Three models for combining information from causal indicators

The `sheafcoef` and `propcnsreg` package

Maarten L. Buis

Institut für Soziologie
Eberhard Karls Universität Tübingen
http://www.maartenbuis.nl

Introduction

- Sometimes we have multiple variables that measure the same latent concept.

Introduction

- Sometimes we have multiple variables that measure the same latent concept.
- For example,
  - a set of question that measure someone's IQ or degree of depression, or

Introduction

- Sometimes we have multiple variables that measure the same latent concept.
- For example,
  - a set of question that measure someone's IQ or degree of depression, or
  - someone's education and occupation may measure someone's socioeconomic status.

Introduction

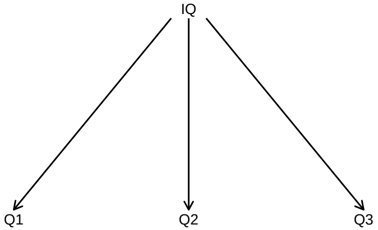
- Sometimes we have multiple variables that measure the same latent concept.
- For example,
  - a set of question that measure someone's IQ or degree of depression, or
  - someone's education and occupation may measure someone's socioeconomic status.
- This is a good thing! But...

Introduction

- Sometimes we have multiple variables that measure the same latent concept.
- For example,
  - a set of question that measure someone's IQ or degree of depression, or
  - someone's education and occupation may measure someone's socioeconomic status.
- This is a good thing! But, we need models to make the best use possible of this information.

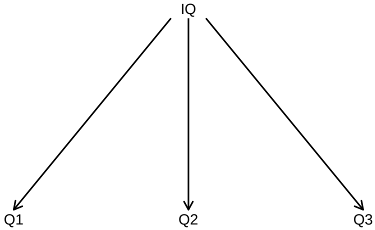
Effect indicators and causal indicators

- Effect indicators are variables that are influenced by the latent variable.



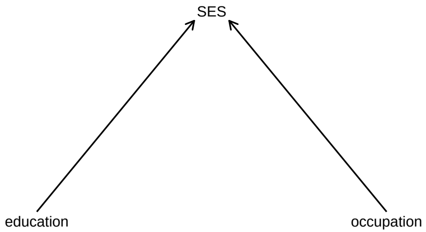
Effect indicators and causal indicators

- Effect indicators are variables that are influenced by the latent variable.
  - For example factor analysis (`factor`)



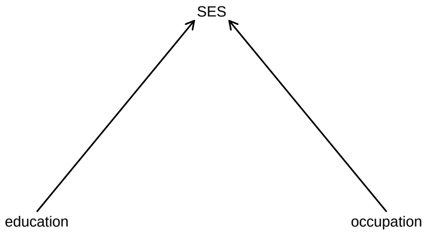
Effect indicators and causal indicators

- Effect indicators are variables that are influenced by the latent variable.
  - For example factor analysis (`factor`)
- Causal indicators are variables that influence the latent variable.



Effect indicators and causal indicators

- Effect indicators are variables that are influenced by the latent variable.
  - For example factor analysis (`factor`)
- Causal indicators are variables that influence the latent variable.
  - For example:
  - sheaf coefficients (`sheafcoef`),
  - parametrically weighted covariates, and
  - MIMIC models (`propcnsreg`).



The basic model

MIMIC

$$
\begin{aligned}
y &= \beta_0 + (\lambda_0 + \lambda_1 z_1)\eta + \varepsilon_y \\
\eta &= \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \varepsilon_\eta
\end{aligned}
$$

The basic model

parametrically weighted covariates

$$
\begin{aligned}
y &= \beta_0 + (\lambda_0 + \lambda_1 z_1)\eta + \varepsilon_y \\
\eta &= \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2
\end{aligned}
$$

The basic model

Sheaf coefficients

$$
\begin{aligned}
y &= \beta_0 + (\lambda_0 \qquad\qquad)\eta + \varepsilon_y \\
\eta &= \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2
\end{aligned}
$$

identification

- The empirical information we use to estimate the γs and λs is that we choose the γs to optimize the effect of η on y.

identification

- The empirical information we use to estimate the γs and λs is that we choose the γs to optimize the effect of η on y.
- The empirical information we use to estimate the variance of ε_η in the MIMIC model is that this model assumes that the total residual variance changes along z_1 according to $var(\varepsilon_y) + (\lambda_0 + \lambda_1 z_1)^2 \times var(\varepsilon_\eta)$

identification

- The empirical information we use to estimate the γs and λs is that we choose the γs to optimize the effect of η on y.
- The empirical information we use to estimate the variance of ε_η in the MIMIC model is that this model assumes that the total residual variance changes along z_1 according to $var(\varepsilon_y) + (\lambda_0 + \lambda_1 z_1)^2 \times var(\varepsilon_\eta)$
- η is a latent variable, so we need to fix its origin and its unit.
  - Fix the origin by setting η to 0 when x_1 and x_2 are both 0
  - Fix the unit by setting the standard deviation of η to 1.

Data preparation

```
. sysuse nlsw88, clear
(NLSW, 1988 extract)

. gen byte occ2 = occupation
(9 missing values generated)

. recode occ2 (2=1) (3 4 11 12 = 2) (5/10= 3) (13=.)
(occ2: 1920 changes made)

. label define occ2 1 "higher services" 2 "lower services" 3 "manual"

. label value occ2 occ2

.
.
. gen byte hs = grade == 12 if grade < .
(2 missing values generated)

. gen byte sc = grade > 12 & grade < 16 if grade < .
(2 missing values generated)

. gen byte c = grade >= 16 if grade < .
(2 missing values generated)

.
. replace tenure = tenure / 10
(2180 real changes made)

. gen white = race == 1 if race < .

.
. gen ln_w = ln(wage)
```

Sheaf coefficients after a linear regression

```
. qui xi: reg ln_w i.occ2 hs sc c
. sheafcoef, latent( _I* ; hs sc c) post
```

ln_w	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p1	.2000228	.0124272	16.10	0.000	.1756516	.224394
a1__Iocc2_2	-1.528682	.1075842	-14.21	0.000	-1.739668	-1.317696
a1__Iocc2_3	-2.600971	.0133063	-195.47	0.000	-2.627067	-2.574876
p2	.144066	.0124393	11.58	0.000	.119671	.168461
a2_hs	.9303067	.2141218	4.34	0.000	.5103867	1.350227
a2_sc	2.205349	.1904522	11.58	0.000	1.831848	2.57885
a2_c	3.031032	.133601	22.69	0.000	2.769024	3.293041
_cons	1.933329	.0378121	51.13	0.000	1.859174	2.007483

```
. test _b[p1] = _b[p2]
 ( 1)  p1 - p2 = 0
       F(  1,  2042) =     6.95
            Prob > F =    0.0084
```

Sheaf coefficients after logistic regression

```
. qui xi: logit union i.occ2 hs sc c
. sheafcoef, latent( _I* ; hs sc c) eform post
```

union	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
p1_e	1.241842	.0855004	14.52	0.000	1.074265	1.40942
a1__Iocc2_2	1.58573	.5031156	3.15	0.002	.5996415	2.571818
a1__Iocc2_3	2.585204	.1054152	24.52	0.000	2.378594	2.791814
p2_e	1.028296	.0661664	15.54	0.000	.8986119	1.15798
a2_hs	-1.15095	5.973281	-0.19	0.847	-12.85837	10.55647
a2_sc	.6553856	7.081814	0.09	0.926	-13.22471	14.53549
a2_c	1.394004	7.161541	0.19	0.846	-12.64236	15.43037
_cons_e	.2045564	.042083	4.86	0.000	.1220752	.2870376

```
(_e) indicates the variables whose coefficients have been exponentiated
. test _b[p1] = _b[p2]
 ( 1)  p1_e - p2_e = 0
           chi2(  1) =    6.02
         Prob > chi2 =    0.0142
```

Syntax of `sheafcoef`

```
sheafcoef,
latent( varlist_1 [ ; varlist_2 [; varlist_3 [...]]]  )
[ eform post iterate(#) level(#) ]
```

Parametrically weighted covariates

```
. propcnsreg ln_w white tenure, lambda(tenure white) ///
>           constrained(hs sc c) nolog
                                                Number of obs   =       2229
                                                LR chi2(8)      =     133.01
Log likelihood = -1607.2184                     Prob > chi2     =     0.0000

Constraint: sd of latent variables = 1
------------------------------------------------------------------------------
        ln_w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
unconstrai~d |
       white |   .2176303   .0366948     5.93   0.000     .1457098    .2895508
      tenure |   .3317353   .0330047    10.05   0.000     .2670473    .3964233
       _cons |   1.252169   .0400622    31.26   0.000     1.173648    1.330689
-------------+----------------------------------------------------------------
constrained  |
          hs |   .6364459   .1429802     4.45   0.000     .3562099    .9166819
          sc |   1.931921   .1414771    13.66   0.000     1.654631    2.209211
           c |    2.75269   .0907335    30.34   0.000     2.574856    2.930525
-------------+----------------------------------------------------------------
lambda       |
      tenure |  -.0429628   .0199328    -2.16   0.031    -.0820303   -.0038952
       white |  -.0938623   .0249237    -3.77   0.000    -.1427118   -.0450128
       _cons |   .3049783   .0251357    12.13   0.000     .2557131    .3542434
-------------+----------------------------------------------------------------
sigma        |
       _cons |   .4976345   .0074532    66.77   0.000     .4830266    .5122424
------------------------------------------------------------------------------
LR test vs. unconstrained model: chi2(4) =       3.22    Prob > chi2 =     0.522

BIC(unconstrained) - BIC(constrained) =     19.91
This difference suggests very strong evidence for the constrained model
```

MIMIC model

```
. propcnsreg ln_w white tenure, lambda(tenure white) ///
>            constrained(hs sc c) mimic nolog
                                                  Number of obs   =       2229
                                                  LR chi2(8)      =     137.63
Log likelihood = -1587.8862                       Prob > chi2     =     0.0000

Constraint: sd of latent variables = 1
------------------------------------------------------------------------------
        ln_w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
unconstrai~d |
       white |   .1154214   .0275711     4.19   0.000      .061383    .1694599
      tenure |    .354109   .0309777    11.43   0.000     .2933937    .4148243
       _cons |   1.290095   .0384749    33.53   0.000     1.214685    1.365504
-------------+----------------------------------------------------------------
constrained  |
          hs |   .7559966   .1473374     5.13   0.000     .4672207    1.044773
          sc |   2.039394   .1383171    14.74   0.000     1.768298    2.310491
           c |   2.805831   .0899889    31.18   0.000     2.629456    2.982206
-------------+----------------------------------------------------------------
lambda       |
      tenure |  -.0658272   .0182428    -3.61   0.000    -.1015825    -.030072
       white |  -.0035393   .0108898    -0.33   0.745    -.0248829    .0178044
       _cons |   .2547694   .0198169    12.86   0.000      .215929    .2936097
-------------+----------------------------------------------------------------
sigma        |
       _cons |   .3016388   .0579338     5.21   0.000     .1880907    .4151869
-------------+----------------------------------------------------------------
sigma_latent |
       _cons |   .4684396   .0384153    12.19   0.000     .3931471    .5437321
------------------------------------------------------------------------------
```

Syntax of `propcnsreg`

`propcnsreg` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] ,

`constrained(`*varlist*`)` `lambda(`*varlist*`)` [

`standardized` `lcons` `unit(`*varname*`)`

`mimic`

`robust` `cluster(`*varname*`)` `level(`*#*`)`

em_maximize_options *maximize_options*]

Conclusion (1)

- Causal indicators require a different strategy to recover the latent variable than effect indicators

Conclusion (1)

- Causal indicators require a different strategy to recover the latent variable than effect indicators
- Models with causal indicators recover the latent variable by scaling the observed indicators to optimize the effect of the latent variable on the dependent variable.

Conclusion (1)

- Causal indicators require a different strategy to recover the latent variable than effect indicators
- Models with causal indicators recover the latent variable by scaling the observed indicators to optimize the effect of the latent variable on the dependent variable.
- A MIMIC model also recovers measurement error by making a parametric assumption on how the total residual variance changes over observed variables.

Conclusion (2)

- Three models have been discussed:

Conclusion (2)

- Three models have been discussed:

  Sheaf coefficients: no measurement error, effect of latent variable is constant

Conclusion (2)

- Three models have been discussed:

  Sheaf coefficients no measurement error, effect of latent variable is constant

  Parametrically weighted covariates no measurement error, effect of latent variable changes over observed variables

Conclusion (2)

- Three models have been discussed:

  Sheaf coefficients no measurement error, effect of latent variable is constant

  Parametrically weighted covariates no measurement error, effect of latent variable changes over observed variables

  MIMIC model measurement error, effect of latent variable changes over observed variables

Conclusion (2)

- Three models have been discussed:
  - Sheaf coefficients: no measurement error, effect of latent variable is constant
  - Parametrically weighted covariates: no measurement error, effect of latent variable changes over observed variables
  - MIMIC model: measurement error, effect of latent variable changes over observed variables
- The model with sheaf coefficients can be estimated using `sheafcoef`,

Conclusion (2)

- Three models have been discussed:
  - Sheaf coefficients: no measurement error, effect of latent variable is constant
  - Parametrically weighted covariates: no measurement error, effect of latent variable changes over observed variables
  - MIMIC model: measurement error, effect of latent variable changes over observed variables
- The model with sheaf coefficients can be estimated using `sheafcoef`,
- the model with parametrically weighted covariates and the MIMIC model can be estimated using `propcnsreg`.

References

Bollen, Kenneth A.
Multiple Indicators: Internal Consistency or No Necessary Relationship.
Quality and Quantity, 18(4): 377-385, 1984.

Bollen, Kenneth A. and Richard Lennox.
Conventional Wisdom on Measurement: A Structural Equation Perspective.
Psychological Bulletin, 110(2): 305-314, 1991.

Hauser, Robert M. and Arthur S. Goldberger.
The Treatment of Unobservable Variables in Path Analysis.
Sociological Methodology 3: 81-117, 1971.

Heise, David R.
Employing nominal variables, induced variables, and block variables in path analysis.
Sociological Methods & Research 1(2): 147-173, 1972.

Yamaguchi, Kazuo.
Regression models with parametrically weighted explanatory variables.
Sociological Methodology 32: 219-245, 2002.