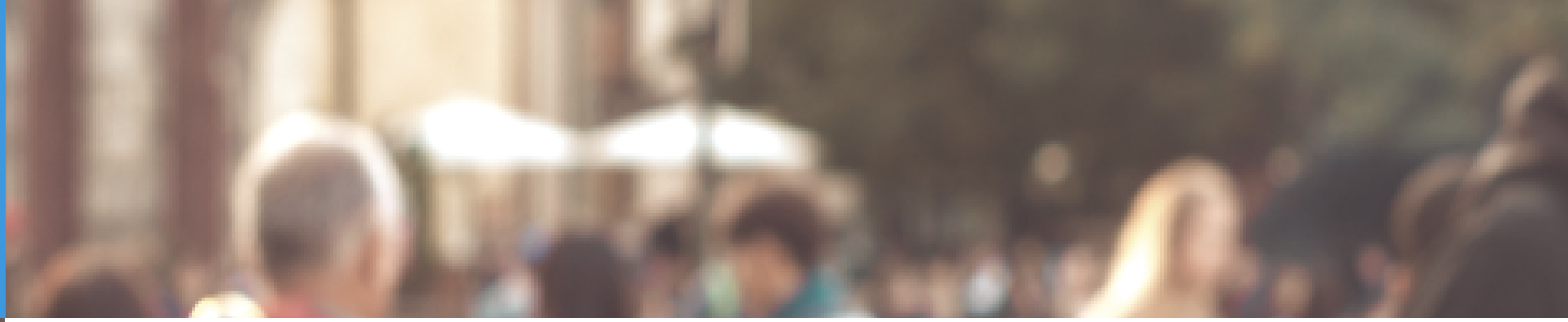


Elements of Structural Modelling

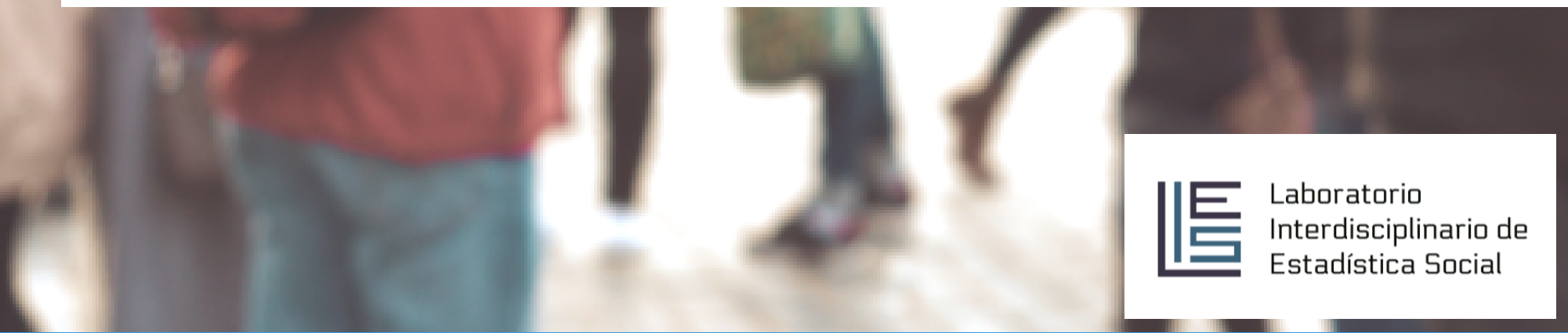
Conditional Independence

Cursillo de Verano 2020, Facultad de Matemáticas, UC

Ernesto San Martín

Laboratorio Interdisciplinario de Estadística Social LIES, UC
The Economics School of Louvain, Université Catholique de Louvain, Belgium

January 2021



Definition

- Let $(M, \mathcal{M}, P)$ be a probability space representing an experimental or observational situation. Let X_i $(i = 1, 2, 3, \dots)$ be random variables defined on $(M, \mathcal{M}, P)$.
- Notational convention: $Y \sqsubset X$ if and only if there exists a measurable function f such that
$$Y = f(X).$$
- It is said that X_1 *and* X_2 *are independent conditionally on* X_3, and it is write as $X_1 \perp\!\!\!\perp X_2 \mid X_3$, if and only if one of the following equivalent conditions is satisfied:
  1. $E(f_1 f_2 \mid X_3) = E(f_1 \mid X_3)\, E(f_2 \mid X_3)$ a.s. for all $f_i \sqsubset X_i$ positive (for $i = 1, 2$).
  2. $E(f_1 \mid X_2, X_3) = E(f_1 \mid X_3)$ a.s for all $f_1 \sqsubset X_1$ positive.
  3. $E\{E(f_{13} \mid X_3) \mid X_2\} = E(f_{13} \mid X_2)$ a.s. for all $f_{13} \sqsubset (X_1, X_3)$ positive.

Counter-example

- When it is said that $Y \perp\!\!\!\perp X \mid Z$, then it is necessary to verify that $E(f \mid X, Z) = E(f \mid Z)$ *for all* $f \sqsubset Y$, and not for a specific Y only.

- To illustrate this aspect, let us consider three discrete random variables X, Y and Z such that X and Z are defined on $\{a, b\}$ and Y is defined on $\{-2, -1, 0, 1, 2\}$. The joint distribution of (X, Y, Z) is given by

Y	$(X,Z) = (a,a)$	$(X,Z) = (a,b)$	$(X,Z) = (b,a)$	$(X,Z) = (b,b)$
-2	0	s_2	0	l_2
-1	r_1	s_1	t_1	l_1
0	r_0	s_0	t_0	l_0
1	r_1	s_1	t_1	l_1
2	0	s_2	0	l_2

such that $s_i > 0$ and $l_i > 0$ for $i = 0, 1, 2$; $r_i > 0$ and $t_i > 0$ for $i = 0, 1$; and

$$2(l_1 + l_2 + r_1 + s_1 + s_2 + t_1) + (l_0 + r_0 + s_0 + t_0) = 1.$$

Counter-example

- It can be verified that $E[Y \mid (X, Z) = (x, z)] = 0$ for all $(x, z) \in \{a, b\} \times \{a, b\}$, then $E(Y \mid X, Z) = 0$.
- The conditional distribution of Y given Z is given by

y	$P(Y = y \mid Z = a)$	$P(Y = y \mid Z = b)$
-2	0	$\omega_b^{-1}(l_2 + s_2)$
-1	$\omega_a^{-1}(r_1 + t_1)$	$\omega_b^{-1}(l_1 + s_1)$
0	$\omega_a^{-1}(r_0 + t_0)$	$\omega_b^{-1}(l_0 + s_0)$
1	$\omega_a^{-1}(r_1 + t_1)$	$\omega_b^{-1}(l_1 + s_1)$
2	0	$\omega_b^{-1}(l_2 + s_2)$

  where $\omega_a = P(Z = a)$ and $\omega_b = P(Z = b)$.
- It follows that $E(Y \mid Z = a) = E(Y \mid Z = b) = 0$ and, therefore, $E(Y \mid Z) = 0$. Consequently,

$$E(Y \mid X, Z) = E(Y \mid Z).$$

Counter-example

- Nevertheless, $Y \not\perp\!\!\!\perp X \mid Z$: the conditional probability $P(Y = y, X = x \mid Z = a)$ is given by

	$Y=-2$	$Y=-1$	$Y=0$	$Y=1$	$Y=2$
$X=a$	0	$\omega_a^{-1} r_1$	$\omega_a^{-1} r_0$	$\omega_a^{-1} r_1$	0
$X=b$	0	$\omega_a^{-1} t_1$	$\omega_a^{-1} t_0$	$\omega_a^{-1} t_1$	0

,

whereas the conditional probability of $P(Y = y, X = x \mid Z = b)$ is given by

	$Y=-2$	$Y=-1$	$Y=0$	$Y=1$	$Y=2$
$X=a$	$\omega_b s_2$	$\omega_b^{-1} s_1$	$\omega_b^{-1} s_0$	$\omega_b^{-1} s_1$	$\omega_b s_2$
$X=b$	$\omega_b l_2$	$\omega_b^{-1} l_1$	$\omega_b^{-1} l_0$	$\omega_b^{-1} l_1$	$\omega_b l_2$

.

- It can be verified that $Y \perp\!\!\!\perp X \mid \{Z = a\}$ if and only if the rank of the matrix corresponding to the joint probability distribution given $\{Z = a\}$ is equal to 1. Therefore, if, for instance,

$$\frac{r_1}{r_0} \neq \frac{t_1}{t_0},$$

then $Y \not\perp\!\!\!\perp X \mid \{Z = a\}$ and, consequently, $Y \not\perp\!\!\!\perp X \mid Z$.

Consequences of the definition

- If $X_3 \sqsubset X_1$, then $X_1 \perp\!\!\!\perp X_2 \mid X_3$ if and only if one of the following two equivalent properties is satisfied:

  1. $E(f_2 \mid X_1) = E(f_2 \mid X_3)$ a.s. for all $f_2 \sqsubset X_2$ positive.
  2. $E[E(f_1 \mid X_3) \mid X_2] = E(f_1 \mid X_2)$ a.s. for all $f_1 \sqsubset X_1$ positive.

Consequences of the definition

- If $X_3 \sqsubset X_1$, then $X_1 \perp\!\!\!\perp X_2 \mid X_3$ if and only if one of the following two equivalent properties is satisfied:
  1. $E(f_2 \mid X_1) = E(f_2 \mid X_3)$ a.s. for all $f_2 \sqsubset X_2$ positive.
  2. $E[E(f_1 \mid X_3) \mid X_2] = E(f_1 \mid X_2)$ a.s. for all $f_1 \sqsubset X_1$ positive.
- If $X_1 \sqsubset X_3$, then $X_1 \perp\!\!\!\perp X_2 \mid X_3$ for all random variable X_2.

Consequences of the definition

- If $X_3 \sqsubset X_1$, then $X_1 \perp\!\!\!\perp X_2 \mid X_3$ if and only if one of the following two equivalent properties is satisfied:
  1. $E(f_2 \mid X_1) = E(f_2 \mid X_3)$ a.s. for all $f_2 \sqsubset X_2$ positive.
  2. $E[E(f_1 \mid X_3) \mid X_2] = E(f_1 \mid X_2)$ a.s. for all $f_1 \sqsubset X_1$ positive.
- If $X_1 \sqsubset X_3$, then $X_1 \perp\!\!\!\perp X_2 \mid X_3$ for all random variable X_2.
- If $X_1 \perp\!\!\!\perp X_2 \mid X_3$ and $X_5 \sqsubset X_1$, then $X_5 \perp\!\!\!\perp X_2 \mid X_3$.

Null events

- Let $(M, \mathcal{M}, P)$ be a probability space representing an experimental or observational situation.
- The completed trivial σ-algebra $\overline{\mathcal{M}_0}$ is given by

$$\overline{\mathcal{M}_0} = \{A \in \mathcal{M} \,:\, P(A) = P^2(A)\}.$$

- So

$$A \in \overline{\mathcal{M}_0} \iff Var(\mathbb{1}_A) = 0 \iff \mathbb{1}_A = E(\mathbb{1}_A)\ P\text{-a.s.}$$

and consequently

$$X \sqsubset \overline{\mathcal{M}_0} \iff Var(X) = 0 \iff X = E(X)\ P\text{-a.s.}$$

Null events

- $Var(X \mid Y) = 0 \iff X = E(X \mid Y)$ a.s. $\iff X \sqsubset \overline{\mathcal{Y}}$; that is,
  $Var(X \mid Y) = 0$ if and only if the information provided by X is, with probability one, already contained in the information provided by Y.

- $X_1 \perp\!\!\!\perp X_1 \mid X_3$ if and only if $V(X_1 \mid X_3) = 0$ if and only if X_1 is a.s. a function of X_3

Fundamental property

- The following properties are equivalent:
  1. $X_1 \perp\!\!\!\perp X_2 \mid X_3$ and $X_1 \perp\!\!\!\perp X_4 \mid X_2, X_3$.
  2. $X_1 \perp\!\!\!\perp (X_2, X_4) \mid X_3$
  3. $X_1 \perp\!\!\!\perp X_4 \mid X_3$ and $X_1 \perp\!\!\!\perp X_2 \mid X_4, X_3$.

Fundamental property

- The following properties are equivalent:
  1. $X_1 \perp\!\!\!\perp X_2 \mid X_3$ and $X_1 \perp\!\!\!\perp X_4 \mid X_2, X_3$.
  2. $X_1 \perp\!\!\!\perp (X_2, X_4) \mid X_3$
  3. $X_1 \perp\!\!\!\perp X_4 \mid X_3$ and $X_1 \perp\!\!\!\perp X_2 \mid X_4, X_3$.
- If $X_1 \perp\!\!\!\perp X_2 \mid X_3$, $X_5 \sqsubset (X_1, X_3)$ and $X_4 \sqsubset (X_2, X_3)$, then
  1. $(X_1, X_5) \perp\!\!\!\perp (X_2, X_4) \mid X_3$.
  2. $X_1 \perp\!\!\!\perp X_2 \mid (X_3, X_4, X_5)$.

Example 1

- Consider the problem of operating a "fair" procedure for the selection of minority group members for university admission. One solution is to require that the probability of such selection should depend only on the academic promise of the candidates, and not on race, sex and so on.
- Let X_1 denote selection $(X_1 = 1)$ or rejection $(X_1 = 0)$, let X_2 denote sex, and let X_3 be a test-score regarded as a good assessment of academic promise.
- It is intended that $X_1 \perp\!\!\!\perp X_2 \mid X_3$, that is, the process generating selection conditionally on sex and the test-score only depends on the test-score.
- By the symmetry of the conditional independence relationship, the above-mentioned assumption is equivalent to $X_2 \perp\!\!\!\perp X_1 \mid X_3$, which means that the process generating the sex conditionally on the selectivity and the test-score only depends on the test-score: that is to say by taking two test-groups, one of successful and another of unsuccessful candidates, and looking to see that the proportions of those getting any particular Z-score who are males are the same in both groups.

Example 2

- Let Y be a function of X (that is, $Y \sqsubset X$) and suppose that the generated outcome X is selected for observation is determined by Y alone. We can express this in terms of conditional independence by defining an indicator variable Q, with $Q = 1$ if the outcome is selected, and $Q = 0$ otherwise.

- Then we are assuming that $Q \perp\!\!\!\perp X \mid Y$. This relation is equivalent to $X \perp\!\!\!\perp Q \mid Y$, which means that the distribution of X given Y is the same for both selected and unselected data.

Example 2

- An application of the previous considerations arises in a statistical approach to medical diagnosis.
- With each individual in a population is associated a pair (S, D), where D signifies his/her disease, and S the full set of symptoms, signs, etc. on which the diagnosis is to be based.
- The selection variable Q represents admission to the center in which the data are collected. This selection is supposed to be governed entirely by the presenting symptoms S_0, a subset of S, together with further non-medical personal information G. Thus,

$$Q \perp\!\!\!\perp (S, D) \mid (S_0, G).$$

The non-medical nature of G may be formalized as

$$G \perp\!\!\!\perp D \mid S,$$

expressing the fact that, for inference about D from (S, G), S is sufficient.

Example

- $Q \perp\!\!\!\perp (S, D) \mid (S_0, G)$ implies that $Q \perp\!\!\!\perp D \mid (S, G)$.
- Using the Fundamental Property, this later condition, along with $G \perp\!\!\!\perp D \mid S$ are equivalent to

$$(Q, G) \perp\!\!\!\perp D \mid S$$

  so that $D \perp\!\!\!\perp Q \mid S$.
- It follows that the distribution of D given S are unaffected by selection.

Example 3

- Example based on Bouckaert and Mouchart (2001), Sure outcomes of random events: a model for clinical trials.

- Motivation:

  We present a simple model of the major events occurring during a randomized clinical trial, the main features of which are the following: (i) the trial is considered after completion; (ii) only two groups of patients are considered: a control group where a non-active drug (also called placebo) is administered and an active group where the drug under trial is administered at one single dose; (iii) two types of outcome are distinguished: outcomes of type I are designated as therapeutic (or main) effects and outcomes of type II are designated as side (or adverse) effects; (iv) for each type of (observable) outcome the model distinguishes two types of (nonobservable) causes: drug-specific (or 'explained') cause and non-specific (or 'residual') cause; (v) all events, whether observable or not, are coded by binary variables, no allowance is made for the 'intensity' of the outcomes or of the causes; (vi) one innovative feature of this model is that the outcomes and the causes are connected by a so-called SORE model (sure outcomes of random events).

Example 3

- Motivation:

In a deterministic model, an 'outcome' (principal effect such as health improvement or side-effect) is, in principle, attributable to one single cause. Consider however a patient with allergic rhinitis included in a double blind drug versus placebo controlled trial, with a new drug expected to bring relief less than 5 hours after intake. Obviously, if the patient feels better after three hours, either the drug was active or it was luck (a word preferred to the pedantic 'placebo effect').

Example 3

- Let us consider three random variables:

$$Z = \begin{cases} 0, & \text{no relief;} \\ 1, & \text{relief.} \end{cases}$$

$$\lambda = \begin{cases} 0, & \text{no drug-induced relief;} \\ 1, & \text{drug-induced relief.} \end{cases}$$

$$\theta = \begin{cases} 0, & \text{no relief by placebo effect;} \\ 1, & \text{relief by placebo effect.} \end{cases}$$

- The situation sketched previously may be modelled by the following relation between the outcome Z and the latent variables θ and λ:

$$Z = \lambda + \theta - \lambda\theta.$$

This is the *sure outcome random events* model or SORE model.

Example 3

- We now consider the problem of modelling the main therapeutic outcome (health improvement or not) and the side (or adverse) outcome observed during a clinical trial, when the possibility of placebo and no-placebo effects is explicitly taken into account.

- We assume that the outcomes are observed without errors, and that treatment allocations have been unblinded for the analysis.

Example 3

- We now consider the problem of modelling the main therapeutic outcome (health improvement or not) and the side (or adverse) outcome observed during a clinical trial, when the possibility of placebo and no-placebo effects is explicitly taken into account.
- We assume that the outcomes are observed without errors, and that treatment allocations have been unblinded for the analysis.
- Observable variables:

$$X = \begin{cases} 0, & \text{placebo group;} \\ 1, & \text{drug-treated group.} \end{cases}$$

$$Y = \begin{cases} 0, & \text{no adverse outcome;} \\ 1, & \text{adverse outcome.} \end{cases}$$

$$Z = \begin{cases} 0, & \text{no health improvement;} \\ 1, & \text{health improvement.} \end{cases}$$

Example 3

- Non-observable events:

$$\lambda = \begin{cases} 0, & \text{no therapeutic pharmacological effect;} \\ 1, & \text{therapeutic pharmacological effect.} \end{cases}$$

$$\nu = \begin{cases} 0, & \text{no toxic pharmacological effect;} \\ 1, & \text{toxic pharmacological effect.} \end{cases}$$

$$\theta = \begin{cases} 0, & \text{no placebo therapeutical effect;} \\ 1, & \text{placebo therapeutical effect.} \end{cases}$$

$$\mu = \begin{cases} 0, & \text{no placebo toxic effect;} \\ 1, & \text{placebo toxic effect.} \end{cases}$$

Example 3

- The model is characterized by the joint distribution of $(X, Y, Z, \lambda, \nu, \theta, \mu)$, which have values in a space with $2^7 = 128$ points.
- The latent variables $(\lambda, \nu, \theta, \mu)$ are included in the model because they correspond to well-known concepts in pharmacology and because they determine observable events in a SORE model:

$$Z = \theta + \lambda - \theta\lambda, \qquad Y = \mu + \nu - \mu\nu.$$

- The logic of the model implies two constrains, namely that drug-related effects are not possible unless drug is administered, that is:

$$\{\lambda = 1\} \Longrightarrow \{X = 1\}, \qquad \{\nu = 1\} \Longrightarrow \{X = 1\}.$$

Or equivalently

$$\lambda = \lambda X, \qquad \nu = \nu X.$$

Example 3

- Next we introduce a set of conditions of stochastic independence; these formalize some structural properties of the problem.

- First, we suppose that allocation of patients to groups is random. More specifically, we first assume that, for the latent causes of the main outcome, placebo effects are independent of group assignment:

$$\theta \perp\!\!\!\perp X.$$

Example 3

- Next we introduce a set of conditions of stochastic independence; these formalize some structural properties of the problem.

- First, we suppose that allocation of patients to groups is random. More specifically, we first assume that, for the latent causes of the main outcome, placebo effects are independent of group assignment:

$$\theta \perp\!\!\!\perp X.$$

- We also assume that in each group of patients, pharmacological and placebo effects are independent, that is

$$\lambda \perp\!\!\!\perp \theta \mid X.$$

Example 3

- Next we introduce a set of conditions of stochastic independence; these formalize some structural properties of the problem.
- First, we suppose that allocation of patients to groups is random. More specifically, we first assume that, for the latent causes of the main outcome, placebo effects are independent of group assignment:

$$\theta \perp\!\!\!\perp X.$$

- We also assume that in each group of patients, pharmacological and placebo effects are independent, that is

$$\lambda \perp\!\!\!\perp \theta \mid X.$$

- Note that the previous two conditions are equivalent to

$$\theta \perp\!\!\!\perp (\lambda, X),$$

which in turn is equivalent to

$$\theta \perp\!\!\!\perp \lambda, \qquad \theta \perp\!\!\!\perp X \mid \lambda.$$

Example 3

- Next, we assume that in each group the drug toxic reactions occur independently of placebo therapeutic or toxic effects and independently of the therapeutic effect of the drug; in other words, the probability of a drug toxic reaction depends only of group allocation and given that one, it is the same whether there is a nocebo reaction or not:

$$\nu \perp\!\!\!\perp (\mu, \theta, \lambda, Z) \mid X.$$

Example 3

- We also assume that the placebo toxic effect is independent of the group allocation, of the therapeutic pharmacological effect and of the therapeutic placebo effect, conditionally to the main therapeutic outcome:

$$\mu \perp\!\!\!\perp (\theta, \lambda, X) \mid Z.$$

This condition is justified s follows:

In this study we aim at developing and formalizing ideas first expressed by Bernheim and Vrana [5]. In that paper, the authors suppose that the occurrence of side-effects has at least three causes: (i) the toxic pharmacological property of the drug; (ii) the toxic placebo (or nocebo) effect; (iii) the pathology of the patient. The first two causes are explicitly dealt with by our model [...] the third cause is dealt with by introducing a statistical dependence between the nocebo effect and the principal outcome. This principal outcome accordingly becomes at the same time an outcome and a modifying factor for the side-effects. There are indeed numerous instances in pathology where side-effects are similar to some unalleviated symptoms of the disease for which the treatment has been initiated or are the excess of some aspect of the therapeutic action, like somnolence in the treatment of insomnia.

The independence between μ and (θ, λ, X) reflects the fact that proneness to side-effects from nocebo depends only on actual health status and not on the mechanism by which this health status was reached.

Example 3

- The impact of the structural specification is the following:

$$
\begin{aligned}
P(X,\theta,\lambda,Z,\mu,\nu,Y) &= P(X)P(\theta \mid X)P(\lambda \mid \theta, X)P(Z \mid \lambda,\theta,X)P(\mu \mid Z,\lambda,\theta,X) \\
&\quad P(\nu \mid \mu, Z,\lambda,\theta,X)P(Y \mid \nu,\mu,Z,\lambda,\theta,X) \\
&= P(X)P(\theta)P(\lambda \mid X)P(Z \mid \lambda,\theta,X)P(\mu \mid Z)P(\nu \mid X) \\
&\quad P(Y \mid \nu,\mu,Z,\lambda,\theta,X)
\end{aligned}
$$