# Invariance in Factor Analysis

Steven W. Nydick

University of Minnesota

May 21, 2012

# Outline

# Univariate Selection

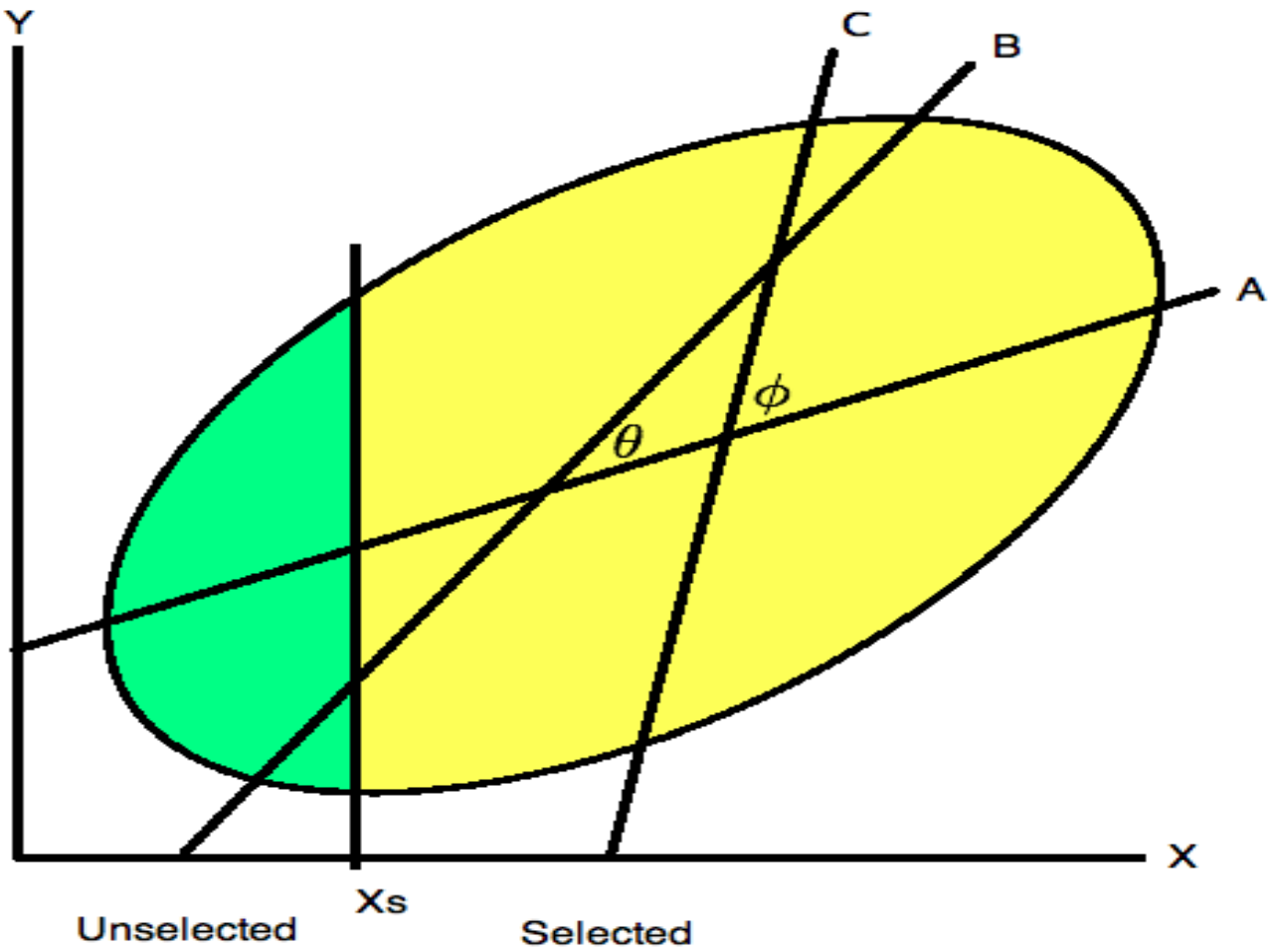
What is univariate selection?

1. We have a population with a population correlation.
2. We retain scores based on one of the variables.
3. How much do we know about the original population?

*How much remains unchained when systematically selecting subsets from a population?*

# Univariate Selection

The idea in pictures (Mulaik, 2009, Figure 14.2):



# Univariate Selection

Using the previous graphic:

1. $X_s$ is the selection line. Everything to the left of $X_s$ is *not selected* (for inclusion), and everything to the right of $X_s$ is *selected*.
   - The full population consists of the entire elipse.
   - The sub-population consists of the yellow area.
2. $A$ is the regression line of $Y$ on $X$.
3. $B$ is the regression line of $X$ on $Y$ (in the full population).
4. $C$ is the regression line of $X$ on $Y$ (in the sub-population).

5. $\theta$ is the angle between $B$ and $A$.
6. $\phi$ is the angle between $C$ and $A$.

# Univariate Selection: Divergences

Consequences of the previous picture (if selecting on $X$):

1. The slope of the $X \sim Y$ regression is *not* invariant under selection from a sub-population.
   - $C$ is steeper than $B$.
2. The angles $\phi$ and $\theta$ relate to the correlation between the population variables and the subpopulation variables.
3. Because $\phi$ is larger than $\theta$, the cosine of $\phi$ is smaller (and hence results in a smaller correlation) than the cosine of $\theta$.

# Univariate Selection: Divergences

**Mulaik p. 410**

As a matter of fact, if we can assume, in a parent population, that two variables are linearly related, we can conclude that the correlation between these two variables will be closer, or at least as close, to zero in any subpopulation selected on one of these variables than the corresponding correlation in the parent population.

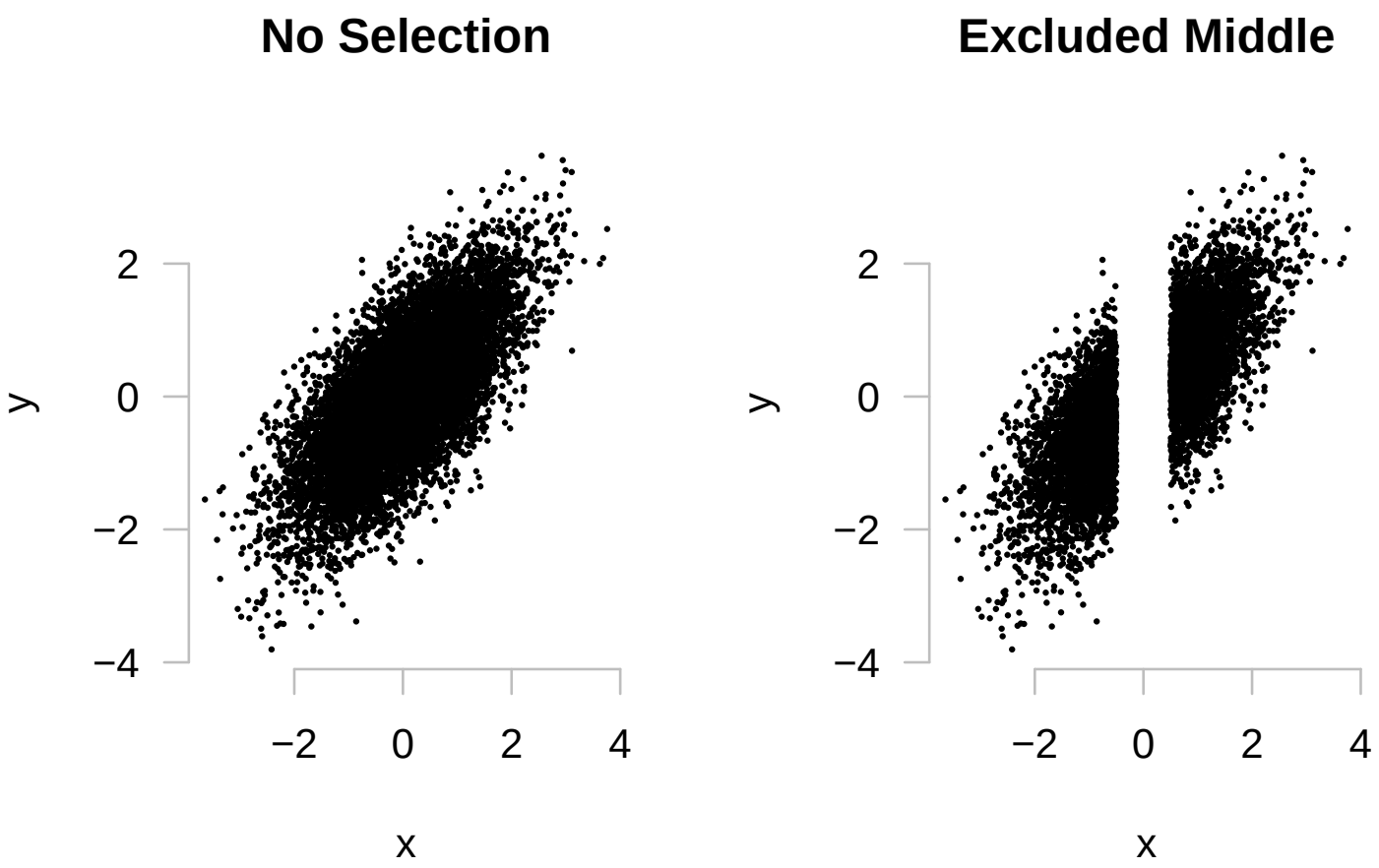
Mulaik is only referring to selection by *chopping off the ends* and not *removing the middle*.

# R-Code Demonstration (Code)

```
> # Set the mean and variance structure:
> mu    <- c(0, 0)
> Sigma <- matrix(c(1.0, 0.7,
+                   0.7, 1.0),
+                 nrow = 2)
> # Simulating data to fit the structure (exactly):
> dat.a <- as.data.frame( mvrnorm(10000,
+                                 mu = mu, Sigma = Sigma,
+                                 empirical = TRUE) )
> names(dat.a) <- c("x", "y")
> # Removing the x's between -.5 and .5:
> dat.b <- with( dat.a,
+                subset( dat.a, (x < -.5) | (x > .5) ) )
```

# R-Code Demonstration (Graphics)

A scatterplot with/without selection:



# R-Code Demonstration (Results)

```
> # Excluded middle raises correlation...
> with(dat.a, cor(x, y)) # should be smaller

[1] 0.7

> with(dat.b, cor(x, y)) # should be larger

[1] 0.7773072

> # ... but shouldn't change regression slope.
> lm(y ~ x, data = dat.a)$coef[2]

  x
0.7

> lm(y ~ x, data = dat.b)$coef[2]

        x
0.7015478
```

# Univariate Selection: Equivalences

Consequences of selecting on $X$:

1. The slope of the $Y \sim X$ regression is invariant under selection from a subpopulation.

$$E(Y|X) = A_0 + A_1 X$$
$$E(y|x) = a_0 + a_1 x$$

where $(Y, X)$ are the parent population & $(y, x)$ are the subpopulation. Therefore

$$A_1 = \left(\frac{S_Y}{S_X}\right) R_{XY} = \left(\frac{s_y}{s_x}\right) r_{xy} = a_1$$

# Univariate Selection: Equivalences

2. Due to homoscedasticity, the variation around the regression line will also be the same regardless of selection.

   Because (in the full and sub-populations)

$$R^2_{XY} = 1 - \frac{S^2}{S^2_Y} \qquad\qquad r^2_{XY} = 1 - \frac{s^2}{s^2_Y}$$

   then (in the full and sub-populations)

$$S^2 = S^2_Y(1 - R^2_{XY})$$
$$s^2 = s^2_y(1 - r^2_{xy})$$

   so (in the full and sub-populations)

$$S^2 = S^2_Y(1 - R^2_{XY}) = s^2_y(1 - r^2_{xy}) = s^2$$

# Correcting for Range Restriction

Therefore, selecting on $X$ does not change the slope of the $Y \sim X$ regression line nor the variance of the points around that regression line. Of course, invariance only happens with *population* data.

Now pretend that we have a selected area (a sub-population), and we desire to predict $R_{XY}$ in the full population. Based on slope equivalences

$$\left(\frac{S_Y}{S_X}\right) R_{XY} = \left(\frac{s_y}{s_x}\right) r_{xy}$$

$$S_Y = \left(\frac{S_X}{R_{XY}}\right)\left(\frac{s_y}{s_x}\right) r_{xy}$$

$$= \frac{S_X\, s_y\, r_{xy}}{s_x R_{XY}}$$

# Correcting for Range Restriction

And based on standard error equivalences

$$S_Y^2(1 - R_{XY}^2) = s_y^2(1 - r_{xy}^2)$$

$$\left(\frac{S_X s_y r_{xy}}{s_x R_{XY}}\right)^2 (1 - R_{XY}^2) = s_y^2(1 - r_{xy}^2)$$

$$\frac{S_X^2 s_y^2 r_{xy}^2}{s_x^2 R_{XY}^2}(1 - R_{XY}^2) = s_y^2(1 - r_{xy}^2)$$

$$\frac{S_X^2 r_{xy}^2}{s_x^2 R_{XY}^2}(1 - R_{XY}^2) = 1 - r_{xy}^2$$

$$\frac{S_X^2 r_{xy}^2}{s_x^2 R_{XY}^2} - \frac{S_X^2 r_{xy}^2}{s_x^2} = 1 - r_{xy}^2$$

$$\frac{S_X^2 r_{xy}^2}{s_x^2 R_{XY}^2} = 1 - r_{xy}^2 + \frac{S_X^2 r_{xy}^2}{s_x^2}$$

# Correcting for Range Restriction

Continuing:

$$\frac{S_X^2 r_{xy}^2}{s_x^2 R_{XY}^2} = 1 - r_{xy}^2 + \frac{S_X^2 r_{xy}^2}{s_x^2}$$

$$\frac{S_X^2 r_{xy}^2}{s_x^2 R_{XY}^2} = \frac{s_x^2 - r_{xy}^2 s_x^2 + r_{xy}^2 S_X^2}{s_x^2}$$

$$\frac{S_X^2 r_{xy}^2}{s_x^2 R_{XY}^2} = \frac{s_x^2 - r_{xy}^2 (s_x^2 - S_X^2)}{s_x^2}$$

$$\frac{S_X^2 r_{xy}^2}{R_{XY}^2} = s_x^2 - r_{xy}^2 (s_x^2 - S_X^2)$$

$$\frac{R_{XY}^2}{S_X^2 r_{xy}^2} = \frac{1}{s_x^2 - r_{xy}^2 (s_x^2 - S_X^2)}$$

# Correcting for Range Restiction

Finishing:

$$\frac{R_{XY}^2}{S_X^2 r_{xy}^2} = \frac{1}{s_x^2 - r_{xy}^2(s_x^2 - S_X^2)}$$

$$R_{XY}^2 = \frac{S_X^2 r_{xy}^2}{s_x^2 - r_{xy}^2(s_x^2 - S_X^2)} = \frac{\left(\frac{S_X^2}{s_x^2}\right) r_{xy}^2}{1 - r_{xy}^2\left(1 - \frac{S_X^2}{s_x^2}\right)}$$

Therefore if we: (1) have a bivariate normal distribution, (2) know the selection mechanism, and (3) know the correlation between $x$ and $y$ in the sub-population, we can find $R_{XY}^2$.

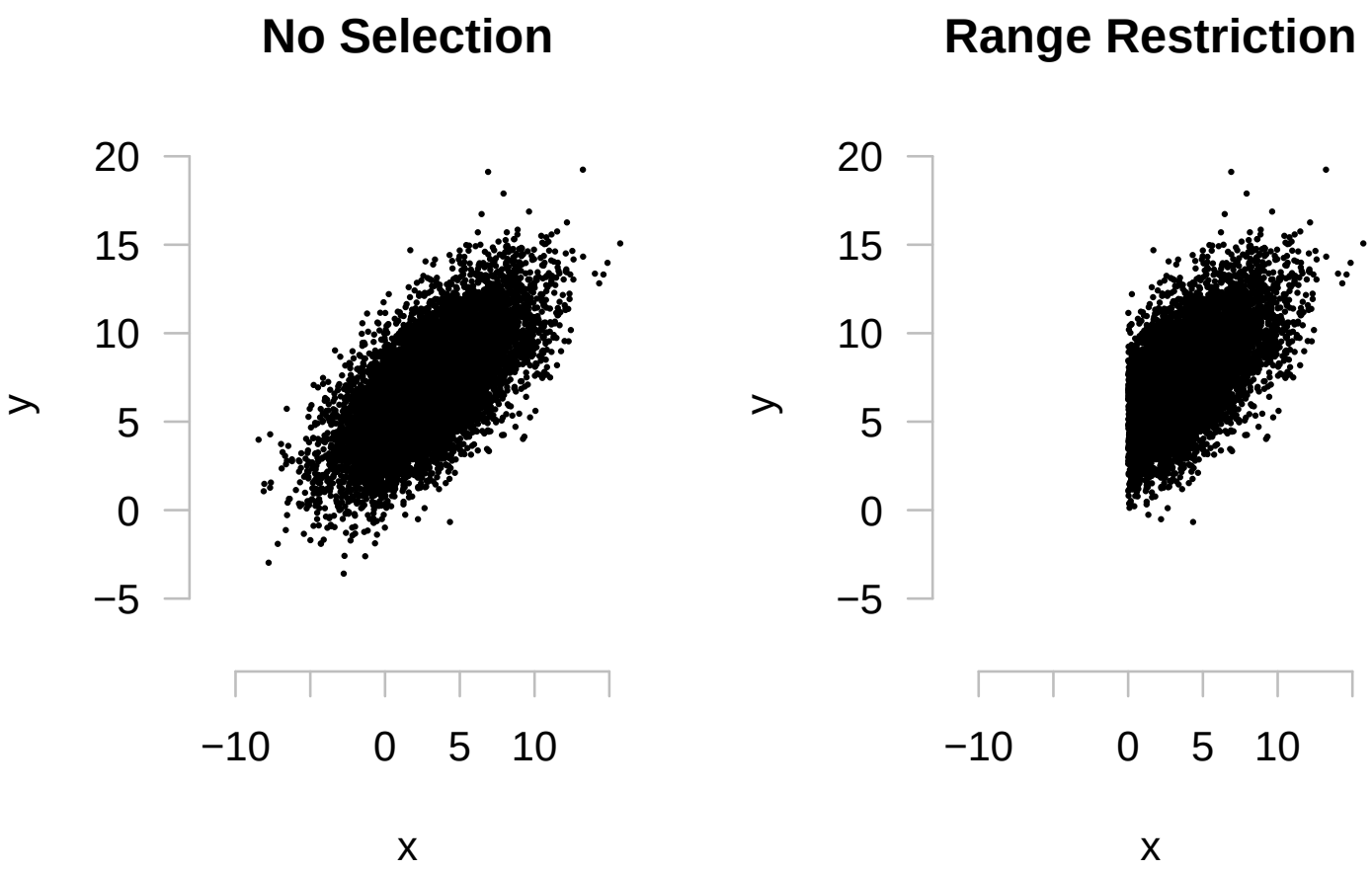
$$R_{XY}^2 = \frac{\left(\frac{S_X^2}{s_x^2}\right) r_{xy}^2}{1 - r_{xy}^2\left(1 - \frac{S_X^2}{s_x^2}\right)}$$

# R-Code Demonstration (Code)

```
> # Set the mean and variance structure:
> mu    <- c(3, 7)
> Sigma <- matrix(c(10,  6,
+                    6,  8),
+                  nrow = 2)
> # Simulating data to fit the structure (exactly):
> dat2.a <- as.data.frame( mvrnorm(1000000,
+                                  mu = mu, Sigma = Sigma,
+                                  empirical = TRUE) )
> names(dat2.a) <- c("x", "y")
> # Removing the x's between -.5 and .5:
> dat2.b <- with( dat2.a,
+                 subset( dat2.a, x > 0 ) )
```

# R-Code Demonstration (Graphics)

A scatterplot with/without selection:



# R-Code Demonstration (Results)

```
> # Calculating variances and sub-population corr:
> Sx2  <- var(dat2.a$x)                # total population
> sx2  <- var(dat2.b$x)                # sub-population
> rxy2 <- with(dat2.b, cor(x, y)^2) # sub-population
> # The actual squared correlation:
> ( Rxy2.act <- with(dat2.a, cor(x, y)^2) )

[1] 0.45

> # The estimated squared correlation:
> ( Rxy2.est <- ((Sx2/sx2)*rxy2)/(1 - rxy2*(1 - Sx2/sx2)) )

[1] 0.4498509

>
> # Are they close?  The same?
```

# Multivariate Selection

How can we generalize selection to more than one predictor variable and (potentially) more than one criterion?

Let

1. $\mathbf{X}$ be a $p \times 1$ random vector with $E(\mathbf{X}) = \mathbf{0}$,
2. $\mathbf{Y}$ be a $q \times 1$ random vector with $E(\mathbf{Y}) = \mathbf{0}$,
3. $E(\mathbf{X}\mathbf{Y}') = \mathbf{\Sigma}_{XY}$, $E(\mathbf{X}\mathbf{X}') = \mathbf{\Sigma}_{XX}$, and $E(\mathbf{Y}\mathbf{Y}') = \mathbf{\Sigma}_{YY}$, and
4. selection statistics be in lowercase (e.g. $\boldsymbol{\sigma}$, $\mathbf{b}$, etc.).

Furthermore, assume that selection only takes place on the predictor(s).

# Multivariate Selection: Equivalences

Imagine the earlier selection picture, but in many more dimensions.

OK. Stop. Not everyone can imagine in $p \cdot q$ dimensions ⌣̈.

Multivariate Extensions:

1. Because all variables are mean centered, the regression "slopes" are identical in the full and sub-populations.

$$\mathbf{B}_{XY} = \boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY} = \boldsymbol{\sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xy} = \mathbf{b}_{xy}$$

2. Moreover, the var/cov matrix of errors-or-prediction are identical in the full and sub-populations.

$$E(\mathbf{E}\mathbf{E}') = E(\mathbf{Y}\mathbf{Y}') - E(\hat{\mathbf{Y}}\hat{\mathbf{Y}}') = E(\mathbf{y}\mathbf{y}') - E(\hat{\mathbf{y}}\hat{\mathbf{y}}') = E(\mathbf{e}\mathbf{e}')$$
$$\boldsymbol{\Sigma}_{EE} = \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY} = \boldsymbol{\sigma}_{yy} - \boldsymbol{\sigma}_{yx}\boldsymbol{\sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xy} = \boldsymbol{\sigma}_{ee}$$

# Multivariate Selection: Equivalences

Assume the following:

1. The error variances matrix is a matrix. We have the same number of variables on the left and right sides of the equation.
2. The functional form of the conditional mean is linear.
3. We have population data.
4. The scale of the variables do not change under selection. The units always mean the same thing, the error variance is the same, and the regression line does not change.
    - i.e., we are working with unstandardized regression coefficients.

# Multivariate Selection: Consequences

If all of the assumptions hold, then

$$\boldsymbol{\Sigma}_{EE} = \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY} = \boldsymbol{\sigma}_{yy} - \boldsymbol{\sigma}_{yx}\boldsymbol{\sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xy} = \boldsymbol{\sigma}_{ee}$$

and

$$\mathbf{B}_{XY} = \boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY} = \boldsymbol{\sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xy} = \mathbf{b}_{xy}$$

so

$$\begin{aligned}
\boldsymbol{\sigma}_{yy} - \boldsymbol{\sigma}_{yx}\boldsymbol{\sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xy} &= \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY} \\
\boldsymbol{\sigma}_{yy} - \mathbf{b}_{xy}'\boldsymbol{\sigma}_{xx}\mathbf{b}_{xy} &= \boldsymbol{\Sigma}_{YY} - \mathbf{B}_{XY}'\boldsymbol{\Sigma}_{XX}\mathbf{B}_{XY} \\
\boldsymbol{\sigma}_{yy} - \mathbf{B}_{xy}'\boldsymbol{\sigma}_{xx}\mathbf{B}_{xy} &= \boldsymbol{\Sigma}_{YY} - \mathbf{B}_{XY}'\boldsymbol{\Sigma}_{XX}\mathbf{B}_{XY} \\
\boldsymbol{\sigma}_{yy} &= \boldsymbol{\Sigma}_{YY} - \mathbf{B}_{XY}'\boldsymbol{\Sigma}_{XX}\mathbf{B}_{XY} + \mathbf{B}_{xy}'\boldsymbol{\sigma}_{xx}\mathbf{B}_{xy} \\
\boldsymbol{\sigma}_{yy} &= \boldsymbol{\Sigma}_{YY} + \mathbf{B}_{XY}'(\boldsymbol{\sigma}_{xx} - \boldsymbol{\Sigma}_{XX})\mathbf{B}_{XY}
\end{aligned}$$

# Multivariate Selection: Consequences

What does the ultimate equation mean?

$$\boldsymbol{\sigma}_{yy} = \boldsymbol{\Sigma}_{YY} + \mathbf{B}'_{XY}(\boldsymbol{\sigma}_{xx} - \boldsymbol{\Sigma}_{XX})\mathbf{B}_{XY}$$

We can determine the criterion var/cov matrix in a sub-population if...

(I) ... we know the var/cov matrix of the predictor variables (in the sub-population), and

(II) ... we know the prediction equation in the full population.

Note that the variable *units* must remain the same.

- Either the full or sub-population can have standardized variables.
- The *other* population cannot have standardized variables.
- A 5 must be a 5 must be a 5 ...

# Multivariate Selection: Consequences

What does the ultimate equation mean?

$$\boldsymbol{\sigma}_{yy} = \boldsymbol{\Sigma}_{YY} + \mathbf{B}'_{XY}(\boldsymbol{\sigma}_{xx} - \boldsymbol{\Sigma}_{XX})\mathbf{B}_{XY}$$

Let $\boldsymbol{\Sigma}_{YY}$ be diagonal: $\mathbf{D}$. Can $\boldsymbol{\sigma}_{yy} = \mathbf{d}$?

Mulaik (2009) says $\boldsymbol{\sigma}_{yy}$ will only be diagonal in rare events.

What would count as a rare event?

# Factorial Invariance

Let's move from regression to factor analysis.

> *Assume that we have two experimental populations, but we are factor analyzing the same set of variables ($\mathbf{Y}$) in each population? Will we obtain the same factors? Will we obtain the same factor loadings? We will obtain the same factor structure?*

Think of the above question as a special case of regression selection.

# Factorial Invariance: Prerequisites

Common Factor Analysis (CFA) usually specifies

$$\mathbf{Y} = \mathbf{\Lambda X} + \mathbf{\Psi E} \tag{1}$$

where

- $\mathbf{Y}$ is an $n \times 1$ random vector of observed variables,
- $\mathbf{X}$ is an $r \times 1$ random vector of $r$ common factors,
- $\mathbf{\Lambda}$ is an $n \times r$ matrix of factor-pattern coefficients,
- $\mathbf{\Psi}$ is an $n \times n$ diagonal matrix of unique-coefficients, and
- $\mathbf{E}$ is an $n \times 1$ random vector of $n$ unique factors.

We can assume: $E(\mathbf{Y}) = E(\mathbf{E}) = \mathbf{0}$ and $E(\mathbf{X}) = \mathbf{0}$.

We should not assume that we are working in the correlation metric.

# Factorial Invariance: Prerequisites

The other fundamental CFA equation is

$$\boldsymbol{\Sigma}_{YY} = \boldsymbol{\Lambda}\boldsymbol{\Phi}_{XX}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2 \tag{2}$$

where

- $\boldsymbol{\Sigma}$ is the $n \times n$ observed score covariance matrix in the full population,
- $\boldsymbol{\Phi}$ is the $r \times r$ common factor covariance matrix in the full population, and
- $\boldsymbol{\Psi}^2$ is the (diagonal) unique factor covariance matrix in the full population.

Note that $\boldsymbol{\Phi}_{XX}$ is not assumed to be diagonal.

# Factorial Invariance: The V Vector

Now consider a random vector ($\mathbf{V}$) such that

1. $\mathbf{V}$ is of length $p \times 1$,
2. $\mathbf{V}$ has been selected on to form a sub-population ($\mathbf{v}$),
3. the regression of $\mathbf{Y}$ *and* $\mathbf{X}$ on $\mathbf{V}$ is linear and homoscedastic,
4. the regression of **both** $\mathbf{Y}$ and $\mathbf{X}$ on $\mathbf{V}$ is linear and homoscedastic, and
5. the $p$ selection variables and $n$ unique factors are uncorrelated.

When will the last assumption hold?

> *The $p$ selection variables and $n$ unique factors will be uncorrelated if the unique variance (of $\boldsymbol{Y}$) is entirely error.*

The last assumption will only be guaranteed to happen if no **specific** factor variance is in the unique factors.

# Factorial Invariance: The V Vector

Assuming that $\mathbf{V}$ satisfies our assumptions, regress $\mathbf{X}$ and $\mathbf{Y}$ on $\mathbf{V}$.

$$\mathbf{Y} = \mathbf{B}'\mathbf{V} + \mathbf{U} = \mathbf{B}'\mathbf{V} + (\mathbf{E} + \mathbf{G})$$
$$\mathbf{X} = \boldsymbol{\beta}'\mathbf{V} + \mathbf{Q}$$

In the above equation

- $\mathbf{B}$ $(n \times p)$ predicts $\mathbf{Y}$ from $\mathbf{V}$,
- $\boldsymbol{\beta}$ $(r \times p)$ predicts $\mathbf{X}$ from $\mathbf{V}$,
- $\mathbf{U} = \mathbf{E} + \mathbf{G}$ (all $n \times 1$) are unique factors and common parts of $\mathbf{Y}$ not related to $\mathbf{V}$, and
- $\mathbf{Q}$ $(r \times 1)$ are aspects of $\mathbf{X}$ (the common factors) not related to $\mathbf{V}$.

Note that $\mathbf{G}$ is in terms of observed, and $\mathbf{Q}$ is in terms of factors.

# Factorial Invariance: The V Vector

Now replace $\mathbf{X}$ in the original equation with $\mathbf{X} = \boldsymbol{\beta}'\mathbf{V} + \mathbf{Q}$.

Then

$$\begin{aligned}
\mathbf{Y} &= \boldsymbol{\Lambda}\mathbf{X} + \mathbf{E} \\
\mathbf{Y} &= \boldsymbol{\Lambda}(\boldsymbol{\beta}'\mathbf{V} + \mathbf{Q}) + \mathbf{E} \\
\mathbf{Y} &= \boldsymbol{\Lambda}\boldsymbol{\beta}'\mathbf{V} + \boldsymbol{\Lambda}\mathbf{Q} + \mathbf{E} \\
\mathbf{Y} &= \mathbf{B}'\mathbf{V} + \mathbf{G} + \mathbf{E}
\end{aligned}$$

So that $\mathbf{B} = \boldsymbol{\beta}\boldsymbol{\Lambda}'$ and $\mathbf{G} = \boldsymbol{\Lambda}\mathbf{Q}$.

# Factorial Invariance: The V Vector

How does $\mathbf{V}$ relate to the multivariate $\mathbf{X}$ from regression?

- $\mathbf{V}$ are the predictor variables of both $\mathbf{Y}$ and $\mathbf{X}$.
- $\mathbf{Y}$ and $\mathbf{X}$ are the criterion variables.

And based on the multivariate selection equations

$$\boldsymbol{\sigma}_{yy} = \boldsymbol{\Sigma}_{YY} + \mathbf{B}'(\boldsymbol{\sigma}_{vv} - \boldsymbol{\Sigma}_{VV})\mathbf{B}$$
$$\boldsymbol{\phi}_{xx} = \boldsymbol{\Phi}_{XX} + \boldsymbol{\beta}'(\boldsymbol{\sigma}_{vv} - \boldsymbol{\Sigma}_{VV})\boldsymbol{\beta}$$

... which just used the earlier equation:

$$\boldsymbol{\sigma}_{yy} = \boldsymbol{\Sigma}_{YY} + \mathbf{B}'_{XY}(\boldsymbol{\sigma}_{xx} - \boldsymbol{\Sigma}_{XX})\mathbf{B}_{XY}$$

# Factorial Invariance: The V Vector

We already assume the following about the CFA model.

$$\boldsymbol{\Sigma}_{YY} = \boldsymbol{\Lambda}\boldsymbol{\Phi}_{XX}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2$$

And based on $\boldsymbol{\Sigma}_{YY}$, we can deduce

$$\begin{aligned}
\boldsymbol{\sigma}_{yy} &= \boldsymbol{\Sigma}_{YY} + \mathbf{B}'(\boldsymbol{\sigma}_{vv} - \boldsymbol{\Sigma}_{VV})\mathbf{B} \\
&= \boldsymbol{\Lambda}\boldsymbol{\Phi}_{XX}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2 + \mathbf{B}'(\boldsymbol{\sigma}_{vv} - \boldsymbol{\Sigma}_{VV})\mathbf{B} \\
&= \boldsymbol{\Lambda}\boldsymbol{\Phi}_{XX}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2 + \boldsymbol{\Lambda}\boldsymbol{\beta}'(\boldsymbol{\sigma}_{vv} - \boldsymbol{\Sigma}_{VV})\boldsymbol{\beta}\boldsymbol{\Lambda}' \\
&= \boldsymbol{\Lambda}\boldsymbol{\Phi}_{XX}\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\boldsymbol{\beta}'(\boldsymbol{\sigma}_{vv} - \boldsymbol{\Sigma}_{VV})\boldsymbol{\beta}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2 \\
&= \boldsymbol{\Lambda}(\boldsymbol{\Phi}_{XX} + \boldsymbol{\beta}'(\boldsymbol{\sigma}_{vv} - \boldsymbol{\Sigma}_{VV})\boldsymbol{\beta})\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2 \\
&= \boldsymbol{\Lambda}\boldsymbol{\phi}_{xx}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2
\end{aligned}$$

# Factorial Invariance: Consequences

Therefore, in the full population, we know

$$\boldsymbol{\Sigma}_{YY} = \boldsymbol{\Lambda}\boldsymbol{\Phi}_{XX}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2$$

and in a sub-population which was formed by directly selecting on a variable relating both to $\mathbf{X}$ and $\mathbf{Y}$, we know

$$\boldsymbol{\sigma}_{yy} = \boldsymbol{\Lambda}\boldsymbol{\phi}_{xx}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2$$

*If the variable are in the same metric and the selection asumptions are upheld, then the factor pattern matrix is invariant.*

# Factorial Invariance: Consequences

But if we scale both full and sub-populations to unit variance, then

$$\begin{aligned}
\mathbf{r}_{yy} &= \mathbf{d}_y^{-1}\boldsymbol{\sigma}_{yy}\mathbf{d}_y^{-1} \\
&= \mathbf{d}_y^{-1}(\boldsymbol{\Lambda}\boldsymbol{\phi}_{xx}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2)\mathbf{d}_y^{-1} \\
&= \mathbf{d}_y^{-1}\boldsymbol{\Lambda}(\boldsymbol{\phi}_{xx})\boldsymbol{\Lambda}'\mathbf{d}_y^{-1} + \mathbf{d}_y^{-1}\boldsymbol{\Psi}^2\mathbf{d}_y^{-1} \\
&= \mathbf{d}_y^{-1}\boldsymbol{\Lambda}(\mathbf{d}_x\mathbf{r}_{xx}\mathbf{d}_x)\boldsymbol{\Lambda}'\mathbf{d}_y^{-1} + \mathbf{d}_y^{-1}\boldsymbol{\Psi}^2\mathbf{d}_y^{-1} \\
&= (\mathbf{d}_y^{-1}\boldsymbol{\Lambda}\mathbf{d}_x)\mathbf{r}_{xx}(\mathbf{d}_x\boldsymbol{\Lambda}'\mathbf{d}_y^{-1}) + \mathbf{d}_y^{-1}\boldsymbol{\Psi}^2\mathbf{d}_y^{-1}
\end{aligned}$$

Therefore, in the full population, we know

$$\mathbf{R}_{YY} = \boldsymbol{\Lambda}\mathbf{R}_{XX}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2$$

but if both full and sub-populations are standardized, then

$$\mathbf{r}_{yy} \neq \boldsymbol{\Lambda}\mathbf{r}_{xx}\boldsymbol{\Lambda}' + \mathbf{d}_y^{-1}\boldsymbol{\Psi}^2\mathbf{d}_y^{-1}$$

# Factorial Invariance: Consequences

More aspects of invariance under selection.

1. If $(\mathbf{\Lambda})_{ij} = 0$, then $(\mathbf{d}_y^{-1}\mathbf{\Lambda}\mathbf{d}_x)_{ij} = 0$, and
2. if $(\mathbf{\Lambda})_{ij} \neq 0$, then $(\mathbf{d}_y^{-1}\mathbf{\Lambda}\mathbf{d}_x)_{ij} \neq 0$, so that
3. "simple structure" is preserved.

But the *covariance* structure matrices are different in both pop'ns.

$$E(\mathbf{Y}\mathbf{X}') = \mathbf{\Lambda}\mathbf{\Phi}_{XX} \neq \boldsymbol{\lambda}\boldsymbol{\phi}_{xx} = E(\mathbf{y}\mathbf{x}')$$

And the *correlation* structure matrices are different in both pop'ns.

$$\mathbf{R}_{YX} = \mathbf{D}_Y^{-1}\mathbf{\Lambda}\mathbf{\Phi}_{XX}\mathbf{D}_X^{-1} \neq \mathbf{d}_Y^{-1}\mathbf{\Lambda}\boldsymbol{\phi}_{XX}\mathbf{d}_X^{-1} = \mathbf{r}_{yx}$$

# R-Code Demonstration (Code)

```
> # Two factors (1/2), and three selection vars (3--5):
> mu    <- c(0, 0, 0, 0, 0)
> Sigma <- matrix(c(1.0, 0.0, 0.7, 0.2, 0.1,
+                   0.0, 1.0, 0.3, 0.8, 0.4,
+                   0.7, 0.3, 1.0, 0.2, 0.1,
+                   0.2, 0.8, 0.2, 1.0, .12,
+                   0.1, 0.4, 0.1, .12, 1.0),
+                 nrow = 5, byrow = TRUE)
> # Simulating data to fit the structure (exactly):
> dat3.1 <- as.data.frame( mvrnorm(100000,
+                                  mu = mu, Sigma = Sigma,
+                                  empirical = TRUE) )
> names(dat3.1) <- c("x1", "x2", "v1", "v2", "v3")
> # Creating a y-hat matrix of 6 variables:
> yhat <- with( dat3.1,
+               data.frame(y1 = 3*x1, y2 = 2*x1,   y3 = 2.5*x1,
+                          y4 = 2*x2, y5 = 1.8*x2, y6 = 1.4*x2) )
```

# R-Code Demonstration (Code)

```
> # Building a matrix of errors, by:
> # --> generating draws from a normal distribution, and
> e <- matrix( rnorm(length(unlist(yhat)), sd = 3),
+              ncol = ncol(yhat) )
> # --> regressing on dat3.a and saving residuals.
> e <- lm(e ~ x1 + x2 + v1 + v2 + v3, data = dat3.1)$resid
> # And adding the residuals to yhat to make y:
> dat3 <- cbind(yhat + e, dat3.1)
> # Standardizing and removing subsets based on v:
> dat3.a <- do.call( data.frame, lapply(dat3, FUN = scale) )
> dat3.b <- with( dat3.a,
+                 subset( dat3.a, v1 > 0 & v2 > 1 & v3 > 0 ) )
> # What percentage of our original data frame is left?
> nrow(dat3.b)/nrow(dat3.a)

[1] 0.05921
```

# R-Code Demonstration (Results)

```
> # Factor analyzing both "observed variable" matrices:
> fa3.a <- fa(dat3.a[ , 1:6], nfa = 2, rot = "promax")
> fa3.b <- fa(dat3.b[ , 1:6], nfa = 2, rot = "promax")
> L1.1 <- matrix(fa3.a$loadings, ncol = 2)
> L2.1 <- matrix(fa3.b$loadings, ncol = 2)
>
> # How do the loadings compare?
```

```
> round(L1.1, 2)
     [,1] [,2]
[1,] 0.70 0.00
[2,] 0.56 0.00
[3,] 0.64 0.00
[4,] 0.00 0.55
[5,] 0.00 0.52
[6,] 0.00 0.42
```

```
> round(L2.1, 2)
      [,1]  [,2]
[1,]  0.61  0.00
[2,]  0.48  0.01
[3,]  0.55 -0.02
[4,] -0.01  0.33
[5,]  0.00  0.38
[6,]  0.00  0.28
```

# R-Code Demonstration (Results)

```
> # Mulaik provided the equation to fix loadings:
> # --> Lamb2 = dy^(-1)*Lamb1*dx
> vy <- diag(diag(cov(dat3.b[ , 1:6])))  # var/cov of y
> vx <- diag(diag(cov(dat3.b[ , 7:8])))  # var/cov of x
> dy <- sqrt(vy); dx <- sqrt(vx)         # sd mat of x/y
> L2.2 <- solve(dy) %*% L1.1 %*% dx
>
> # How do the loadings compare?
```

```
> round(L2.2, 2)
     [,1] [,2]
[1,] 0.64 0.00
[2,] 0.49 0.00
[3,] 0.56 0.00
[4,] 0.00 0.39
[5,] 0.00 0.37
[6,] 0.00 0.29
```

```
> round(L2.1, 2)
      [,1]  [,2]
[1,]  0.61  0.00
[2,]  0.48  0.01
[3,]  0.55 -0.02
[4,] -0.01  0.33
[5,]  0.00  0.38
[6,]  0.00  0.28
```

# Across Populations: Prerequisites

Given two populations, we want to decide whether

1. the same factors account for the underlying relationships in both populations or
2. different factors account for the relationships in each population.

# Across Populations: Things to Consider

Remember, the factor pattern matrix ($\mathbf{\Lambda}$) is invariant.

1. If the units mean the same thing in both groups.
   - We use covariance matrices instead of correlation matrices.
   - We standardize in a reference group and apply the same linear transformation to the other group.
   - We standardize across both groups.
2. If we use correlation matrices and transform the factor pattern.
   - $\mathbf{\Lambda}_2 = \mathbf{d}_y^{-1}\mathbf{\Lambda}_1\mathbf{d}_x$
   - $\mathbf{\Lambda}_1 = \mathbf{d}_y\mathbf{\Lambda}_2\mathbf{d}_x^{-1}$
   - $\mathbf{d}_y$ and $\mathbf{d}_x$ are the standard deviations associated with group 2 when group 1 is standardized.
   - $\mathbf{d}_y$ and $\mathbf{d}_x$ are the ratios of the group 2 variances to group 1 variances regardless of standardization.

# Across Populations: Things to (Not) Consider

But do not

1. compare two factor patterns when both are based off of correlation matrices,
2. compare the factor structure matrices,
3. rotate the factor pattern using an orthogonal rotation, or
4. pick populations selected on variables that correlate with the unique components of $\mathbf{Y}$.

Just remember:

> *Compare pattern (not structure), do not use orthogonal rotation, and make sure that the error is error. Then the population pattern matrices will be invariant.*

# References


- Mulaik, S. A. (2009). *Foundations of factor analysis, second edition.* Boca Raton, FL: CRC Press.