

NOTES TO SEM LECTURES (in yellow: not explicitly discussed)

Harry BG Ganzeboom

Melbourne, October 27 2017

Lecture 3: Multi-group (cross-national, cross-national) analysis in SEM and GSEM

A traditional concern in cross-national analysis has been the invariance of (measurement) models between different societies. The traditional wisdom in this literature has been that you need to measure your concept in an identical way, before you can even start to compare parameters across contexts. (The contexts could be countries, but also men and women, different ethnic groups, different time periods, cohorts, etc.). I take issue with this wisdom on the following grounds¹:

- The invariance-testing literature is overly concerned with testing the null-hypothesis of NO difference in measurement models. Notice the difference in expectations: usually, we would hope to reject the H0, in invariance testing we would hope to NOT to reject H0. Small differences to the H0 situation should not bother us (in particular not when we have much power in our model), but the invariance-testing literature they do. One should rather look at variations in quantities of interest (such as means, structural coefficients) under different invariance scenario's, instead of the fit or test statistics.
- Measurement invariance is more important to estimating latent means than for estimating latent structural coefficients. In fact, I would recommend always to interpret half or fully standardized models, on the assumption that your different groups represent the underlying units in a conceptually similar way. This point of view is particularly relevant when comparing countries, using nationally representative samples. In completely standardized models, latent means vanish, all of the interest is in the latent correlational structures.
- In many applications a good case can be made for allowing differences in measurement models. E.g. what is exactly the problem if in one country the crude occupation measure would be a better measure of latent occupational status than the detailed measure, but in another country it would be the other way around? I see no problem in comparing structural coefficients, in particular not if these are completely standardized.

Invariance

The invariance-testing literature distinguishes four scenario's:

- Configural invariance: the same pattern of factor loadings (in a multifactorial structure) is applied.

¹ These issues are discussed in a similar way by: Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement Equivalence in Cross-National Research. *Annual Review of Sociology*, 40(1), 55–75.
<http://doi.org/10.1146/annurev-soc-071913-043137>

- Metric invariance: the measurement coefficients ('factor loadings') are constrained to be the same. This would make for the strict comparability of latent variances and structural coefficients.
- Scalar invariance: the residual variances of the measurement effects are constrained to be the same. This would make for the strict comparability of latent means.
- Error invariance: the residual variances of the latent variables are constrained to be the same.

Stata SEM includes two set of tools to implement these constraints: by keywords, and by group specification.

The keywords can follow a **ginvariant** statement on the options part:

scoef	Structural coefficients
scons	Structural intercepts
mcoef	Measurement coefficients
mcons	Measurement intercepts
serrvar	Structural error covariances
merrvar	Measurement error covariances
smerrvar	Covariances between structural and measurement errors
meanex	means of exogenous variables
covex	covariances of exogenous variables
all	all of the above
none	none of the above

The default is: **sem (..) , group(..) ginvariant(mcoeff mcons)**

The group specific model can be generated in the following way:

```
by cnr: sem (FOCC -> zfisei) (FOCC -> zfosei) (OCC1 -> zisei1) (OCC1 -> zosei1), var(FOCC@1)
var(OCC1@1) covar(e.zfosei*e.zosei1) iterate(50)
```

(The model does not converge for LV – Latvia. There may be something fishy about the fosei data in this country.)

Alternatively, we could have run:

```
sem (FOCC -> zfisei) (FOCC -> zfosei) (OCC1 -> zisei1) (OCC1 -> zosei1), var(FOCC@1) var(OCC1@1)
covar(e.zfosei*e.zosei1) iterate(50) group(cnr)
```

we can build up the model until (which illustrates how to make implement invariance by group: specification:

```
sem (FOCC -> zfisei@c) (FOCC -> zfosei@d) (OCC1 -> zisei1@c) (OCC1 -
> zosei1@d) , var(FOCC@1) var(OCC1@1) iterate(50) group(cnr)
covar(1: e.zfosei*e.zosei1@a) covar(2: e.zfosei*e.zosei1@a) covar(3:
e.zfosei*e.zosei1@a) covar(4: e.zfosei*e.zosei1@a) covar(5:
e.zfosei*e.zosei1@a) covar(6: e.zfosei*e.zosei1@a) covar(7:
e.zfosei*e.zosei1@a) covar(8: e.zfosei*e.zosei1@a) covar(9:
e.zfosei*e.zosei1@a) covar(10: e.zfosei*e.zosei1@a) ginvariant(mcons
mcoef merrvar) var(e.zfisei@ee) var(e.zfosei@ee) var(e.zisei1@ee)
var(e.zosei1@ee)
```

Table 3.1 gives an overview of some of the models estimates on the double occupation measurement model. Note in particular the step towards metric invariance (almost insignificant) and then the step towards scalar invariance (very significant). Despite the spectacular difference between these steps, the structural parameter of interest (the latent correlation between father's and first occupation) hardly changes, however we constrain the measurement part of the model. Not only the rank order of countries remains essentially unchanged, also the numerical values of this correlation, even if we constrain the measurement of father's and first occupation to be the same, or the measurement quality of crude and detailed measurement.

Multilevel modelling in GSEM

Stata SEM cannot do multilevel models, but generalized GSEM claims it can. However, GSEM is primarily targeted at doing another thing, which is the generalized linear model (GLM) for (observed) response variables and allows for a link function and a variety of distribution of the error terms. This allows you to use (binomial, multinomial, ordered) logistic regression or poisson regression to model the relationship between latent and observed variables. However, GSEM cannot do:

- MLMV
- Group comparisons
- Standardized solutions
- Multilevel with latent variables

Of course, the last option is in particular what we had hoped for. What remains is possibility to estimate multilevel models on observed models. Stata's main alternative here is ME (mixed estimation), which is considerably easier to handle. GSEM would offer opportunities to multilevel estimates in mediation models, but this is not very much of an extension. To show how you fit a very simple multilevel model with a cross-level interaction between $fisei \rightarrow ise1$ and a mystery country level variable hg , consider the following models:

```
generate fisei_hg = fisei*hg
regr ise1 fisei fisei_hg i.cnr
```

```

regr iseil fisei fisei_hg i.cnr, cluster(cnr)

xtreg iseil fisei fisei_hg i.cnr

xtreg iseil fisei fisei_hg i.cnr, cluster(cnr)

mixed iseil fisei fisei_hg || cnr:

mixed iseil fisei fisei_hg || cnr: , robust cluster(cnr)

gsem (iseil <- fisei fisei_hg M1[cnr])

gsem (iseil <- fisei fisei_hg M1[cnr]), vce(cluster cnr)

```

Using meta-analysis as an alternative to multi-level models

The multi-level model for cross-country comparisons has recently come under attack for the case of low N (such as 10 countries). Bryan & Jenkins (2016)² basically repeat the warnings about low-N situations that you cannot trust analytical statistics, based on normality assumptions. Remember the warning in your introductory statistics course that the z-test should be replaced by a t-test if $N < 30$ (or even $N < 100$), and calculate exact statistics for even smaller N. Bryan & Jenkins (2016) recommend to give up on the complicated machinery of multilevel programming for low N cases and replace it by meta-analysis, consisting of the following steps:

- Obtain level-1 coefficients in a individual level analysis, which could be obtained by splitting the file, or some group() specification in SEM.
- Write the estimates, together with their associated SE, to a level-2 file.
- Analyze the level-2 data using $1/(se*se)$ as a weight. [You can also use the context N as analytical weight.]

Meta-analysis is commonly used for combining results from independent research project, e.g. a series of randomized trials in medical testing. Stata has specific ados file for it: **metan** and **metareg**, but it also works without.

cntry	bbb	se	hg	www1	www2	n0	n1
BE	0.317	0.037	3	730.46	27.03	763	651
CH	0.429	0.029	8	1189.06	34.48	887	792
CZ	0.335	0.037	7	730.46	27.03	849	728
HU	0.496	0.035	10	816.33	28.57	776	693
IT	0.340	0.039	6	657.46	25.64	730	651

² Bryan, M. L., & Jenkins, S. P. (2016). Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review*, 32(1), 3–22. <http://doi.org/10.1093/esr/jcv059>. The point was earlier made by: Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86–92. <http://doi.org/10.1027/1614-2241.1.3.86>

LV	0.180	0.042	2	566.89	23.81	706	578
NL	0.280	0.033	5	918.27	30.30	826	747
RU	0.322	0.036	4	771.61	27.78	1027	728
TR	0.362	0.042	9	566.89	23.81	1202	730
UA	0.279	0.031	1	1040.58	32.26	1378	1138

bbb: effect of fisei → isei1

se: associated SE

hg: mysterious variable, predicting social reproduction

n0 Initial N per country (pairwise)

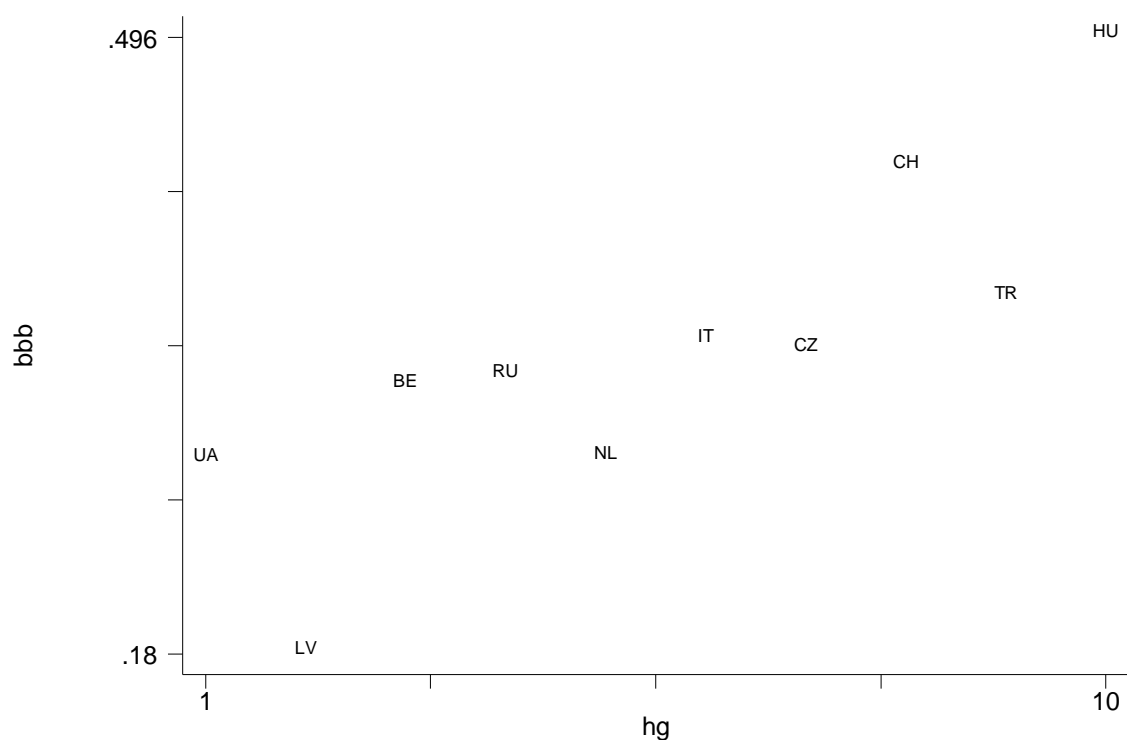
n1 Final N per country (listwise)

www1 $1/(se*se)$

www2 $1/se$

We can see the very strong relationship between bbb and hg in the following graph:

gr7 bbb hg, s([cntry])



A meta-analytical model can be obtained by:

bootstrap, reps (2000): regr bbb hg

regr bbb hg [w=www1]

(This recovers more or less the coefficients of the multilevel analysis, with similar SE / t-values.)

Meta-analysis can be applied on all sorts of coefficients, including effects between latent variables. Additional features of the procedure are:

- It is quick and easy to understand/
- It can accommodate a number of complicated models at level-2 data that otherwise would be beyond reach (such as XT panel models).
- It can accommodate simulation based sampling distributions, using **bootstrap** of **jackknife**.
- It makes you feel modest, because you suddenly see that you have no more than 10 cases to work on, and you would need an effect as strong as $r > 0.65$ to see some significant trend in your data. Nobody told you this in multilevel class.

Table 3.1: Structural parameters (latent correlation between father's and first occupation), estimated with various constraints of the measurement model.

	A	B	C	D	E	F	G
Constraints	none	mcoef	covar	mcoef ncons merrvar	FOCC == OCC1	osei == isei	errors Invariant
DF	??	54	63	99	101	102	104
L2	??	87.1	95.1	428.1	465	466	473.20
CH	0.564	0.539	0.541	0.562	0.568	0.569	0.567
HU	0.556	0.522	0.522	0.552	0.562	0.561	0.560
TR	0.480	0.484	0.497	0.456	0.455	0.458	0.455
CZ	0.446	0.438	0.441	0.429	0.441	0.443	0.442
BE	0.404	0.387	0.386	0.398	0.400	0.403	0.402
IT	0.402	0.394	0.395	0.403	0.411	0.412	0.411
RU	0.396	0.389	0.390	0.376	0.395	0.395	0.395
NL	0.388	0.402	0.399	0.373	0.380	0.379	0.377
UA	0.363	0.373	0.374	0.369	0.360	0.361	0.358
LV	0.246	0.237	0.233	0.244	0.241	0.243	0.242

Covar(crude) Var var 0.028 0.032 0.028 0.026 0.025

A: Configural invariance

B: Metric invariance

C: Metric invariance + invariant method effect

D: Scalar invariance

E: Scalar invariance + invariance of measurement model for the two occupations

E: Scalar invariance + invariance of measurement model for the two occupations +
invariance for the two indicators