

Structural Equation Modeling: Part 2

Paul D. Allison, Ph.D.

Upcoming Seminar:
August 16-September 13, 2021, On Demand

6. FORMATIVE INDICATORS

Formative Indicators

The standard approach is to presume that multiple indicators are **caused** by some latent variable--**reflective indicators**. Correlations among reflective indicators tell us something about the strength of their relationship with the latent variable(s) on which they depend.

In some cases, we might want to assume **formative indicators** which *cause* the latent variable. Problem: correlations among the indicators don't give us any useful information. Consequently, these models often have identification problems.

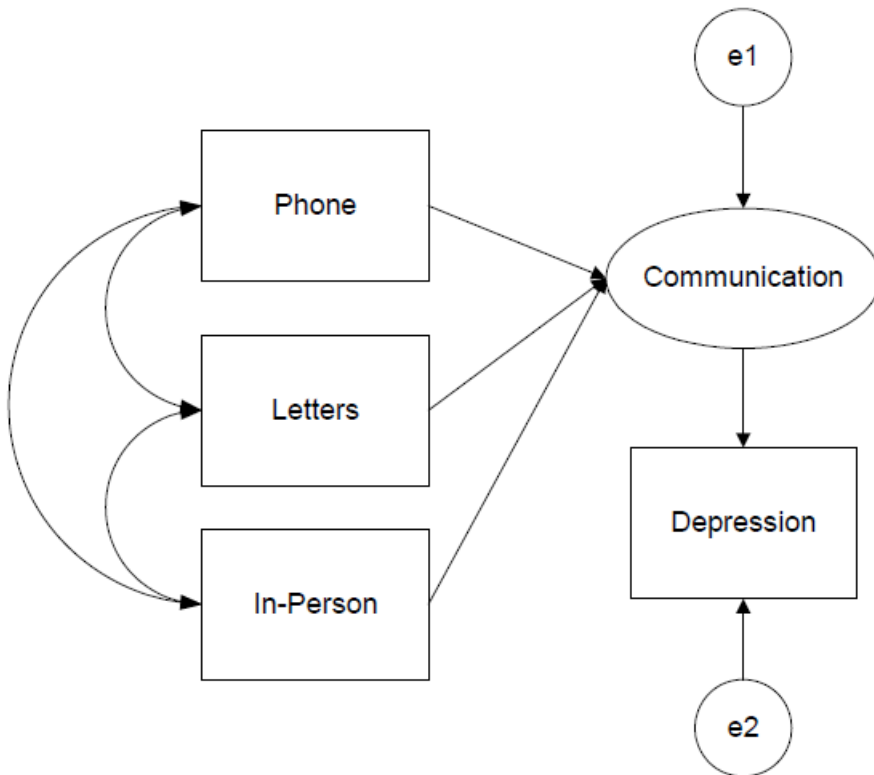
Example: You want to measure the “volume of communication between parent and adult child”. You want to know how this variable affects level of depression for the parent.

You have the following indicators of communication:

- Number of phone calls per week
- Number of letters or e-mails exchanged per month
- Time spent in face-to-face conversation

Formative Indicators 2

Is it reasonable to assume that these three measures are *caused* by some latent variable, “communication”? Probably not. Rather, they are alternative methods of communication that may even be negatively correlated. We could represent this as:

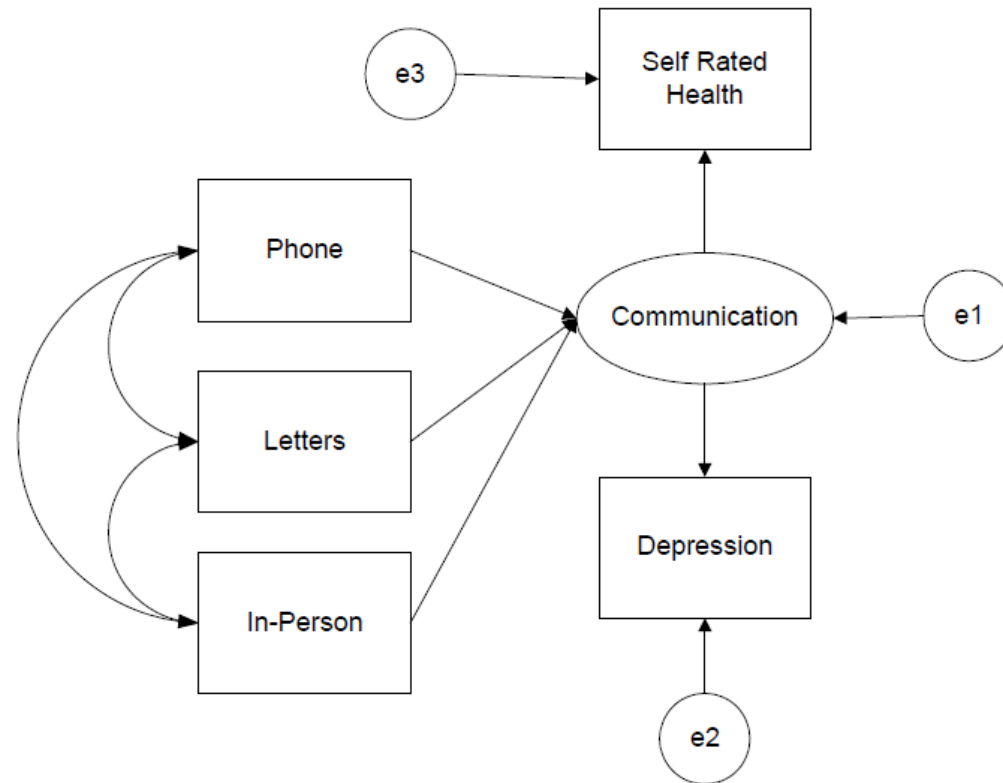


First select one of the "indicators" as a reference indicator, with a coefficient of 1. But this model is still not identified.

How to make it identified? Two alternatives:

- Set $\text{Var}(e1)=0$. Assumes that the three indicators exhaust the means of communication.
- Add another consequence of communication.

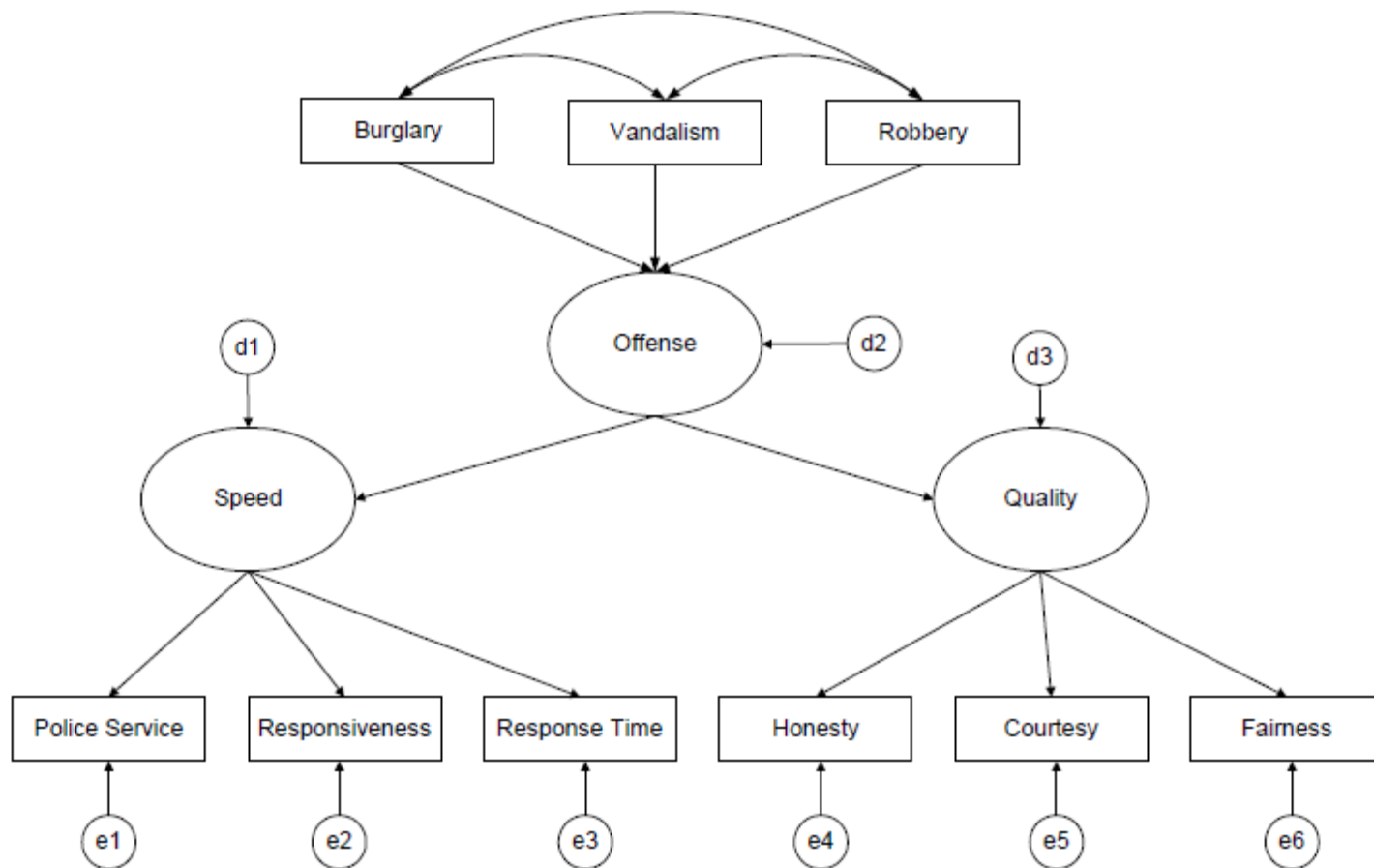
Formative Indicators 3



Now the model is identified. This type of model is known as a MIMIC model (multiple indicator, multiple cause). Note that it requires the assumption that self-rated health and depression have a partial correlation of 0, controlling for the latent variable communication

Police Example

11,000 people who had been crime victims were surveyed about their experience with police. McIver, John P., Edward G. Carmines, and Richard A. Zeller. "Multiple indicators." *Measurement in the Social Sciences* (1980): 162-186.



Mplus for Police

DATA:

FILE IS c:\data\police.txt;

TYPE IS CORR; NOBS IS 11000;

VARIABLE:

NAMES ARE service response time honesty courtesy
fairness burglary vandal robbery;

MODEL:

speed BY service response time;

quality BY honesty courtesy fairness;

offense BY speed quality;

offense ON burglary vandal robbery;

OUTPUT: STDYX;

Mplus Police Results

WARNING: THE LATENT VARIABLE COVARIANCE MATRIX (PSI) IS NOT POSITIVE DEFINITE. THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE FOR A LATENT VARIABLE, A CORRELATION GREATER OR EQUAL TO ONE BETWEEN TWO LATENT VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO LATENT VARIABLES. CHECK THE TECH4 OUTPUT FOR MORE INFORMATION.
PROBLEM INVOLVING VARIABLE SPEED.

MODEL FIT INFORMATION

Chi-Square Test of Model Fit

| | |
|--------------------|---------|
| Value | 203.783 |
| Degrees of Freedom | 22 |
| P-Value | 0.0000 |

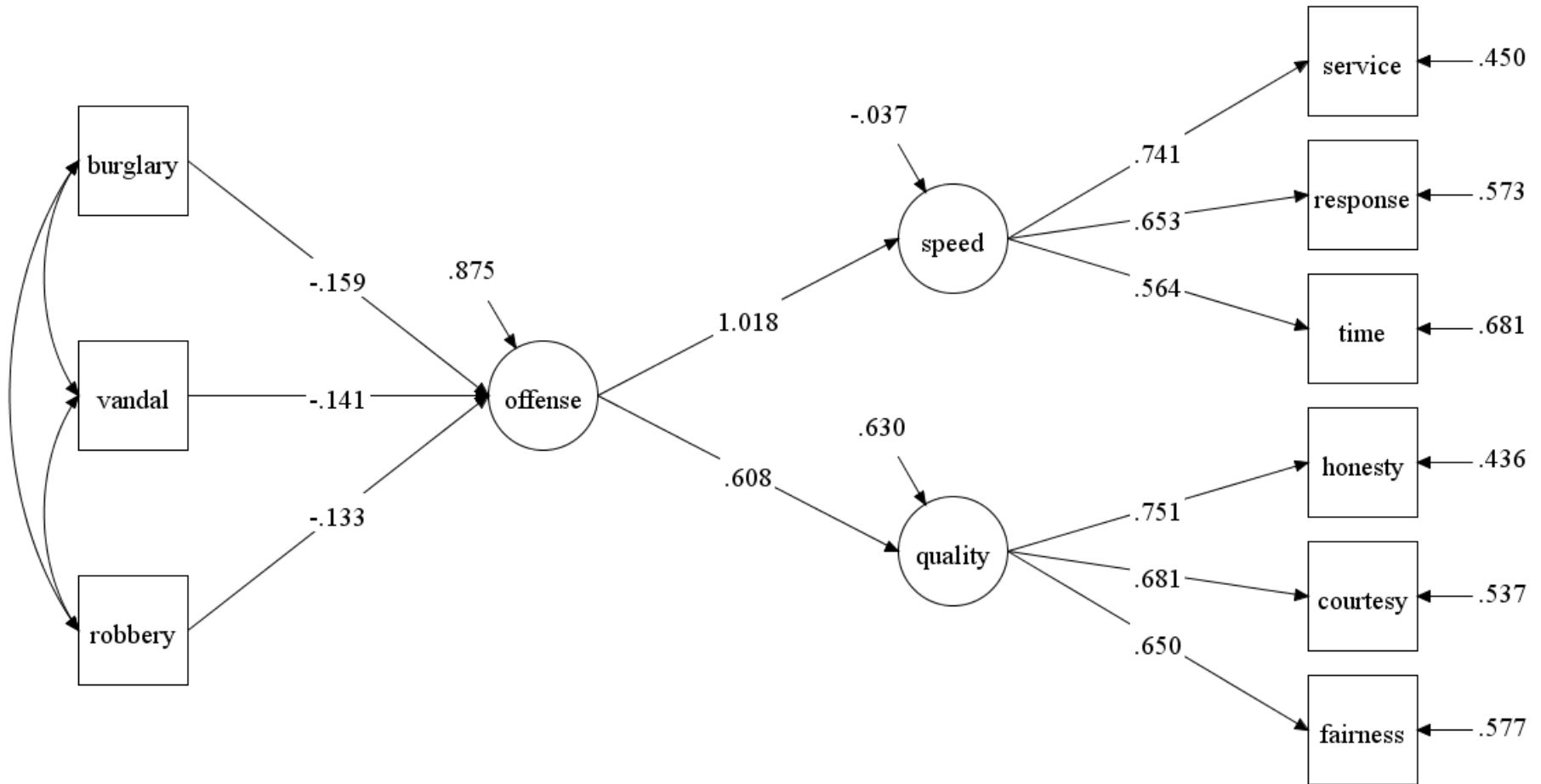
RMSEA (Root Mean Square Error Of Approximation)

| | |
|--------------------------|-------------|
| Estimate | 0.027 |
| 90 Percent C.I. | 0.024 0.031 |
| Probability RMSEA <= .05 | 1.000 |

CFI/TLI

| | |
|-----|-------|
| CFI | 0.989 |
| TLI | 0.983 |

Mplus Police Standardized Diagram



Stata for Police

```
clear
ssd init service response time honesty courtesy fairness burglary
    vandal robbery
ssd set obs 373
ssd set cor 1.00 \ .50 1.00 \ .41 .35 1.00 \ .33 .29 .30 1.00 \ .28 .26 .27 .52 1.00 \ .30 .27 .29 .48
.44 1.00 \ -.24 -.19 -.17 -.13 -.11 -.15 1.00 \ .23 -.19 -.16 -.11 -.09 -.13 .58 1.00 \ -.20 -.18 -.14 -.15 -.10
-.13 .47 .42 1.00
save "c:\data\police.dta"
```

```
use "c:\data\police.dta"
sem (Speed -> service response time)
    (Quality -> honesty courtesy fairness)
    (Offense -> Speed Quality)
    (Offense <- burglary vandal robbery)
```

But this does not converge.

lavaan for Police

```
lower<- '
1.00
.50 1.00
.41 .35 1.00
.33 .29 .30 1.00
.28 .26 .27 .52 1.00
.30 .27 .29 .48 .44 1.00
-.24 -.19 -.17 -.13 -.11 -.15 1.00
-.23 -.19 -.16 -.11 -.09 -.13 .58 1.00
-.20 -.18 -.14 -.15 -.10 -.13 .47 .42 1.00 '
corr<-getCov(lower,names=c("service","response","time","honesty", "courtesy",
    "fairness","burglary","vandal","robbery"))
police.mod<- '
speed =~ service+response+time
quality =~ honesty+courtesy+fairness
offense =~ speed+quality
offense ~ burglary+vandal+robbery '
police.fit<-sem(police.mod,sample.cov=corr, sample.nobs=11000)
summary(police.fit, stand=T)
```

Get a Heywood problem.

SAS for Police

```
DATA my.police(TYPE=CORR);
INPUT service response time honesty courtesy fairness burglary
      vandal robbery;
DATALINES;
1.00 .....
.50 1.00 .....
.41 .35 1.00 .....
.33 .29 .30 1.00 .....
.28 .26 .27 .52 1.00 ....
.30 .27 .29 .48 .44 1.00 ...
-.24 -.19 -.17 -.13 -.11 -.15 1.00 ..
-.23 -.19 -.16 -.11 -.09 -.13 .58 1.00 .
-.20 -.18 -.14 -.15 -.10 -.13 .47 .42 1.00
PROC CALIS DATA=my.police PLOTS=PATHDIAGRAM NOBS=11000;
PATH speed -> service response time = 1,
      quality -> honesty courtesy fairness = 1,
      offense -> speed quality =1,
      offense <- burglary vandal robbery;
RUN;
```

If you only have correlations or covariances, you can often get by without the TYPE or NAME variables.

Mplus with Reflective Indicators

DATA:

FILE IS c:\data\police.txt;
TYPE IS CORR; NOBS IS 11000;

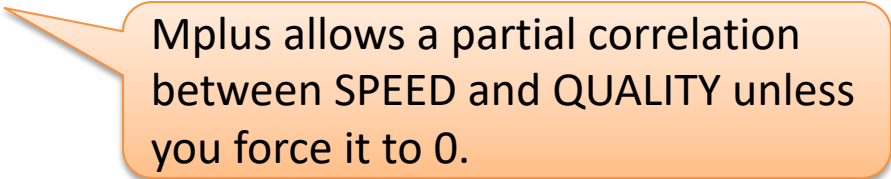
VARIABLE:

NAMES ARE service response time honesty courtesy
fairness burglary vandal robbery;

MODEL:

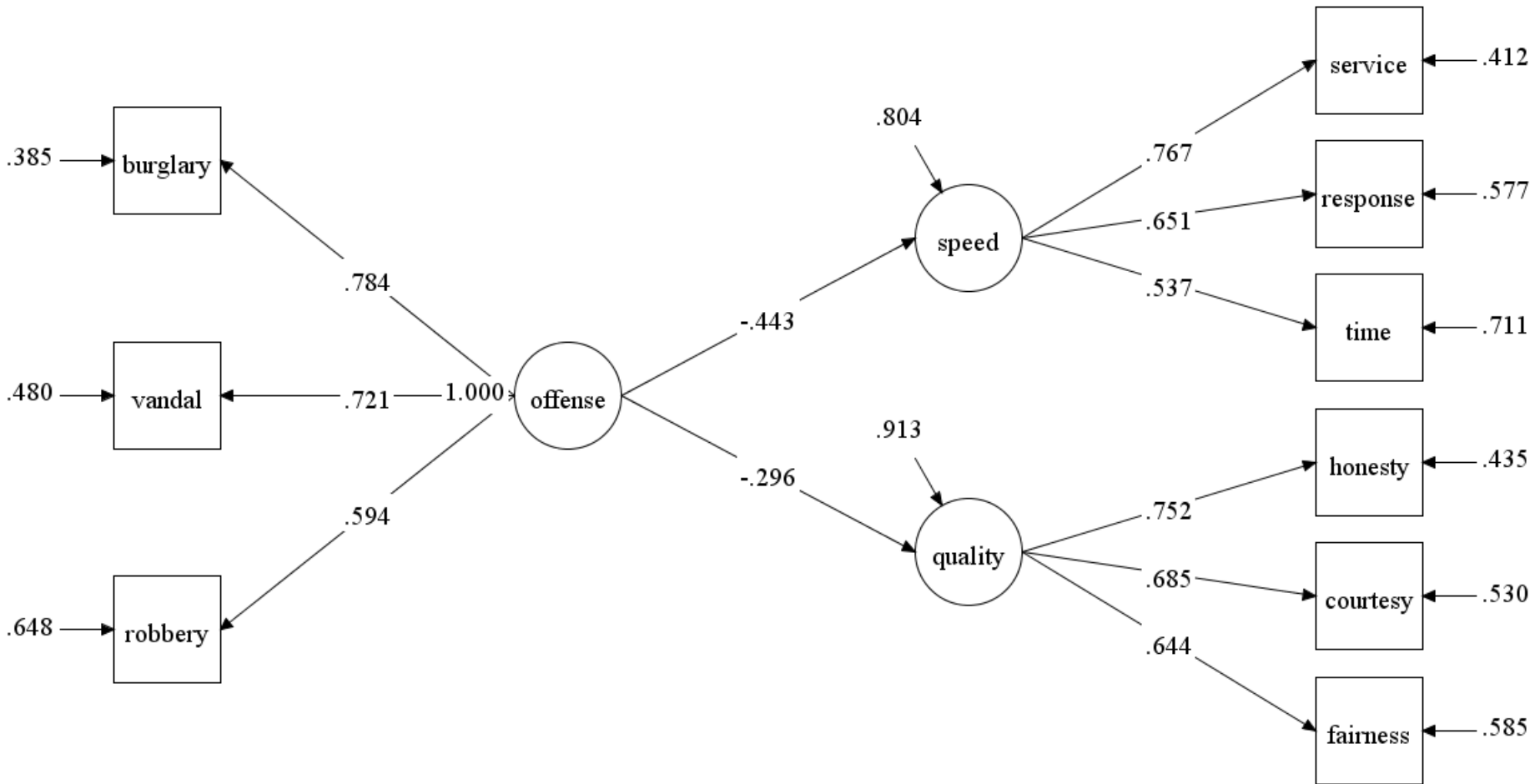
speed BY service response time;
quality BY honesty courtesy fairness;
offense BY burglary vandal robbery;
speed quality ON offense;
speed WITH quality@0;

OUTPUT: STDYX;



Mplus allows a partial correlation between SPEED and QUALITY unless you force it to 0.

Reflective Police Diagram



No Heywood problem, but not a great fit (RMSEA=.09). You can get a great fit by allowing a partial correlation between SPEED and QUALITY.

What to Do When a Model Doesn't Fit

Possibly nothing, especially if the sample size is large and the N -invariant measures of fit look good.

Isolate the problem:

- Remove all restrictions on relationships among latent variables by fitting the corresponding CFA model. If it fits OK, then the problem is in the structural part. The chi-square for the structural part is the difference between the CFA chi-square and the overall chi-square.
- If the problem is in the structural part, try adding additional paths.
- If you have four or more indicators for some latent variables, try fitting measurement models for just those variables, one at a time.
- Try fitting a model to subparts of the whole model, e.g., if you have five latent variables, try models for two at a time.
- Problems in measurement part can sometimes be solved by deleting entire variables.

Modification indices and residuals can be useful, but use cautiously. Keep in mind the possibility that additional latent variables may be needed (which won't be obvious from modification indices).

7. ALTERNATIVE ESTIMATION METHODS

Alternative Estimation Methods

Most SEM packages offer estimation methods that are alternatives to ML. We'll begin with a method that can be implemented with either summary data or individual data.

Generalized Least Squares

Here the function being minimized is

$$tr[(\mathbf{S}^{-1}(\mathbf{S} - \hat{\Sigma}))^2]$$

Like ML, under multivariate normality, GLS estimators are consistent, asymptotically normal and asymptotically efficient. However, these properties also hold under somewhat less restrictive assumptions.

- To implement this method in Mplus, use the ESTIMATOR=GLS option on the ANALYSIS command.
- In SAS, put METHOD=GLS on PROC statement.
- In lavaan, put **estimator="GLS"** in the sem function.
- Not available in Stata.