From measurement model to structural model

In the previous video we looked at how we can use latent variable models in particular confirmatory factor analysis to specify what we call the measurement model, the measurement part of a structural equation model. In that component we are seeking to get good measures of the concepts that we're interested in, in our theoretical model, we were wanting to assess how well we are able to measure concepts to correct for error and assess the overall fit of those models. In this video we will extend from the measurement model to include what we call structural equations which specify our theory of how our concepts are related to one another. The steps in fitting structural equation models we've already looked at in the previous videos the idea of establishing a satisfactory measurement model, so we have a clear idea of the concepts that we're interested in that form our theoretical model, we use latent variable models, confirmatory factor analysis and so on, to get good measures of those concepts.
Now we want to specify how those concepts are related to one another and this is really the part of the modelling where we are testing our causal theory, and what this really involves is specifying regression paths between the measured concepts. Now why don't we call these the structural equations and well the terminology here is a little bit loose it really comes from econometrics where the idea of a structural equation is one which embodies the formal theory of economics within the equation. The same is not so true in the context that we are considering here general structural equation modelling, but it still maintains the same idea that the structural equations should embody our a priori theory of how the concepts are related, but in general we won't have a strong a priori theory as we would do in the econometrics context.

So we are going to fit regression paths between the measured concepts and then we will estimate the parameters of the model and this will yield beta weights for the regression equations in the normal way that we would get in a OLS regression for example and we would test hypotheses about the direction and magnitude and significance of those model parameters which would be suggested by our theory and then the last step would be to assess the fit of the model. If we have a well-fitting model then we can be confident that our estimates of the model parameters are consistent and unbiased. So those are the basic steps and we've already seen the measurement part of the model so here really we're just extending what we saw in the previous video using confirmatory factor analysis to incorporate regressions between the measured concepts. So this is a very simple structural equation model it's really equivalent to a bivariate regression model where we are regressing a dependent variable on a single independent variable but in this context we now have latent variables rather than directly observed variables.
So e1 here is the dependent variable and Si one is the independent variable, the exogenous variable, and we have a disturbance term in that for that regression and we're going to estimate some value for beta which would then tell us what the relationship between e1 and Xi..n this example. So a very simple structural equation model just extending a regression path between two latent variables.
Now we can make this system of structural equations more complex, we've done that now by adding in two more latent variables and we now have a covariance between the two exogenous variables, that means that what we're saying is that with these are they're exogenous variables they're correlated with one another but we aren't making any statements about our theory of the causal relationships between those variables, but we have a regression of e1 on Xi1 and of e2 on Xi2 and we also have a regression of e1 on Xi2. So we are including a number of different regression weights here and that's something that is different

than we would be used to in a standard OLS regression context because we have variables which are now at both endogenous and exogenous variables, so e1 in this example is a dependent variable in one equation, but it's also a predictor an independent variable in the second equation because we can see we have a regression of e2 on e1. So what we can also see here is that we've introduced indirect effects that we have potentially in effect here running from Xi2 to e2 which runs through e1. So we can estimate that sort of relationship and we'll come to look at that in a bit more detail later.

So here we have a an example of a more complicated model, but this should relate to our theoretical expectations about what generated the data that we're analysing, and so we would recover the parameter estimates for the beta weights in this model and we would check the fit of the data using the kind of indices that we saw in the previous videos and if we have a well-fitting model then we would make some inferences about the theory that led us to develop this model. Does the model fit the data, do the parameter estimates support or reject the hypotheses. Another kind of model that we can fit as a structural equation model is what's referred to as a multiple indicators multiple causes model and there are a couple of things to note about this, you can see we have one latent endogenous variable, that is e1, but the predictor variables in this model are just observed variables they're not latent. So something to note there is that a structural equation model does not have to comprise solely of latent variables it can be a mix of both directly observed and latent variables. The sort of use and context of this mimic model as it’s known is often to assess the measurement properties of a latent variable and what's interesting to note here in this path diagram is that there is a regression path between the observed variable x1 and z1.
So z1 is predicting the latent variable but it's also got a regression path into the observed indicator of that latent variable. Now we would use this as a way of testing whether that x1 variable, the observed indicator, is that perhaps functioning differently for different subpopulations. So we can imagine perhaps that the three observed variables here are perhaps scores on a test, and if we find a significant beta weight running between Z1 and X1 that would tell us that depending on your score on Z1 you would have a higher or a lower probability of getting a correct answer to X1, and that would be over and above your score on the latent variable, which in this case would be some kind of ability score and that would tell us that there is a problem with that item, it's what we refer to as differential item function. That some groups in the population will have a higher probability of getting that item correct even controlling for their level of ability, and this can generalize to attitude scales as well as ability scales. So that's the sort of context that one might use this kind of mimic model in.

Now we've seen that the process that we follow through in instruction equation modelling is to specify the measurement model, get a good fitting model, then look at estimating regression paths between the latent variables which we use to test our theory, then we check the fit of the structural model and if that fits then we say well our model fits the data and we have a theory which is supported by the data. Now that might lead us to conclude and that we've got the true model, this is the correct data generating model, but we need to be very careful about avoiding the temptation to make that strong conclusion. That's because with this kind of model with an observational data there are many different potential generating mechanisms that could have produced the data and could give you equally good or perhaps even better fit.
Now here's a very simple way of demonstrating that point. Here we have a simple bivariate regression of one latent variable measuring social trust on a variable latent variable measuring political trust and this is using data from the European Social Survey, and you can see that from the fit statistics there that this is a well-fitting model and we have a significant

beta weight there, so we find that the two concepts that political trust has a significant effect on or relationship with, to be more neutral on social trust. So this might lead us to conclude that this is the true model but of course we can simply reverse the direction of the path between social and political trust. This tells us something completely different, in fact we reverse the causal direction that we see here and but the fit of the model and the beta weight indeed is exactly the same with this completely opposite model. So this is a caution about the strength of the conclusions that we can draw even when our model fits the data well and is not necessarily the true data generating model.