Identification Problems in Model Construction: Its Epistemological Scope

Virtual IOMW 2020

Ernesto San Martín

Interdisciplinary Laboratory of Social Statistics

February 2021 (Joint work with D. Avello)

Laboratorio Interdisciplinario de Estadística Social

Structural Modelling

Laboratorio Interdisciplinario de Estadística Social

- Structural models connect theories and facts/data (Frisch, 1934, 1948; Haavelmo, 1944).
- Structural modelling was/is developed under the motto no measurement without theory (Koopmans, 1947).

Structural Modelling

Laboratorio Interdisciplinario de Estadística Social

- Koopmans (1949):

  Statistical inference unsupported by economic theory applies to whatever statistical regularities and stable relationships can be discerned in the data. Such purely empirical relationships when discernible are likely to be due to the presence and persistence of the underlying structural relationships, and (if so) could be deduced from a knowledge of the latter. However, the direction of this deduction cannot be reversed -from the empirical to the structural relationships– except possibly with the help of a theory which specifies the form of the structural relationships, the variables which enter into each, and any further details supported by prior observation or deduction therefrom. The more detailed these specifications are made in the model, the greater scope is thereby given to statistical inference from the data to the structural equations. We propose to study the limits to which statistical inference, from the data to the structural equations (other than definitions), is subject, and the manner in which these limits depend on the support received from economic theory.

Classical Test Theory

Laboratorio Interdisciplinario de Estadística Social

- Standard presentation:

$$Y = \tau + \epsilon, \qquad \tau \perp \epsilon, \qquad E(\epsilon) = 0.$$

Classical Test Theory

Laboratorio Interdisciplinario de Estadística Social

- Standard presentation:

$$Y = \tau + \epsilon, \qquad \tau \perp \epsilon, \qquad E(\epsilon) = 0.$$

- How τ and ϵ are interpreted?

Classical Test Theory

- Standard presentation:

$$Y = \tau + \epsilon, \qquad \tau \perp \epsilon, \qquad E(\epsilon) = 0.$$

- How τ and ϵ are interpreted?

- This question can be answered once the parameter of interest is made explicit ...

Classical Test Theory

Laboratorio Interdisciplinario de Estadística Social

- Standard presentation:

$$Y = \tau + \epsilon, \qquad \tau \perp \epsilon, \qquad E(\epsilon) = 0.$$

- How τ and ϵ are interpreted?

- This question can be answered once the parameter of interest is made explicit ... which leads to present correctly the CTT according to Spearman.

Classical Test Theory

Laboratorio Interdisciplinario de Estadística Social

- Standard presentation:

$$Y = \tau + \epsilon, \qquad \tau \perp \epsilon, \qquad E(\epsilon) = 0.$$

- How τ and ϵ are interpreted?
- This question can be answered once the parameter of interest is made explicit ... which leads to present correctly the CTT according to Spearman.
- This is done taking into account a geometrical perspective –which is the opposite of the arithmetical perspective: other perspective of measurement.

Fundamental problem

- Why two observations Y_1 and Y_2 are correlated?

Fundamental problem

Laboratorio Interdisciplinario de Estadística Social

- Why two observations Y_1 and Y_2 are correlated?
- It is always (in L^2) true that

$$
\begin{aligned}
cov(Y_1, Y_2) &= cov[E(Y_1 \mid \theta), E(Y_2 \mid \theta)] + cov[Y_1 - E(Y_1 \mid \theta), Y_2 - E(Y_2 \mid \theta)] \\
&= cov[E(Y_1 \mid \theta), E(Y_2 \mid \theta)]
\end{aligned}
$$

if the weak version of the Axiom of Local Independence (WALI) is assumed:

$$Y_1 \perp Y_2 \mid \theta.$$

Fundamental problem

Laboratorio Interdisciplinario de Estadística Social

- WALI is equivalent (in L^2) to the zero partial correlation between Y_1 and Y_2 controlled by θ, that is,

$$
\begin{aligned}
cor[Y_1 - E(Y_1 \mid \theta), Y_2 - E(Y_2 \mid \theta)] &= \frac{cor(Y_1, Y_2) - cor(Y_1, \theta)\, cor(Y_2, \theta)}{\sqrt{1 - cor^2(Y_1, \theta)}\, \sqrt{1 - cor^2(Y_2, \theta)}} \\
&= cor(Y_1, Y_2 \cdot \theta) \\
&= 0
\end{aligned}
$$

provided that the conditional expectations are affine linear functions.

Spearman's identification argument

- Let Y_1, Y_2, Y_3 be three observations that satisfy the WALI:

$$\begin{array}{c}
Y_1 \perp Y_2 \mid \theta \iff cor(Y_1, Y_2 \cdot \theta) = 0 \iff cor(Y_1, Y_2) = cor(Y_1, \theta)\, cor(Y_2, \theta) \\
Y_1 \perp Y_3 \mid \theta \iff cor(Y_1, Y_3 \cdot \theta) = 0 \iff cor(Y_1, Y_3) = cor(Y_1, \theta)\, cor(Y_3, \theta) \\
Y_2 \perp Y_3 \mid \theta \iff cor(Y_2, Y_3 \cdot \theta) = 0 \iff cor(Y_2, Y_3) = cor(Y_2, \theta)\, cor(Y_3, \theta).
\end{array}$$

- It follow that

$$\begin{aligned}
cor(Y_1, \theta)^2 &= \frac{cor(Y_1, Y_2)}{cor(Y_2, \theta)} \frac{cor(Y_1, Y_3)}{cor(Y_3, \theta)} \\
&= \frac{cor(Y_1, Y_2) cor(Y_1, Y_3)}{cor(Y_2, Y_3)}.
\end{aligned}$$

- This is the tetrachoric relation.

Discussion

- In this example, statistical regularities are represented through the correlations of the observables.
- Theoretically, it is assumed that there exists a third variable that produces such a correlation (WALI).
- In which sense? Under WALI it is possible to identify

$$cor^2(Y_1, \theta).$$

- Note that the sign of this correlation is identified provided some

$+cor(Y_1, Y_2)$	$+cor(Y_1, Y_3)$	$+cor(Y_3, Y_2)$
$-cor(Y_1, Y_2)$	$-cor(Y_1, Y_3)$	$+cor(Y_3, Y_2)$
$+cor(Y_1, Y_2)$	$-cor(Y_1, Y_3)$	$-cor(Y_3, Y_2)$
$-cor(Y_1, Y_2)$	$+cor(Y_1, Y_3)$	$-cor(Y_3, Y_2)$

Acknowledgements

Partially funded by the FONDECYT Grant 1181261