimation or underestimation was more penalized. Make a may be in many of the cases that in some depending on the some range of the of the parameters over estimation may be maybe more penalized than under estimation, but if the range is change; obviously, the effect gets interchanged in the sense under estimate underestimation gets more penalized than over estimation.

So, we will consider that in this case when you are doing the estimation and the testing problem, will consider the all the expected values should equal to 0 in the sense the error term based on which we are going to do would in the long run have an error which expected value 0. We will consider the variances existing for the errors and the various existing for the X to be such that, they would be non-stochastic; that means, they would not be changing with respect to time. Once found out the errors variance would basically be keep kept fix at the same level. So, we will consider the excess and epsilon 1 epsilon 2 are all independent and everything would be normal in the sense that we will consider the normality to be true as the distribution based on which we will do our calculations accordingly.

Now, normality to be true is a very simplistic assumption, but the result which you get by trying to utilize the normality would definitely give you some idea and that how the problems can be solved and will consider that normality to be true for the other models also.

(Refer Slide Time: 22:33).



A General Two-Stage Model

$$\begin{aligned}\mathbf{Y}_i &= \beta\mathbf{Y}_i + \mathbf{\Gamma}\mathbf{X}_i + \epsilon_i\\ \mathbf{F}_i &= \begin{pmatrix}\mathbf{X}_i\\ \mathbf{Y}_i\end{pmatrix}\\ \mathbf{D}_i &= \mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i\end{aligned}$$

- $\mathbf{D}_i$ (the data) are observable. All other variables are latent.
- $\mathbf{Y}_i = \beta\mathbf{Y}_i + \mathbf{\Gamma}\mathbf{X}_i + \epsilon_i$ is called the *Latent Variable Model*
- The latent vectors $\mathbf{X}_i$ and $\mathbf{Y}_i$ are collected into a "factor" $\mathbf{F}_i$. This is *not* a categorical independent variable, the usual meaning of factor in experimental design.
- $\mathbf{D}_i = \mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i$ is called the *Measurement Model.*

NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 594

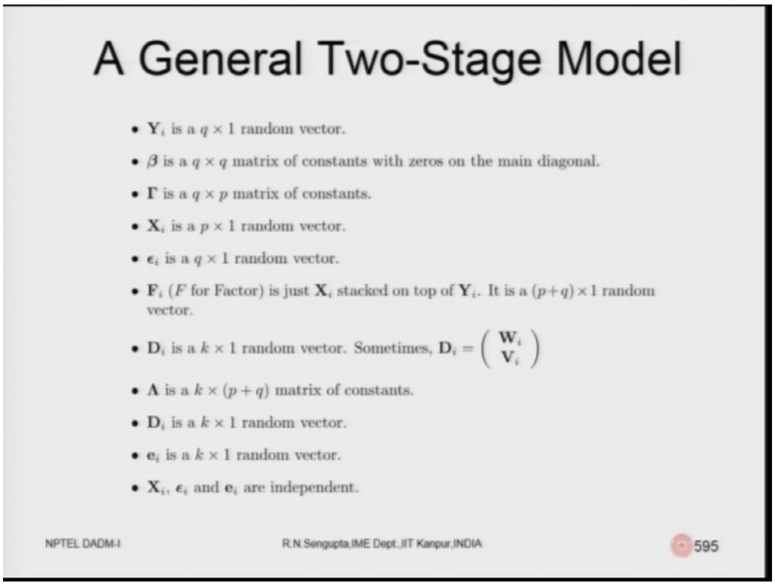
So, now will consider a general 2 stage model for the for the structural equation modelling. So, here D is the data which are observable and the suffix part I jk, I am going to now skip because that will depend on the number. So, we will consider D the errors as the data which are observable and you can find out and you can see and you can measure all the variables will consider at a latent.

So, based on that will do the calculations. ill consider the equation on the relationship between the Y is to be in such a way that Y would be dependent on the 2 sets of variables, which are basically called the Xs and the Ys. So, they are basically the latent factors and X and Ys are collected into a vector factor F. So, we are basically trying to analyze the or the factors have been divided into 2 sets X and then Ys and they would basically dependent on whether the stochastic is there for some of them whether the stochastic is not there for some of them.

So, those are reading the third point it would mean the latent factors X 1 and Y 1 are collected into the factors F and with the suffix I means the reading number; so, that each reading is being divided into 2 to 2 factors. So, this is not a categorical independent variable the usual meaning of factor is basically, they are the experimental design variables based on which we are doing their studies. So, we will basically have one set of the data's would basically have a measurement model, which means some observable part is actually as it is and some more basically be the so, called measurement errors which you have.

So, measurement errors would be denoted by e with the suffix i depending on the reading number. And your actual model would basically be given Y would be given by the by the weightages depending on the factors. So, the weightages which will basically have is basically beta and gamma. So, these are vectors. So, the beta factors would basically with the factors, which will be affecting the level of variability which will be there in the first set of factors which is X and gamma would be the effects or the weights which you will give to the second set of for Y for F which is basically the Xs values which you have. So, you are giving weights to both X and Y and depending on the errors which you have, you will basically try to model it accordingly try to break the errors into observable part on observable part and basically the white noise which is there.

(Refer Slide Time: 25:17)



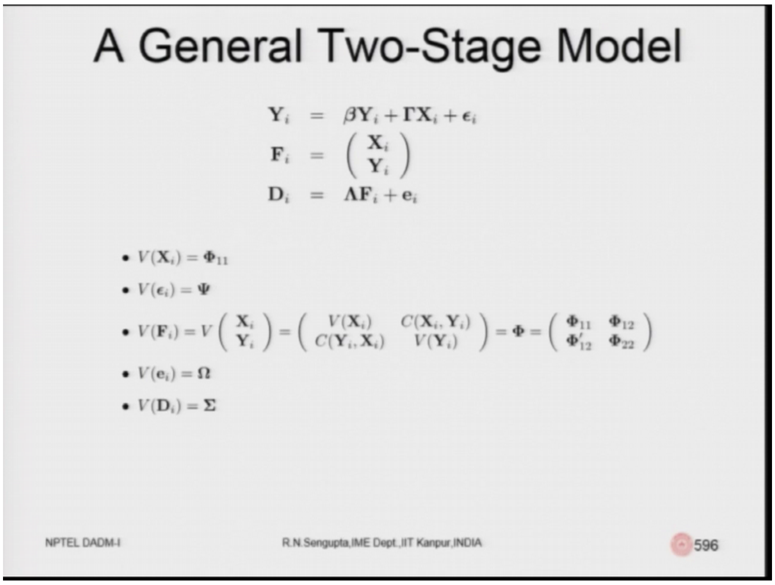
A General Two-Stage Model

- $\mathbf{Y}_i$ is a $q \times 1$ random vector.
- $\boldsymbol{\beta}$ is a $q \times q$ matrix of constants with zeros on the main diagonal.
- $\boldsymbol{\Gamma}$ is a $q \times p$ matrix of constants.
- $\mathbf{X}_i$ is a $p \times 1$ random vector.
- $\boldsymbol{\epsilon}_i$ is a $q \times 1$ random vector.
- $\mathbf{F}_i$ (F for Factor) is just $\mathbf{X}_i$ stacked on top of $\mathbf{Y}_i$. It is a $(p+q) \times 1$ random vector.
- $\mathbf{D}_i$ is a $k \times 1$ random vector. Sometimes, $\mathbf{D}_i = \begin{pmatrix} \mathbf{W}_i \\ \mathbf{V}_i \end{pmatrix}$
- $\boldsymbol{\Lambda}$ is a $k \times (p+q)$ matrix of constants.
- $\mathbf{D}_i$ is a $k \times 1$ random vector.
- $\mathbf{e}_i$ is a $k \times 1$ random vector.
- $\mathbf{X}_i$, $\boldsymbol{\epsilon}_i$ and $\mathbf{e}_i$ are independent.

NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 595

So, we will basically now describe them accordingly. So, Y is a q cross 1 random vector. So, q is the number of random variables which are there. Beta is a q cross q matrix of constant with zeros on the on the main diagonal and the, of the diagonal element would be they would be given by the factors or the weightages which are there. The gamma and Xs are basically the q cross p and p cross 1 matrices are constant and random vectors. The epsilon are the q cross 1 random vectors which we have which is basically considering the white noise. The F are the factors which is just tact into two's different sets of X i s and Y i s they would basically be p plus q cross 1. So, because p factors are coming from X and q factors are coming from Y d is the random variable corresponding to the measurements of the datas and correspondingly you will basically have also consider the Xs the epsilons and the es e e being the other set of random variables which would be the white noise are all independent of each other if they are not; obviously, it will affect that how you are able to find out the errors and find out the correlation coefficients accordingly.

(Refer Slide Time: 26:37)



A General Two-Stage Model

$$\begin{aligned}\mathbf{Y}_i &= \beta\mathbf{Y}_i + \mathbf{\Gamma}\mathbf{X}_i + \boldsymbol{\epsilon}_i \\ \mathbf{F}_i &= \begin{pmatrix}\mathbf{X}_i \\ \mathbf{Y}_i\end{pmatrix} \\ \mathbf{D}_i &= \mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i\end{aligned}$$

- $V(\mathbf{X}_i) = \mathbf{\Phi}_{11}$
- $V(\boldsymbol{\epsilon}_i) = \mathbf{\Psi}$
- $V(\mathbf{F}_i) = V\begin{pmatrix}\mathbf{X}_i \\ \mathbf{Y}_i\end{pmatrix} = \begin{pmatrix}V(\mathbf{X}_i) & C(\mathbf{X}_i, \mathbf{Y}_i) \\ C(\mathbf{Y}_i, \mathbf{X}_i) & V(\mathbf{Y}_i)\end{pmatrix} = \mathbf{\Phi} = \begin{pmatrix}\mathbf{\Phi}_{11} & \mathbf{\Phi}_{12} \\ \mathbf{\Phi}'_{12} & \mathbf{\Phi}_{22}\end{pmatrix}$
- $V(\mathbf{e}_i) = \mathbf{\Omega}$
- $V(\mathbf{D}_i) = \mathbf{\Sigma}$

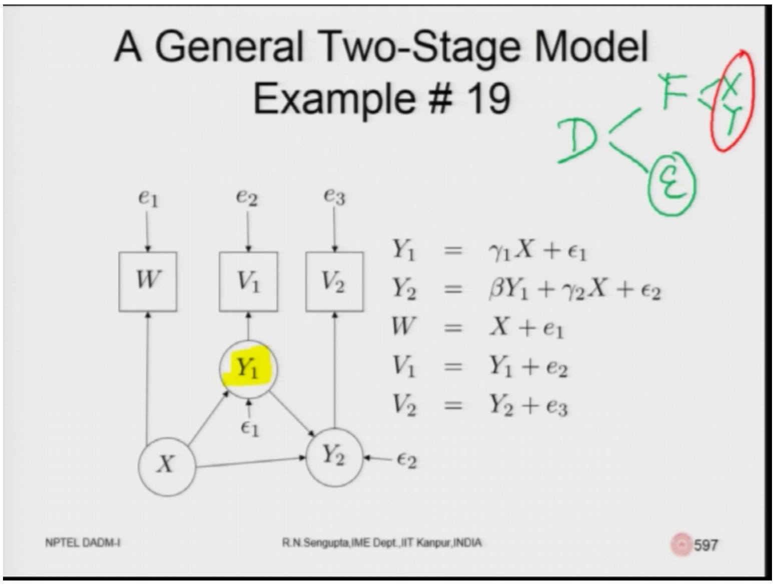
NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 596

So, here is the general 2 stage models and we will try to basically analyze that what are the errors based on which you will try to solve the problem.

So, this is Y is the FX which you want to study; F with the factors which have been divided into X and Y and D is the general set of data measurements which you have basically happened. We will consider the variability to be there for X epsilon and an e and they would be termed accordingly considering that we consider initially in the normal distribution to be true and independent structure being true between the factors and the variabilities as I mentioned. So, X epsilon and e would basically be independent of each other.

Now, the problems which will have is that, I will come to the main errors later on. So, the problem would look like this.

(Refer Slide Time: 27:25).



A General Two-Stage Model
Example # 19

$$
\begin{aligned}
Y_1 &= \gamma_1 X + \epsilon_1\\
Y_2 &= \beta Y_1 + \gamma_2 X + \epsilon_2\\
W &= X + e_1\\
V_1 &= Y_1 + e_2\\
V_2 &= Y_2 + e_3
\end{aligned}
$$

NPTEL DADM-I R.N.Sengupta,IME Dept.,IIT Kanpur,INDIA 597

So, you have basically. So, this is Y which you want to study. So, Y would basically be coming out from FX of X n Y is the actual factors which are start studying. So, basically you will have the data, which will have basically the factors F this being divided into X and Y and there would be an error. So, each so, this is the white noise and; obviously, if I consider they would be errors in the measurements of X and Y which are the factors themselves.

So, each factor measurements have an error plus there is an error corresponding to the white noise which you are considering. So, I will I will basically concentrate on the different type of errors which are there and discuss that in the last class. So, considering that we have will be will be discussing structural equation modelling in the last class I will give the hints and the errors based on which we are able to study and analyze structural equation modelling rather than solving the problem it will be much better if I tell you what are the mean point based on which you can model the structural equation modelling. With this I will close this second last, but one class and we will have one in one lecture to wrap up this course and give you some hints that how you can basically analyze and take the information set which you have been studying for DADM 1 in a much better way to analyze different type of problems. I know that I am again repeating which I did in the last class lecture; the covering the whole set of topics which are actually there in DADM 1 is a huge task because considering that DADM 1 would be a whole semester or 2 semesters course we are trying to crash it in 30 hours.

So, we have tried our level best if there are differently queries from your end we will try to answer it in the forum and hope we will be able to help the students with all the queries as they come; and after the they solve the 12th assignment and also take the examinations I am sure things would be much clearer and how to read and how to solve the problems accordingly. Have a nice day and.

Thank you very much.