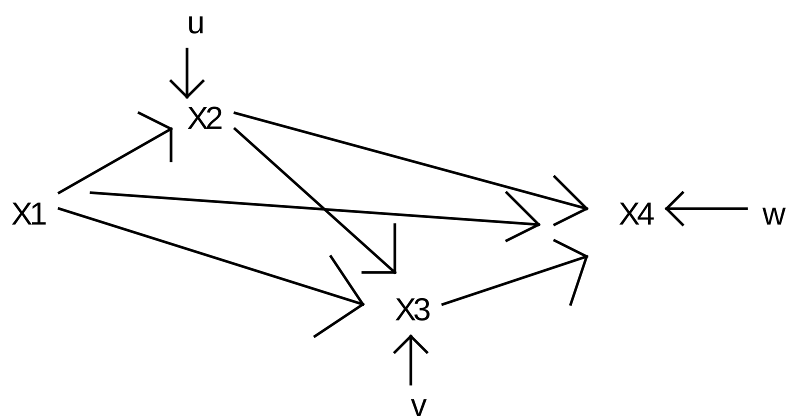
Structural Coefficients in Recursive Models/ Evils of Standardization

Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/
Last revised February 22, 2015

RECURSIVE DEFINED. A model is said to be recursive if all the causal linkages run "one way", that is, no two variables are reciprocally related in such a way that each affects and depends on the other, and no variable feeds back upon itself through any indirect concatenation of causal linkages. The model we have been looking at is recursive:



The model would be non-recursive if, for example, X4 also affected X3, i.e. the causation ran in both directions. Non-recursive models are much more difficult to work with, and we'll discuss them later in the course.

RECURSIVE MODELS WITH STANDARDIZED VARIABLES. We have been examining a 4-variable recursive model in which variables were standardized. The b's in such models are referred to as the *path coefficients*. The advantages of standardized variables are:

1. Certain algebraic steps are simplified

2. Sewell Wright's rule for expressing correlations in terms of path coefficients can be applied without modification

3. Continuity is maintained with the earlier literature on path analysis and causal models in Sociology

4. It shows how an investigator whose data are only available in the form of a correlation matrix can, nevertheless, make use of a clearly specified model in interpreting those correlations.

Nevertheless, standardization should generally be avoided. Standardization tends to obscure the distinction between the structural coefficients of the model and the several variances and covariances that describe the joint distribution of the variables in a certain population. To illustrate this, we will first see what happens when variables are not standardized.

RECURSIVE MODELS WITH UNSTANDARDIZED VARIABLES. We will continue to assume that all X's have a mean of 0. (The only thing this affects is the intercepts.) We will not assume that variances all = 1. Under these conditions,

$$E(X_i^2) = E(X_iX_i) = V(X_i) = \sigma_i^2$$

$$E(X_iX_j) = Cov(X_iX_j) = \sigma_{ij}$$

To get the normal equations, we proceed as before: Multiply each structural equation by the predetermined variables and then take expectations. In addition, to get the variances, we multiply the structural equation by the DV of the equation and take expectations. Hence,

(1) For X2, the structural equation is

$$X_2 = \beta_{21}X_1 + u$$

The only predetermined variable is X1. Hence, if we multiply both sides of the above equation by X1 and then take expectations, we get the normal equation

$$\begin{aligned} E(X_1X_2) &= \beta_{21}E(X_1^2) + E(X_1u) = \\ \sigma_{21} &= \beta_{21}\sigma_1^2 \end{aligned}$$

When variables are standardized, $\sigma_1^2 = 1$ and $\sigma_{21} = \rho_{21}$, but that isn't true when the variables are not standardized.

The variance of X2 is

$$\begin{aligned} E(X_2X_2) &= \beta_{21}E(X_1X2) + E(uX_2) = \\ \sigma_2^2 &= \beta_{21}\beta_{21}\sigma_1^2 + \sigma_u^2 \\ &= \beta_{21}^2\sigma_1^2 + \sigma_u^2 \end{aligned}$$

(2) For X3, the structural equation is

$$X_3 = \beta_{31}X_1 + \beta_{32}X_2 + v$$

There are two predetermined variables, X1 and X2. Taking each in turn, the normal equations are

$$\begin{aligned} E(X_1X_3) &= \beta_{31}E(X_1^2) + \beta_{32}E(X_1X_2) + E(X_1v) = \\ \sigma_{13} &= \beta_{31}\sigma_1^2 + \beta_{32}\sigma_{12} = \beta_{31}\sigma_1^2 + \beta_{32}\beta_{21}\sigma_1^2 \end{aligned}$$

Doing the same thing for X2 and X3, we get

$$E(X_2 X_3) = \beta_{31} E(X_1 X_2) + \beta_{32} E(X_2^2) + E(X_2 v) =$$
$$\sigma_{23} = \beta_{31}\sigma_{12} + \beta_{32}\sigma_2^2 = \beta_{31}\beta_{21}\sigma_1^2 + \beta_{32}(\beta_{21}^2\sigma_1^2 + \sigma_u^2)$$

We can proceed similarly to get the variance of X3 and the normal equations for X4. (The mathematical simplicity of standardized variables should be fairly apparent by now!)

The key thing to note is that the variances and covariances are functions of (at most) three kinds of quantities:

1. the variance of the exogenous variable

2. the variance(s) of one or more disturbances

3. a nonlinear combination of structural coefficients

For example, in the model we have been working with, there is

1 variance of the exogenous variable

3 variances of the disturbances

6 structural coefficients

That gives us 10 parameters altogether. Note that there are also 10 variances and covariances among the four X variables.

We can suppose without contradiction that one of these components may change without any of the others having to change. If any of them changes, however, the observable variances and covariances will, in general, change.

That is, suppose we have 2 populations. Suppose that the structural coefficients are the same in both and the variances of the disturbances are the same. If only σ_{11} differs,

- All the other variances and covariances will differ

- All the correlations will differ

- All the standardized path coefficients will differ

Hence, if we look only at the standardized path coefficients, it will appear that the two populations differ completely; when in reality, the only thing that differs is the variance of the exogenous variable.

This is why we use the term "structural coefficients" — because structural coefficients don't change when other parameters change.

AN EXAMPLE. In the following example, note that

- Hypothetical regression analyses are presented for 2 populations, side by side
- The metric coefficients are the same for each population
- The standardized coefficients substantially differ.
- In this example, only one "structural" parameter differs between the two populations. In population 1, the s.d. of X1 is 1, whereas in population 2 the s.d. of X1 is 2. Specifically, the parameters are

Parameter	Population 1	Population 2
σ_{11}	1	4
β_{21}	9	9
β_{31}	12	12
β_{32}	2	2
β_{41}	30	30
β_{42}	2	2
β_{43}	1	1
σ_{uu}	500	500
σ_{vv}	6,000	6,000
σ_{ww}	90,000	90,000

Here are the descriptive statistics for the two populations. Notice how much they differ. Standard deviations and correlations are all higher in population 2, even though the only structural parameter that differs across populations is the variance of X1.

Population 1

Descriptive Statistics

	Mean	Std. Deviation	N
X1	.000000	1.0000000	100
X2	.000000	24.1039400	100
X3	.000000	94.3398100	100
X4	.000000	331.7891000	100

Correlations

		X1	X2	X3	X4
Pearson Correlation	X1	1.000	.373	.318	.235
	X2	.373	1.000	.558	.338
	X3	.318	.558	1.000	.394
	X4	.235	.338	.394	1.000

Population 2

Descriptive Statistics

	Mean	Std. Deviation	N
X1	.000000	2.0000000	100
X2	.000000	28.7054000	100
X3	.000000	107.7033000	100
X4	.000000	358.2401000	100

Correlations

		X1	X2	X3	X4
Pearson Correlation	X1	1.000	.627	.557	.435
	X2	.627	1.000	.673	.468
	X3	.557	.673	1.000	.502
	X4	.435	.468	.502	1.000

Now, note the similarities and differences when we estimate the regression models. Structural coefficients are the same, but most other parameters differ.

```
Multiple R           .37338
R Square             .13941
Adjusted R Square    .13063
Standard Error     22.47447

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           1          8018.99825       8018.99825
Residual            98         49499.99418        505.10198

F =      15.87600       Signif F =  .0001

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

X1             8.999999    2.258770    .373383     3.984  .0001
(Constant)      .000000    2.247447                 .000 1.0000


End Block Number   1   All requested variables entered.

* * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 2    Dependent Variable..   X3

Block Number  1.  Method:  Enter      X1       X2

Variable(s) Entered on Step Number
   1..    X2
   2..    X1

Multiple R           .57083
R Square             .32584
Adjusted R Square    .31194
Standard Error     78.25414

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           2        287100.04549     143550.02274
Residual            97        593999.92984       6123.71062

F =      23.44167       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

X1            12.000003    8.477988    .127200     1.415  .1601
X2             2.000000     .351726    .511003     5.686  .0000
(Constant)      .000000    7.825414                 .000 1.0000

End Block Number   1   All requested variables entered.
```

```
Multiple R           .62706
R Square             .39320
Adjusted R Square    .38701
Standard Error     22.47448

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           1         32075.99826      32075.99826
Residual            98         49500.00066        505.10205

F =      63.50400       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

X1             9.000000    1.129385    .627060     7.969  .0000
(Constant)      .000000    2.247448                 .000 1.0000


End Block Number   1   All requested variables entered.

* * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 2    Dependent Variable..   X3

Block Number  1.  Method:  Enter      X1       X2

Variable(s) Entered on Step Number
   1..    X2
   2..    X1

Multiple R           .69481
R Square             .48276
Adjusted R Square    .47209
Standard Error     78.25415

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           2        554400.06262     277200.03131
Residual            97        594000.01964       6123.71154

F =      45.26667       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

X1            11.999998    5.048221    .222834     2.377  .0194
X2             2.000000     .351726    .533046     5.686  .0000
(Constant)      .000000    7.825415                 .000 1.0000

End Block Number   1   All requested variables entered.
```

```
* * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 3    Dependent Variable..   X4

Block Number  1.  Method:  Enter      X1       X2       X3

Variable(s) Entered on Step Number
   1..    X3
   2..    X1
   3..    X2

Multiple R           .42713
R Square             .18244
Adjusted R Square    .15689
Standard Error    304.65145

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           3       1988316.12327     662772.04109
Residual            96       8910000.55774      92812.50581

F =       7.14098       Signif F =  .0002

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

X1            29.999994   33.344791    .090419      .900  .3705
X2             2.000001    1.581139    .145297     1.265  .2090
X3             1.000000     .395285    .284337     2.530  .0130
(Constant)      .000000   30.465145                 .000 1.0000
```

```
* * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 3    Dependent Variable..   X4

Block Number  1.  Method:  Enter      X1       X2       X3

Variable(s) Entered on Step Number
   1..    X3
   2..    X1
   3..    X2

Multiple R           .54655
R Square             .29872
Adjusted R Square    .27680
Standard Error    304.65141

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           3       3795262.90617    1265087.63539
Residual            96       8909998.04938      92812.47968

F =      13.63058       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

X1            30.000009   20.217564    .167485     1.484  .1411
X2             1.999999    1.581139    .160258     1.265  .2090
X3             1.000000     .395285    .300646     2.530  .0130
(Constant)      .000000   30.465141                 .000 1.0000
```

EVILS OF STANDARDIZATION. From the above, and from our previous work, and from the homework to come, we can note the following problems with standardized variables:

- If the original metric is “meaningful,” (e.g. income in dollars as opposed to, say, an arbitrarily scaled 9 point attitudinal index), the standardized (path) coefficients are generally less intuitively meaningful than the structural (metric) coefficients

- Comparisons across populations can easily be distorted with path coefficients. Similarly, comparisons of parameters within a model can be distorted. All the sorts of hypothesis testing we have been doing about equality of parameters within a model and across populations generally are not meaningful with path coefficients

- If the dependent variable is measured with random error, the path coefficients will be biased downward in magnitude. The structural coefficients will not be. (Consider the simple case of when a flawed Y is regressed on a single X.)

- Suppose a model is perfectly specified, but the weighting of cases is wrong, e.g. there are a disproportionately large number of minorities in the sample. Metric/structural coefficients will not be biased by the improper weighting, but path/standardized coefficients will. Or, suppose that, across time, the minority population grows relative to the majority population. The path coefficients can change even though the structural coefficients do not.

NOTE: Metric coefficient does not necessarily = structural coefficient. It is only structural if the model is correctly specified. Mis-specified models can also distort comparisons within models and across populations even if the coefficients are not standardized.