

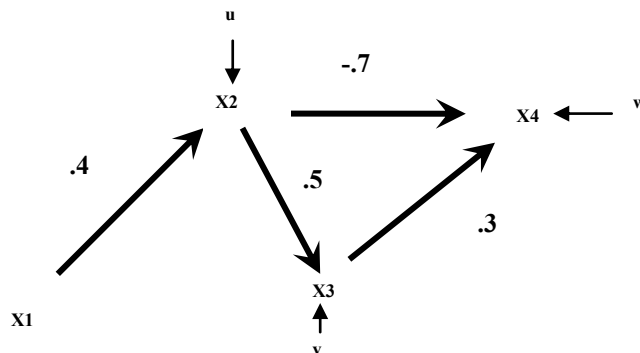
**Soc 63993, Advanced Social Statistics II**  
**Homework No. 7**  
**Nonlinear effects/ Intro to path analysis**

**Problem 1.** The files *nonlinhw.sps* and *nonlinhw.sav* will generate the computer runs you need for this problem. Copy them from the course web page. [Note: The output looks a little different in different versions of SPSS, but as far as I can tell the key elements stay the same.]

There are 4 variables in *nonlinhw.sav*: X1 (the IV), Y1, Y2, and Y3 (the DVs). The SPSS program does scatterplots of X1 versus each DV. It also uses the SPSS Curvefit and Regression routines to fit models to these data. For each DV in turn, you are to do the following:

- Examine each scatterplot. Explain why the relationship is nonlinear and what type of nonlinearity appears to be present.
- Discuss what problems result from a linear (mis)specification. The SPSS Curvefit results will probably help you here. Curvefit shows you what the linear prediction is and compares that with the actual observed data and (in some cases) also shows you the predictions under a nonlinear model.
- Various variables are computed and then incorporated into a multiple regression. Explain why, i.e. how does the approach taken make it possible to deal with nonlinear effects?
- Present a substantive example, real or hypothetical, that the model you have estimated might be appropriate for. Explain why it is appropriate. Do not use any of the examples already given in class.
- For Y3 only, two different Curvefits are presented. Explain why, based on the graphics only, it would be difficult to decide which nonlinear specification was most appropriate, and how theory might help you to choose.

**Problem 2.** A sociologist believes that the following model describes the relationships between X1, X2, X3 and X4. All variables are in standardized form. The hypothesized value of each path is included in the diagram.



a. Write out the structural equation for each endogenous variable, using both the names for the paths (e.g.  $\beta_{42}$ ) and the estimated value of the path coefficient.

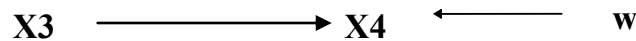
b. Part of the correlation matrix is shown below. Determine the complete correlation matrix. (Remember, variables are standardized. You can use either normal equations or Sewell Wright, but you might want to use both as a double-check.) Optionally, confirm your answer using either SPSS or Stata, i.e. enter the correlations into one of the programs and confirm that you get the right estimates for the betas.

	x1	x2	x3	x4
x1	1.0000			
x2	0.4000	1.0000		
x3	?	?	1.0000	
x4	-0.2200	?	?	1.0000

c. Decompose the correlation between X3 and X4 into

- Correlation due to direct effects
- Correlation due to indirect effects
- Correlation due to common causes

d. Suppose the above model is correct, but instead the researcher believed in and estimated the following model:



What conclusions would the researcher likely draw? In particular, what would the researcher conclude about the effect of changes in X3 on X4? Why would he make these mistakes? Discuss the consequences of this mis-specification.