Comparing coefficients of nested nonlinear probability models using `khb`

Ulrich Kohler

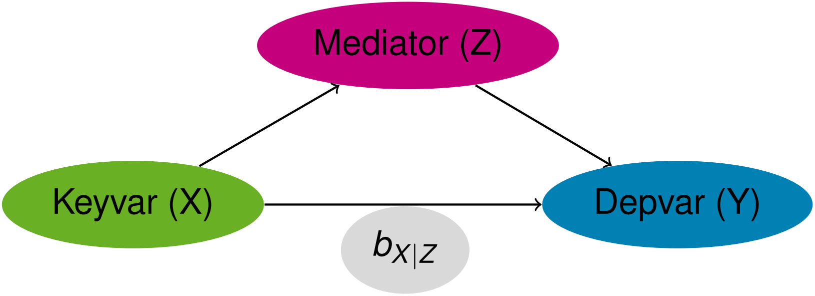
Joined Work with Kristian B. Karlson and Anders Holm

9th German Stata Users Group Meeting
Bamberg
01 July 2011

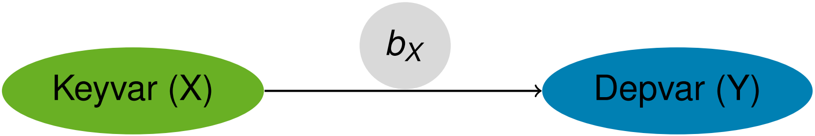
Outline

Outline

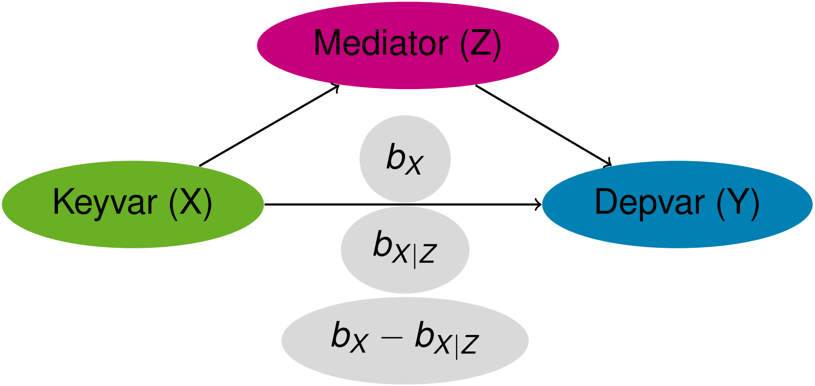
Reasons to compare



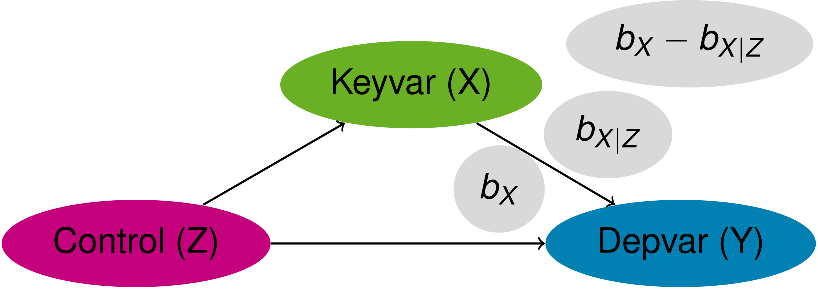
Reasons to compare



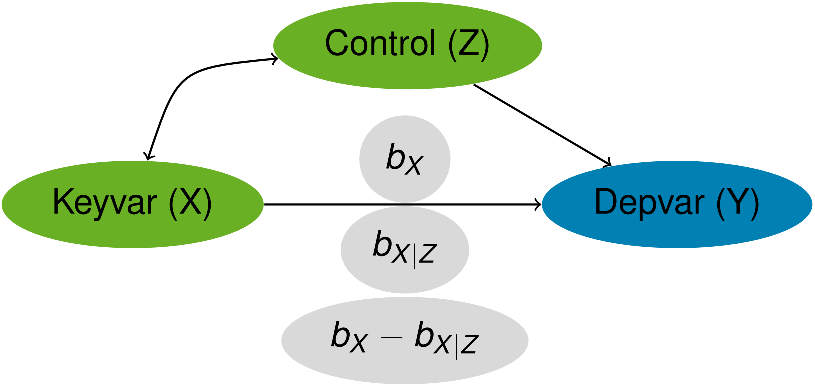
Reasons to compare



Reasons to compare



Reasons to compare



The problem

We are interested in obtaining $\beta_R - \beta_F$ from the following models for latent Y^*:

$$Y^* = \alpha_F + \beta_F X + \gamma_F Z + \delta_F C + \epsilon \quad (1)$$

$$Y^* = \alpha_R + \beta_R X + \delta_R C + \varepsilon \quad (2)$$

Having ovserved Y with value 0 if $Y^* < \tau$ and 1 if $Y^* \geq \tau$ we can obtain the logit/probit estimates with

$$b_F = \frac{\beta_F}{\sigma_F} \quad \text{and} \quad b_R = \frac{\beta_R}{\sigma_R} \quad (3)$$

Note: We identify the underlying coefficients of interest relative to a scale unknown to us.

Outline

General idea

The KHB-method extracts from Z the information that is not contained in X. This is done by calculating the residuals of a linear regression of Z on X, i.e,

$$R = Z - (a + bX) \quad , \tag{4}$$

where a and b are the estimated regression parameters of a linear regression.

Instead of using equation (2) we then use

$$Y^* = \widetilde{\alpha}_R + \widetilde{\beta}_R X + \widetilde{\gamma}_R R + \widetilde{\delta}_R C + \epsilon \quad . \tag{5}$$

Difference of coefficients

As R and Z differ only in the component in Z that is correlated with X, model (1) is no more predictive than model (5), and consequently the residuals have the same standard deviation so that

$$\widetilde{\sigma}_R = \sigma_F \tag{6}$$

As $\widetilde{\beta}_R = \beta_R$ we can write

$$\widetilde{b}_R - b_F = \frac{\widetilde{\beta}_R}{\widetilde{\sigma}_R} - \frac{\beta_F}{\sigma_F} = \frac{\beta_R - \beta_F}{\sigma_F} \quad . \tag{7}$$

Hence, the difference obtained reflects the difference searched divided by some common scale.

Derived statistics

Confounding ratio

$$\frac{\widetilde{b}_R}{b_F} = \frac{\frac{\beta_R}{\sigma_F}}{\frac{\beta_F}{\sigma_F}} = \frac{\beta_R}{\beta_F} \quad , \tag{8}$$

Counfounding percentage

$$100 \cdot \frac{\widetilde{b}_R - b_F}{\widetilde{b}_R} = 100 \cdot \frac{\frac{\beta_R}{\sigma_F} - \frac{\beta_F}{\sigma_F}}{\frac{\beta_R}{\sigma_F}} = 100 \cdot \frac{\beta_R - \beta_F}{\beta_R} \quad , \tag{9}$$

Significance test for the difference in effects

- Analyitcally derived standard errors for the difference in effects exist.
- Based on the delta method (Sobel, 1982).
- Simple for one X and ond Z but fairly complicated for situations with more than one X, Z.
- Karlson et al. (2010) has more details; also see our Stata Journal publications (in Press)

Outline

Syntax

`khb` *model-type depvar key-vars* || *mediator-vars* [*if*] [*in*] [, *options*]

model-type can be any of `regress`, `logit`, `ologit`, `probit`, `oprobit`, `cloglog`, `slogit`, `scobit`, `rologit`, `clogit`, and `mlogit`.

key-vars may contain factor variables

`aweights`, `fweights`, `iweights`, and `pweights` are allowed if they are allowed for the specified model type.

Options (most important ones)

options	description
`concomitant(`*varlist*`)`	concomitants
`disentangle`	disentangle difference of effects
`summary`	summary of decomposition
`vce(`*vcetype*`)`	robust or cluster clustvar
`ape`	decomposition using avg. partial effects
`verbose`	show restricted and full model
`keep`	keep residuals of mediators

Outline

Preliminaries

- Examples from educational sociology
- Subset of Danish National Longitudinal Survey (DLSY).
- Reproduce analysis presented by Karlson and Holm (2011).

```
. use dlsy_khb, clear
. describe
Contains data from dlsy_khb.dta
  obs:         1,896
 vars:             8                          17 Jan 2011 10:26
 size:        49,296 (99.9% of memory free)
------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
edu             byte   %20.0g      edu        Educational attainment
upsec           byte   %10.0g      yesno      Complete upper secondary
                                                education (Gymnasium)
univ            byte   %13.0g      yesno      Complete University education
fgroup          byte   %9.0g       fgroup     Father´s social group/class
fses            float  %9.0g                  Father´s SES, standardized with
                                                mean 0 and sd 1
abil            double %10.0g                 Standardized ability measure,
                                                with mean 0 and sd 1
intact          byte   %9.0g       yesno      Intact family
boy             byte   %9.0g       yesno      Boy
------------------------------------------------------------------------------
Sorted by:
```

Basic use

```
. khb logit univ fses || abil, c(intact boy)
Decomposition using the KHB-Method
Model-Type:  logit                                  Number of obs     =    1896
Variables of Interest: fses                         Pseudo R2         =    0.19
Z-variable(s): abil
Concomitant: intact boy
------------------------------------------------------------------------------
        univ |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
fses         |
     Reduced |   .5459815   .0779806     7.00   0.000     .3931424    .6988206
        Full |   .3817324   .0778061     4.91   0.000     .2292353    .5342295
        Diff |   .1642491   .0293249     5.60   0.000     .1067734    .2217247
------------------------------------------------------------------------------
```

Confounding ratio/percentage

```
. khb logit univ fses || abil, c(intact boy) summary notable
Decomposition using the KHB-Method
Model-Type:  logit                                 Number of obs     =    1896
Variables of Interest: fses                        Pseudo R2         =    0.19
Z-variable(s): abil
Concomitant: intact boy
Summary of confounding
```

Variable	Conf_ratio	Conf_Pct	Resc_Fact
fses	1.4302727	30.08	1.0602422

Option ape

```
. khb logit univ fses || abil, c(intact boy) ape summary
Decomposition using the APE-Method
Model-Type:  logit                                 Number of obs     =    1896
Variables of Interest: fses                        Pseudo R2         =    0.19
Z-variable(s): abil
Concomitant: intact boy
------------------------------------------------------------------------------
        univ |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
fses         |
     Reduced |   .0384906   .0054429     7.07   0.000     .0278226    .0491585
        Full |   .0269113   .0054476     4.94   0.000     .0162343    .0375884
        Diff |   .0115792   .0020667     5.60   0.000     .0075286    .0156298
------------------------------------------------------------------------------
 Note: Standard errors of difference not known for APE method
Summary of confounding
    Variable | Conf_ratio    Conf_Pct   Dist_Sens
-------------+------------------------------------
        fses |  1.4302727       30.08   .95931864
--------------------------------------------------
```

Disentangle contributions of mediators

```
. khb logit univ fses || abil intact boy, s d not
Decomposition using the KHB-Method
Model-Type:  logit                                  Number of obs     =     1896
Variables of Interest: fses                         Pseudo R2         =     0.19
Z-variable(s): abil intact boy
Summary of confounding

    Variable | Conf_ratio    Conf_Pct   Resc_Fact
-------------+--------------------------------------
        fses |  1.5207722       34.24   1.1317064
----------------------------------------------------

Components of Difference

  Z-Variable |      Coef    Std_Err      P_Diff  P_Reduced
-------------+-----------------------------------------------
fses         |
        abil |  .1661177   .0301003       83.56      28.61
      intact |   .020142   .0144611       10.13       3.47
         boy |  .0125359    .011524        6.31       2.16
-------------------------------------------------------------
```

More than one key variable

```
. khb logit univ boy intact || abil, c(fses) s
Decomposition using the KHB-Method
Model-Type:  logit                                 Number of obs     =     1896
Variables of Interest: boy intact                  Pseudo R2         =     0.19
Z-variable(s): abil
Concomitant: fses
```

univ	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
boy						
Reduced	1.06178	.1848087	5.75	0.000	.6995613	1.423998
Full	.9821406	.1848351	5.31	0.000	.6198704	1.344411
Diff	.0796391	.133004	0.60	0.549	-.1810438	.3403221
intact						
Reduced	1.129767	.7386976	1.53	0.126	-.3180536	2.577588
Full	1.08391	.7386558	1.47	0.142	-.3638292	2.531648
Diff	.0458575	.1328438	0.35	0.730	-.2145116	.3062266

Summary of confounding

Variable	Conf_ratio	Conf_Pct	Resc_Fact
boy	1.0810873	7.50	1.0033213
intact	1.0423075	4.06	1.03542

Categorical variables

```
. xtile catabil = abil, n(4)
. tab catabil, gen(catabil)
. khb logit univ i.fgroup || catabil2-catabil4, c(intact boy) s d
```

Ordered outcome

```
. forv i = 1/3 {
  2.          quietly eststo: khb ologit edu fses || abil, out(`i´) ape s
  3. }

. esttab, scalars("ratio_fses Conf.-Ratio" "pct_fses Conf.-Perc.")
------------------------------------------------------------
                      (1)             (2)             (3)
                      edu             edu             edu
------------------------------------------------------------
fses
Reduced            -0.103***       0.0643***       0.0385***
                 (-11.33)         (10.72)          (9.27)

Full              -0.0755***       0.0472***       0.0283***
                  (-8.02)          (7.76)          (7.23)

Diff              -0.0272***       0.0170***       0.0102***
                  (-6.50)          (6.44)          (5.95)
------------------------------------------------------------
N                    1896            1896            1896
Conf.-Ratio         1.360           1.360           1.360
Conf.-Perc.         26.48           26.48           26.48
------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Multinomial outcome

```
. forv i = 2/3 {
  2.          quietly eststo: khb mlogit edu fses || abil,  out(`i´) base(1) s
  3. }

. esttab, scalars("ratio_fses Conf.-Ratio" "pct_fses Conf.-Perc.")
```

	(1) edu	(2) edu
fses		
Reduced	0.423***	0.779***
	(7.63)	(9.30)
Full	0.313***	0.552***
	(5.70)	(6.68)
Diff	0.109***	0.227***
	(5.93)	(6.04)
N	1896	1896
Conf.-Ratio	1.349	1.411
Conf.-Perc.	25.88	29.15

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Outline

References

Karlson, K. B. and A. Holm. 2011. Decomposing primary and secondary effects: A new decomposition method. *Research in Stratification and Social Mobility* 29: XXXX.

Karlson, K. B., A. Holm, and R. Breen. 2010. Comparing regression coefficients between models using logit and probit: a new method. Unpublished paper (currently under review).

Sobel, M. E. 1982. Asymptotic confidence intervals for indirect effects in structural equation models. In *Sociological Methodology 1982*, ed. L. S., 290–312. Washington D.C.: American Sociological Association.