G. Rohwer
Selection, Choice and Causal Interpretations

There are 10 chapters:

1. Descriptions and generalizations
1.1 Descriptive statistical statements

Definition:

Descriptive statistical statements are statements about the frequency distribution of properties (or quantities derived thereof) in a specified set of units.

Formal framework: statistical variables. Symbolic notation:

$$X : \Omega \longrightarrow \mathcal{X}$$

X is the name of the variable, Ω is the reference set, a finite set of actually observed or assumed cases, and $\mathcal{X}$ is the property space (domain).

If the reference set consists of not actually observed cases, it is nevertheless required, for descriptive statements, that one can reasonably assume that the cases do exist, or have existed in the past.

Simply stated: one cannot make descriptive statements about the future.

1.2 Defining descriptive generalization

Starting point: $X : \Omega \longrightarrow \mathcal{X}$, representing the observations.

Ω, the set of observed cases, is then considered as subset of another set, Ω^*, for which one can assume an analogously defined statistical variable:

$$X^* : \Omega^* \longrightarrow \mathcal{X}$$

having the same property space as X.

This framework allows one to define: A *descriptive generalization* consists in using the observed values of X for making descriptive statements about the distribution of X^* in Ω^*.

It is noteworthy that the desired generalization has the same linguistic form as the statistical statements derived from the observations; there only is a change in the reference set.

1.3 Data-generating and fact-generating processes

I will not discuss here problems of statistical inference.

It is obvious, however, that the justification of a descriptive generalization must be based on the data generating process that has generated the observations.

Note that I use the following distinction:

– The term ‘data generating process’ is used to refer to a process that generates data, that is, information about already existing facts.
– In contrast, when referring to processes that generate new facts (outcomes), I use the term ‘fact-generating process’.

1.3 Data-generating and fact-generating processes (cont’d)

As an example think of a learning frame in which students can acquire capabilities of a specified kind, and assume that individual learning results can be captured by values of a variable, say Y.

One can firstly think of a fact-generating process in which each student eventually acquires a particular capability.

Afterwards, a data-generating process can take place, that is, a process in which a researcher represents students’ capabilities by particular values of Y.

1.4 Limitations of descriptive generalization

Descriptive generalization intends to enlarge the knowledge about statistical facts, meaning here statistical distributions as they are actually realized in specified populations.

This very interest requires a narrow understanding of ‘population’.

Limitations become obvious when the justification of descriptive generalization is based on probability sampling.

This requires that Ω can be viewed as a probability sample from Ω^*.

Consequently, Ω^* can only consist of units having a positive selection probability *when and where* the sample is drawn.

1.4 Limitations of descriptive generalization (cont’d)

Particular difficulties arise when the interest concerns historical processes.

The basic question then is, How to define a population of processes?

From a methodological point of view, such populations are best defined as cohorts.

Being interested in descriptive generalizations, this requires to adopt a historical perspective that is confined to mostly completed processes.

1.5 Modal generalization with rules

I now consider a different kind of generalization where the goal is, not a descriptive statement about a set of units, but a predictive rule. (For more about this distinction, see Rohwer 2014.)

I use the term ‘rule’ in a general sense for statements having the form

If ..., then ...

Different kinds of rules can be distinguished w.r.t. the *modalities* used in formulating the *then*-part; for example: If ..., then ... is possible, or probable, or necessary, or normatively required.

Empirical research is primarily interested in *predictive rules*.

Example: Let ω denote an individual who has finished school in Germany: If at least one of ω’s parents has finished school with an Abitur, then it is highly probable that also ω has an Abitur.

Note that this is a *generic* rule, meaning that its object is specified only by values of variables.

1.5 Modal generalization with rules (cont'd)

Important distinction:

- A *static predictive rule* formulates a relationship between properties of a unit.

  The general form is: If ω has property x, then ω (probably) has property y.

  This kind of predictive rule is exemplified by the above example.
- A *dynamic predictive rule* relates to a fact-generating process that generates an outcome that is to be predicted.

  Example: If ω (a generically specified individual) regularly participates in the instructions, she will (probably) be successful in the final exam.

1.6 Formulating predictive rules with variables

In the following, I only consider predictive rules which include a probabilistic qualification of the prediction.

When formulating such rules with variables, a first question concerns how to understand the probabilistic qualification.

There are two forms:

a) Qualitative: If $X{=}x$, then $Y{=}y$ is probable (in some qualified sense).
b) Quantitative: If $X{=}x$, then $\Pr(Y{=}y) = \ldots$ [a specific, actually given or assumed, numerical value].

Empirical research with statistical methods regularly uses quantitative formulations.

1.6 Formulating predictive rules with variables (cont'd)

The presupposition of quantifiable probabilities allows one to use mathematical functions for formulating the relationship between the *if*- and the *then*-part of the rule.

As a general form one can use

$$x \longrightarrow \Pr[Y \mid X{=}x]$$

to be read as a function that assigns to each value x in the domain of X a conditional probability distribution of Y.

Note: I use square brackets when referring to the distribution of a variable (to the left of the conditioning bar).

1.6 Formulating predictive rules with variables (cont'd)

If Y is a discrete variable, one can also use specific functions having the form

$$x \longrightarrow \Pr(Y{=}y \mid X{=}x)$$

for each value y in the domain of Y.

Another often used special form is

$$x \longrightarrow \mathrm{E}(Y \mid X{=}x)$$

which formulates the relationship with conditional expectations of Y.

Starting from such general formulations, one can think of more specific parametric forms.

Whatever the finally chosen functional form, these forms must be distinguished from numerically specified functions which actually allow one to calculate values of the function.

1.7 Predictive rules vs. descriptive statements

Predictive rules must be distinguished from descriptive statistical statements.

- While descriptive statistical statements concern a reference set of particular units, a predictive rule concerns a generic unit which is only specified by values of variables.
- Correspondingly, there is a conceptual difference between frequencies,

  $$\mathrm{P}(Y{=}y \mid X{=}x)$$

  which presuppose a finite reference set, and probabilities,

  $$\Pr(Y{=}y \mid X{=}x)$$

  which concern a generic unit.

  I therefore use different symbols: P for frequencies, and Pr for probabilities.

1.7 Predictive rules vs. descriptive statements (cont'd)

A random generator can serve to illustrate the distinction. I use 'throwing a die' as an example.

The random generator can be defined by a rule, e.g.,

> *If the die is thrown, there are six possible outcomes, each can occur with the same probability (1/6).*

This rule is to be distinguished from a descriptive statement about frequencies of outcomes in an actually realized set of throws.

Assume the die is thrown 100 times. Results can be represented by a statistical variable $Z : \Omega \longrightarrow \mathcal{Z} := \{1, \ldots, 6\}$. $\mathrm{P}[Z]$, the distribution of Z, must be distinguished (numerically and conceptually) from the probability distribution which is used in the formulation of the rule describing the random generator.

1.7 Predictive rules vs. descriptive statements (cont'd)

Since predictive rules are different from descriptive statements (and different from analytical truths), they cannot be true or false.

They can only be pragmatically justified, that is, with arguments showing that, and how, a rule can help people in their activities.

Since there is a formal equivalence of frequency and probability functions, researchers often ignore the conceptual distinction and present their observed frequencies in terms of probabilities.

This should be avoided in order to remind of the distinction between descriptive statements and predictive rules.

1.8 Statistical and modal variables

The conceptual distinction between descriptive statements and predictive rules suggests to make a corresponding distinction between the kinds of variables involved.

As already explained, descriptive statistical statements are derived from statistical variables which are known, or assumed, to represent realized properties of existing units.

This is also true for conditional frequencies:

$\mathrm{P}(Y{=}y \mid X{=}x)$ is derived from a statistical variable, (X, Y), which is defined for a particular reference set.

1.8 Statistical and modal variables (cont'd)

When considering instead conditional probabilities,

$$\Pr(Y = y \mid X = x)$$

one must recognize that there are two different conceptual frameworks:

1) One can assume that the conditional probabilities are derived from a random variable (X, Y).

This understanding presupposes the existence of joint and marginal probability distributions of the two variables.

2) The situation is different when conditional probabilities serve to formulate probabilistic predictive rules.

1.8 Statistical and modal variables (cont'd)

In the second framework, X is used to formulate a hypothetical assumption, and so it is neither a random variable (having a probability distribution) nor a statistical variable (having a statistical distribution).

Consequently, also Y has no unconditional distribution, but can only be viewed as a random variable for specified values of X.

In order to remind of the second context, I speak of *modal variables* and use a special notation: $\ddot{X}$ instead of X, and $\dot{Y}$ instead of Y.

The symbolic notation for a probabilistic predictive rule then becomes

$$x \longrightarrow \Pr[\dot{Y} \mid \ddot{X} = x]$$

to be read as:

IF $\ddot{X}$ has the value x, THEN one can consider $\dot{Y}$ as a random variable with distribution $\Pr[\dot{Y} \mid \ddot{X} = x]$.

2. Descriptive models
2.1 Descriptive models based on statistical variables

Descriptive models are tools for describing distributions of statistical variables.

Starting point is a statistical variable, $X : \Omega \longrightarrow \mathcal{X}$, often consisting of several components.

A descriptive model aims to describe the distribution of X, denoted by $P[X]$, or aspects of this distribution, by using a simpler mathematical form.

Example: describing the distribution of students' 'ability scores' (observed in a sample or posited in a population) by a normal distribution.

2.2 Regression models with statistical variables

If X consists of two or more components, one is often interested in descriptions of conditional distributions.

This is done with regression functions and regression models.

The starting point is given by a two-dimensional statistical variable, say

$$(X, Y) : \Omega \longrightarrow \mathcal{X} \times \mathcal{Y}$$

A general *regression function* is a function

$$x \longrightarrow P[Y \mid X = x]$$

which assigns to each value $x \in \mathcal{X}$ the conditional frequency distribution of Y, as given by the statistical variable (data).

2.2 Regression models with statistical variables (cont'd)

In order to create a descriptive regression model, one uses a simpler mathematical representation of the conditional distribution, say

$$g(x; \theta) \approx P[Y \mid X = x]$$

where θ is a parameter vector.

A general *regression model* is then given by the function $x \longrightarrow g(x; \theta)$.

Special regression models are used to represent aspects of $P[Y \mid X = x]$.

Of widespread use is regression with mean values: $m(x; \theta) \approx M(Y \mid X = x)$.

Example: $M(Y \mid X = x) \approx \alpha + x\beta$

This model approximates the conditional mean value of Y by a linear function of the values of X that are used as conditions.

2.3 Descriptive models as tools

Descriptive models are primarily tools for comprehending aspects of complex data sets.

As suggested by R. A. Fisher (1922: 311), this is a primary task of statistical methods:

> *Briefly, and in its most concrete form, the object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data.*

2.3 Descriptive models as tools (cont'd)

Descriptive models are also useful tools for descriptive generalizations.

Since these models relate to statistical variables, there is no conceptual difference whether the reference set is a sample or a population.

Starting from a model intended to describe a population allows one to think in terms of estimating its parameters with the information from a sample.

Notice that the term 'estimation' has a clear meaning in this context: It means that one aims to find values of model parameters which are defined by their hypothetical calculation for the complete population.

This entails that already their definition depends not only on the specified model, but also on a particular method for calculating its parameters.

Example: Estimation by minimizing (a) squared or (b) absolute deviations. These methods are part of different model definitions.

3. Functional models
3.1 Functional models: introduction

Functional models (as I use this term) are tools for thinking about relationships between variables. (For a more comprehensive introduction to this kind of models, see Rohwer 2010.)

The basic formal tool are functions (mathematically understood) which connect variables.

So one can speak of 'functional relationships' between variables, and the models are also called 'functional models'.

Two kinds of functional relationships must be distinguished.

3.1 Functional models: introduction (cont'd)

Consider two variables, X with domain $\mathcal{X}$ and Y with domain $\mathcal{Y}$.

– A *deterministic* functional relationship consists of a function

$$x \longrightarrow y = f(x)$$

which assigns to each value $x \in \mathcal{X}$ exactly one value $f(x) \in \mathcal{Y}$.

– A *probabilistic* functional relationship consists of a function

$$x \longrightarrow \Pr[Y|X{=}x]$$

which assigns to each value $x \in \mathcal{X}$ a conditional probability distribution.

3.1 Functional models: introduction (cont'd)

Notice that $\Pr[Y|X{=}x]$ is itself a function.

If Y is a discrete variable, this function can be written as

$$y \longrightarrow \Pr(Y{=}y \,|\, X{=}x) \tag{1}$$

to be interpreted as the probability of $Y{=}y$ given that $X{=}x$.

Since (1) is formally identical with a probabilistic predictive rule, functional models can also be understood as tools for formulating probabilistic predictive rules.

3.2 A general notion of functional models

a) The structure of the model is given by a directed acyclic graph.
b) To each node of the graph corresponds a variable. Variables with indegree zero are called exogenous variables and marked by two dots. All other variables are called endogenous variables and marked by a single dot.
c) For each endogenous variable, there is a deterministic or probabilistic function showing how the variable (its values or probability distribution) depends on values of the immediately preceding variables.
d) Without further assumptions, exogenous variables do not have an associated distribution.

This is a formal framework. The arrows between variables have no specified meaning. In many applications they can be understood as indicating some kind of dependence relation.

3.3 Illustration with a simple example

The example concerns the educational outcome of a generic child.

Model 3.1

$$\begin{tikzcd} \ddot{X} \arrow[r] \arrow[d] & \dot{Y} \\ \dot{Z} \arrow[ur] & \end{tikzcd}$$

For simplicity, the variables are assumed to be binary and defined as follows:

$\dot{Y}$	child's educational outcome (1 successful, 0 otherwise)
$\ddot{X}$	parents' educational level (1 high, 0 low)
$\dot{Z}$	school type (1 or 2)

3.3 Illustration with a simple example (cont'd)

The model contains two probabilistic functions:

$$x \longrightarrow \Pr[\dot{Z} \,|\, \ddot{X}{=}x]$$
$$(x,z) \longrightarrow \Pr[\dot{Y} \,|\, \ddot{X}{=}x, \dot{Z}{=}z]$$

Since $\dot{Z}$ and $\dot{Y}$ are binary variables, it suffices to consider the functions

$$x \longrightarrow \Pr(\dot{Z}{=}1 \,|\, \ddot{X}{=}x)$$
$$(x,z) \longrightarrow \Pr(\dot{Y}{=}1 \,|\, \ddot{X}{=}x, \dot{Z}{=}z)$$

The first function is intended to show how the probability of attending a specified school type depends on the parents' educational level.

The second function is intended to show how the probability of educational success depends both on the parents' educational level and on the school type.

3.4 Assuming distributions for exogenous variables?

Exogenous variables of a functional model do not have an associated distribution.

There sometimes are reasons for assuming distributions for exogenous variables. I briefly mention three situations.

1. Using the model for a statistical explanation w.r.t. a particular reference set.

   Distributions of the model's exogenous variables are then identified with the distributions of the corresponding statistical variables.

3.4 Assuming distributions for exogenous variables? (cont'd)

2. Using the model for predicting the outcome for an individual that is (only) known to belong to a particular reference set.

   One then employs a reduced model that is derived from the original model by integrating over the distributions of the unobserved exogenous variables.

   Assume, for example, that one wants to use the model of the previous subsection for predicting a child's educational success. Not knowing the educational level of the parents, one cannot use the model. However, substituting $\ddot{X}$ by a variable $\dot{X}$ with a probability distribution $\Pr[\dot{X}|\ddot{Z}{=}z]$, one can derive a *reduced model*

   $$\Pr(\dot{Y}{=}1 \,|\, \ddot{Z}{=}z) = \sum\nolimits_x \Pr(\dot{Y}{=}1 \,|\, \ddot{Z}{=}z, \dot{X}{=}x)\, \Pr(\dot{X}{=}x|\ddot{Z}{=}z)$$

   which only requires knowledge of the child's school type.

3.4 Assuming distributions for exogenous variables? (cont'd)

3. Using the model for predicting the value of an exogenous variable based on knowing values of endogenous variables.

   In order to apply Bayesian inference, one must begin with a prior distribution for the exogenous variables.

   For example, knowing that a child had successfully completed the school, make a guess about her school type.

3.5 Defining effects of explanatory variables

Assume that $\dot{Y}$ depends on an exogenous variable $\ddot{X}$.

To think of an effect of $\ddot{X}$ means to compare

$$\Pr[\dot{Y}|\ddot{X}=x'] \quad \text{and} \quad \Pr[\dot{Y}|\ddot{X}=x'']$$

for (at least) two values, x' and x'', of $\ddot{X}$.

This comparison concerns conditional distributions and cannot, in general, be summarized by a single number.

One therefore often uses a simplified definition which only compares expected values:

$$\mathrm{E}(\dot{Y}|\ddot{X}=x'') - \mathrm{E}(\dot{Y}|\ddot{X}=x') \tag{2}$$

This suffices if $\dot{Y}$ is a binary variable.

3.5 Defining effects of explanatory variables (cont')

One has to take into account that $\dot{Y}$ most often depends on further variables.

Then effects cannot simply be attributed to a change in $\ddot{X}$, but are *context-dependent*.

Formally, assume that $\dot{Y}$ also depends on $\ddot{Z}$. The effect of a change in $\ddot{X}$ must then be written as

$$\mathrm{E}(\dot{Y}|\ddot{X}=x'',\ddot{Z}=z) - \mathrm{E}(\dot{Y}|\ddot{X}=x',\ddot{Z}=z) \tag{3}$$

and, in general, depends on the *covariate context* specified by $\ddot{Z}=z$.

In this situation, already the question 'What is THE effect of $\ddot{X}$?' would be misleading.

3.6 Interaction of explanatory variables

If the effect defined in (3) depends on the value of $\ddot{Z}$ one can say that both variables interact.

A more general definition: $\ddot{X}$ and $\ddot{Z}$ are *interactive conditions* for the distribution of $\dot{Y}$ if the effect of a change in $\ddot{X}$ [$\ddot{Z}$] depends on values of $\ddot{Z}$ [$\ddot{X}$].

The formulation shows that the presence of interaction also depends on the definition of 'effect'.

3.6 Interaction of explanatory variables (cont'd)

Consider a logit model without an explicitly defined multiplicative interaction term:

$$\Pr(\dot{Y}=1\,|\,X=x, Z=z) = \frac{\exp(\alpha + x\beta_x + z\beta_z)}{\exp(\alpha + x\beta_x + z\beta_z)}$$

There is an interaction when using definition (3), but not when using odds ratios for a definition of effects:

$$\frac{\frac{\Pr(\dot{Y}=1\,|\,X=x'',Z=z)}{\Pr(\dot{Y}=0\,|\,X=x'',Z=z)}}{\frac{\Pr(\dot{Y}=1\,|\,X=x',Z=z)}{\Pr(\dot{Y}=0\,|\,X=x',Z=z)}} = \frac{\exp(\alpha + x''\beta_x + z\beta_z)}{\exp(\alpha + x'\beta_x + z\beta_z)} = \exp((x''-x')\beta_x)$$

3.6 Interaction of explanatory variables (cont'd)

Interaction must be distinguished from stochastic relationships between explanatory variables. For example, think of Model 3.1 with an additional arrow from $\ddot{X}$ to $\dot{Z}$ (which is then an endogenous variable).

Interaction is independent of how values of the interacting variables come into being.

Numerical example:

with interaction			without interaction			
x	z	$\mathrm{E}(\dot{Y}\|\ddot{X}=x,\ddot{Z}=z)$	x	z	$\mathrm{E}(\dot{Y}\|\ddot{X}=x,\ddot{Z}=z)$	
0	0	0.5	0	0	0.5	(4)
0	1	0.7	0	1	0.7	
1	0	0.8	1	0	0.7	
1	1	0.9	1	1	0.9	

4. Causal interpretations
4.1 Causally relevant variables

Consider a functional relationship

$$x \longrightarrow \Pr[\dot{Y}|\ddot{X}=x] \qquad (\text{often: } x \longrightarrow \mathrm{E}(\dot{Y}|\ddot{X}=x))$$

What is entailed by the idea that this can be viewed as a causal relationship? I propose that there are basically two claims:

(1) that one can refer to a fact-generating process generating values of $\dot{Y}$, and

(2) that this process depends on values of $\ddot{X}$.

I then call $\ddot{X}$ a variable which is *causally relevant* for $\dot{Y}$.

These two claims are at the core of viewing causation as a generative process (Goldthorpe 2001, Blossfeld 2009).

In order to think about how processes generating values of $\dot{Y}$ depend on $\ddot{X}$ one must be more specific about the meaning of the values of $\ddot{X}$.

4.2 Thinking of causes as events

I begin with assuming that values of $\ddot{X}$ represent possible events. In the most simple case, $\ddot{X}$ is a binary variable, and $\ddot{X}=1$ means that an event of a specified kind has occurred, and $\ddot{X}=0$ means that such an event has not (yet) occurred.

One can then distinguish two kinds of relationships between $\ddot{X}$ and a process generating values of $\dot{Y}$:

(a) The event initiates a process that generates a value of $\dot{Y}$. This presupposes that the occurrence of the event (= an event of this kind) is a necessary condition for the generation of a value of $\dot{Y}$.

As an example, think of $\dot{Y}$ as the outcome of a student's participating in a learning frame. The student's *beginning* to participate in the learning frame can be considered as an event that initiates a process that eventually generates a value of $\dot{Y}$.

4.2 Thinking of causes as events (cont'd)

(b) The event occurs while a process that eventually generates a value of $\dot{Y}$ already takes place. So one can think of two such processes: one during which the event did occur, and another one in which the event did not occur. The impact of the event, if it occurs, must then be understood as modifying an ongoing process.

As an example, one can think that the student becomes severely ill while participating in the learning frame.

Most often, already the definition of a process requires to refer to an event that initiates the process.

The causal relevance of that event is then easily stated: Its occurrence is a necessary condition for the process to take place.

4.2 Thinking of causes as events (cont'd)

Variables representing the occurrence of events will be called *event variables*.

These need not be binary variables which refer to just one event type. In general, if $\dot{X}$ is an event variable, its domain will be denoted by

$$\mathcal{X} = \{0, 1, \ldots, m\}$$

with values having the following meaning:

If $x > 0$, $\dot{X} = x$ means that an event of the type x has occurred; $\dot{X} = 0$ means that no event of the specified kinds has yet occurred.

The definition shows that using event variables at least implicitly requires a temporal view.

4.3 Causally relevant conditions

Causally relevant variables need not be event variables.

As another kind one can think of variables representing conditions on which a process depends. I then speak of *context variables*.

To illustrate, I use our standard example in which a student's educational success, $\dot{Y}$, depends on the school type, $\dot{X}$, and on the parents' educational level, $\ddot{Z}$.

The functional model looks as follows:

Model 4.1

$$\begin{tikzcd} \dot{X} \arrow[r, two heads] & \dot{Y} \\ \ddot{Z} \arrow[u, two heads] \arrow[ur, two heads] & \end{tikzcd}$$

4.3 Causally relevant conditions (cont'd)

In this model, one can think of $\dot{X}$ as a variable representing the occurrence of an event:

$$\dot{X} = \begin{cases} 1 & \text{if the student starts participating in learning frame } \sigma_1 \\ 2 & \text{if the student starts participating in learning frame } \sigma_2 \end{cases}$$

In contrast, $\ddot{Z}$, recording the parents' educational level, is a context variable. Its values can sensibly be understood as characterizing the context in which the process that generates a value of $\dot{X}$ takes place.

Similarly, one can understand the causal relevance of $\ddot{Z}$ for processes generating values of $\dot{Y}$.

The leading idea is that parents' activities through which they support a child's education depend on their own educational level.

4.3 Causally relevant conditions (cont'd)

How to think of the causal relevance of $\dot{X}$? There are two considerations.

First, thinking of the event $\dot{X} = j$, it can be understood as initiating a process that eventually generates a value of $\dot{Y}$. In this view, $\dot{X}$ is causally relevant because without $\dot{X}$'s taking a positive value a process generating a value of $\dot{Y}$ cannot take place.

Second, as soon as one of the possible events did occur, it can be viewed as having generated a specific context. If $\dot{X} = j$, it is the context σ_j, a particular learning frame, in which the process generating a value of $\dot{Y}$ takes place.

This can be generally stated: As soon as an event variable has a positive value, the variable can be considered as a context variable for a process that begins at the point in time when the event occurred.

Note: When an event variable has taken a particular value, it cannot change anymore. This is different with context variables.

4.4 Comparative and dynamic effects

To speak of a causally relevant variable, say $\dot{X}$, presupposes a functional model in which the variable has a particular place and can be functionally related to other variables representing possible effects.

However, it is not the model, understood as a system of mathematical functions, that provides the causal meaning.

To give a variable a causal meaning requires considerations which cannot be expressed in terms of mathematical functions.

The functional model is silent about the meaning of its functional relationships (think of our example). But given a causal interpretation, it can be used to formally define causal effects.

There are two possibilities.

4.4 Comparative and dynamic effects (cont'd)

A) Comparative effects.

One specifies a variable, say $\dot{Y}$, representing the outcomes of interest, and considers all variables on which $\dot{Y}$ functionally depends.

Assume that these are the variables $\dot{X}$ and $\dot{Z}$. Both can then be used to define effects.

For example, an effect of $\dot{X}$ can be defined by

$$\begin{aligned} &\Delta^s(\dot{Y}; \dot{X}[x', x''], \dot{Z} = z) = \\ &\qquad \mathrm{E}(\dot{Y} | \dot{X} = x'', \dot{Z} = z) - \mathrm{E}(\dot{Y} | \dot{X} = x', \dot{Z} = z) \end{aligned} \tag{5}$$

This definition compares the expectation of $\dot{Y}$ in two situations: one in which $\dot{X} = x'$ and another one in which $\dot{X} = x''$, and further presupposes that $\dot{Z} = z$ in both situations.

In this sense, the definition formulates a *comparative effect*, and can be used for all kinds of causally relevant variables.

4.4 Comparative and dynamic effects (cont'd)

B) Dynamic effects.

If $\dot{X}$ is an event variable, also another effect definition becomes possible:

$$\Delta^d(\dot{Y}; \dot{X}[j], \dot{Z} = z) = \mathrm{E}(\dot{Y} | \dot{X} = j, \dot{Z} = z) - \mathrm{E}(\dot{Y} | \dot{X} = 0, \dot{Z} = z) \tag{6}$$

This definition formulates a *dynamic effect*; it compares a situation where the event $\dot{X} = j$ occurred with a situation where no event (of the kinds specified by $\dot{X}$) occurred.

If the occurrence of an event specified by $\dot{X}$ is a necessary condition for $\dot{Y}$'s getting a value, I use the convention that $\mathrm{E}(\dot{Y} | \dot{X} = 0, \dot{Z} = z) = 0$.

To illustrate, consider the example introduced in Section 4.3. A comparative effect compares the educational outcomes in the two learning frames, σ_1 and σ_2.

In contrast, a dynamic effect compares the outcome of $\dot{X} = j$ with $\dot{X} = 0$, and since a positive value of $\dot{X}$ is a necessary condition for a value of $\dot{Y}$, the dynamic effect is given by $\mathrm{E}(\dot{Y} | \dot{X} = j, \dot{Z} = z)$.

4.5 Functional and causal mechanisms

Without presupposing possible effects one cannot think of 'causes'.

In a statistical approach to causality, possible effects are conceptualized by an outcome variable, $\dot{Y}$.

In social research, being interested in processes generating values of $\dot{Y}$, it is seldom reasonable to consider only a single causal condition, say $\ddot{X}$.

In most applications one has to take into account further causally relevant conditions.

The question 'What is the causal effect of $\ddot{X}$ on $\dot{Y}$?', without further qualification, is then not appropriate.

4.5 Functional and causal mechanisms (cont'd)

An alternative is to think of 'mechanisms'.

Here I use the term in this sense: A *mechanism* is an explicitly defined framework for thinking of processes generating values of an outcome variable.

More specifically, I use the term *functional mechanism* to denote a functional model that shows how an outcome variable depends on other variables.

It will be called a *causal mechanism* if at least some of the functional relationships can be given a causal interpretation.

4.5 Functional and causal mechanisms (cont'd)

Given these definitions, a mechanism is a (formal) framework and must be distinguished from the processes which, possibly, take place according to the rules of the mechanism.

This entails that a mechanism is not by itself a dynamic entity.

While one can sensibly think that a process can generate an outcome, this cannot be said of a mechanism.

But note that only the mechanism has an explicit representation (in terms of variables). To think of processes that actually generate outcomes requires a causal interpretation of the mechanism.

5. Models and references to actors
5.1 Primary and secondary actors

Models in social research most often concern processes which depend on the behavior of human actors.

I call these the *primary actors*.

In contrast, I speak of *secondary actors* when referring to those who construct and use models.

For example, think of the models dealing with students' educational outcomes in different learning frames. Primary actors are the students, their parents, the teachers; in general, all actors to which one refers when interpreting the models and reflecting about causal relationships. In contrast, the secondary actors are those who construct, discuss, and possibly use these models for one reason or another.

Note that the distinction presupposes the reference to a model. Only w.r.t. a model can one distinguish primary and secondary actors in the proposed sense.

5.2 Explanatory and treatment models

There are many reasons why one could be interested in models. Here I want to mention just one distinction that also suggests to distinguish two kinds of models.

1) One interest concerns the primary actors, the conditions and outcomes of their behavior.

   Models are then constructed as tools for understanding and explaining conditions and outcomes of the behavior of the primary actors. I then speak of *explanatory models*.

2) In contrast, secondary actors could be interested in the possibility of interventions supporting their economic and/or political goals.

   Models are then constructed as tools for assessing the possible effects of treatments. I then speak of *treatment models*.

5.3 Functional models of experiments

Discussions of causally interpretable models often presuppose an interest in effects of treatments. Grounded in a long tradition, treatment models are preferably related to an experimental context. I briefly mention some ways in which functional models can be used as a formal framework.

A first possibility is illustrated by the following model.

Model 5.1

$$\begin{tikzcd} \ddot{X} \arrow[r, two heads] & \dot{Y} \\ \ddot{Z} \arrow[ur, two heads] & \end{tikzcd}$$

Values of $\ddot{X}$ represent the treatments whose causal effects are of primary interest. $\ddot{Z}$ records further conditions which, presumably, are causally relevant for $\dot{Y}$. Both are exogenous variables because their values are deliberately fixed by the experimenter.

5.3 Functional models of experiments (cont'd)

In another kind of experiment, the experimenter randomly selects some of the conditions for the experiment but still deliberately generates values of the treatment variable.

This can be represented by the following model.

Model 5.2

$$\begin{tikzcd} & \ddot{X} \arrow[r, two heads] & \dot{Y} \\ \ddot{E} \arrow[r, two heads] & \dot{Z} \arrow[ur, two heads] & \end{tikzcd}$$

$\ddot{X}$ is still an exogenous variable, but $\dot{Z}$ is now an endogenous variable which depends in a specified way on $\ddot{E}$, an event variable initiating the experiment.

5.3 Functional models of experiments (cont'd)

In both situations, assuming that the experiment concerns processes depending on the behavior of primary actors, the experimenter is a secondary actor.

There also is then a potential conflict.

Being interested in predictable effects of treatments, the experimenter has reason to control and regulate the behavior of the primary actors as far as possible.

5.4 Randomly assigned causal conditions

There is a further kind of experiment in which also values of the treatment variable, $\dot{X}$, are randomly generated. The functional model then looks as follows:

Model 5.3

$$\begin{tikzcd} & \dot{X} \arrow[r, two heads] & \dot{Y} \\ \ddot{E} \arrow[ur, two heads] \arrow[r, two heads] & \dot{Z} \arrow[ur, two heads] & \end{tikzcd}$$

The set-up entails that, conditional on $\ddot{E}=1$ (initiation of the experiment), $\dot{X}$ and $\dot{Z}$ are independent.

In a standard notation, this can be written as

$$\dot{X} \perp\!\!\!\perp \dot{Z} \mid \ddot{E} = 1$$

5.4 Randomly assigned causal conditions (cont'd)

The independence relation allows one to express effects of $\dot{X}$ as follows:

$$\Delta^s(\dot{Y};\dot{X}[x',x''],\ddot{E}=1) = \sum_z \Delta^s(\dot{Y};\dot{X}[x',x''],\dot{Z}=z,\ddot{E}=1)\Pr(\dot{Z}=z|\ddot{E}=1) \quad (7)$$

This shows that a randomized experiment allows one to compare effects of different treatments (= values of $\dot{X}$) in a balanced way.

This means: the distribution of further possibly relevant variables, $\dot{Z}$, does not depend on the values of $\dot{X}$, the treatments, that one intends to compare.

But note that this does not entail that effects of $\dot{X}$ are independent of $\dot{Z}$. If $\dot{X}$ and $\dot{Z}$ interact in the generation of values of $\dot{Y}$, the effect defined in (7) still depends on the distribution of $\dot{Z}$.

In any case, it is a mean effect w.r.t. the distribution of any further causally relevant conditions.

5.4 Randomly assigned causal conditions (cont'd)

Notice that a reference to the variable $\ddot{E}$ is required in order to think of an experimental context which includes an experimenter who at least initiates the experiment.

Without the variable $\ddot{E}$, Model 5.3 only contains random variables getting their values from processes not represented in the model.

Also note that this is an intervention model in a very specific sense: randomization is used as a fact-generating process, that is, a process generating events (values of $\dot{X}$).

5.5 Two contexts for randomization

Is randomization useful in social research? In thinking about this question one should distinguish between two different contexts for randomization.

1) One can think of data-generating processes. The most relevant application concerns the selection of units (for further observation).

   There are good arguments that samples should be generated randomly, that is, according to in some way fixed and known selection probabilities.

2) An essentially different context is experimentation. Performing an experiment requires, first of all, a fact-generating process.

   In this context, randomization not only, if at all, concerns the selection of a sample of units for the experiment; but also is a method of creating facts ('treatments') whose possible effects one intends to study.

5.5 Two contexts for randomization (cont'd)

While randomization in data generation is certainly useful in social research, the value of randomized experiments seems questionable.

The argument is not that such experiments are seldom possible.

The relevant point is that randomization would change, in an essential way, the processes to be studied.

The example depicted in Model 4.1 can show this. In this example, one can be interested in effects of the learning frames ($\dot{X}$).

The model realistically assumes that the selection of learning frames depends on the student's family background (represented by $\ddot{Z}$).

This is a fact-generating process. Randomization would substitute this by another fact-generating process that randomly assigns students to learning frames.

Consequently: quite different kinds of models.

5.6 Randomization and interaction effects

Even if randomization is possible and does not destroy the object of modeling, the method can easily be misleading if causally relevant variables interact.

Example where X is randomized w.r.t. Z.

x	z	$y=0$	$y=1$	cases
0	0	70	30	100
1	0	100	100	200
0	1	30	70	100
1	1	100	100	200

Conditional effects:

$$\mathrm{E}(\dot{Y}|\ddot{X}=1,\dot{Z}=0) - \mathrm{E}(\dot{Y}|\ddot{X}=0,\dot{Z}=0) = 0.5 - 0.3 = 0.2$$

$$\mathrm{E}(\dot{Y}|\ddot{X}=1,\dot{Z}=1) - \mathrm{E}(\dot{Y}|\ddot{X}=0,\dot{Z}=1) = 0.5 - 0.7 = -0.2$$

Marginal effect resulting from randomization:

$$\mathrm{E}(\dot{Y}|\ddot{X}=1) - \mathrm{E}(\dot{Y}|\ddot{X}=0) = 0.5 - 0.5 = 0$$

5.6 Randomization and interaction effects (cont'd)

The marginal effect heavily depends on the distribution of Z.
Same example with a different distribution of Z:

x	z	$y=0$	$y=1$	cases
0	0	35	15	50
1	0	30	30	60
0	1	60	140	200
1	1	120	120	240

Same conditional effects, but marginal effect resulting from randomization:

$$\mathrm{E}(\dot{Y}|\ddot{X}=1) - \mathrm{E}(\dot{Y}|\ddot{X}=0) = \frac{150}{300} - \frac{155}{250} = -0.12$$

6. Two understandings of 'potential outcomes'
6.1 Rule-based understanding of potential outcomes

The approach to understanding causal relationships that was sketched in Chapter 4 uses functional models as a formal framework.

This is appropriate when one is interested in causally interpretable rules. Such rules – I briefly speak of *causal rules* – concern potential outcomes which can be linked to different values of causally relevant variables.

Such a rule often has the form

$$(x,z) \longrightarrow \mathrm{E}(\dot{Y}|\dot{X}=x,\dot{Z}=z) \quad (8)$$

It is a causal rule if $\dot{X}$ and $\dot{Z}$ can be interpreted as variables which are causally relevant for the generation of values of $\dot{Y}$.

The dependent variable, $\dot{Y}$, represents *potential outcomes* which can be expected in a generic situation where $\dot{X}$ and $\dot{Z}$ have specified values.

6.2 Descriptive notion of potential outcomes

In the statistical literature, one also finds a descriptive notion of potential outcomes which is not based on rules.

This notion relates to a specified set of particular units, say Ω, and presupposes three (or more) statistical variables.

A variable

$$X:\Omega \longrightarrow \mathcal{X}$$

represents causally relevant factors (conditions or events).

As is often done in the literature, in order to simplify notations, I assume that X is a binary variable:

$$X = \begin{cases} 1 & \text{if a specified causal factor is present} \\ 0 & \text{otherwise} \end{cases}$$

6.2 Descriptive notion of potential outcomes (cont'd)

The interest concerns outcomes in situations where $X = 1$ or $X = 0$.

It is assumed that these outcomes can be represented by statistical variables

$$Y_j : \Omega \longrightarrow \mathcal{Y} \tag{9}$$

($j = 0$ or 1) having the following meaning: If $X(\omega) = j$, the outcome of interest has the value $Y_j(\omega)$.

One can then formally define, for each unit $\omega \in \Omega$, a causal effect $Y_1(\omega) - Y_0(\omega)$. Of course, these individual causal effects cannot be observed.

One therefore aims to estimate an *average causal effect* which can be defined for Ω by

$$\mathrm{M}(Y_1) - \mathrm{M}(Y_0) \tag{10}$$

6.2 Descriptive notion of potential outcomes (cont'd)

Observations only allow estimation of conditional mean values, $\mathrm{M}(Y_j|X=j)$.

So the question arises under which conditions one can think of these conditional mean values as unbiased estimates of $\mathrm{M}(Y_j)$.

A sufficient condition would be that X and Y_j are approximately independent; formally: $Y_j \perp\!\!\!\perp X$ (for $j = 0, 1$).

This independence condition suggests that a critical question concerns the generation of values of X.

If possible, such values should be randomly assigned to the members of Ω.

This would justify to consider $\Omega_j := \{\omega \in \Omega | X(\omega) = j\}$ as a simple random sample from Ω, and therefore to assume that Y_j has approximately the same distribution in Ω_j and Ω.

6.3 Including further causally relevant variables

As described in the previous section, the descriptive approach aims to define a causal effect that can be attributed to a single variable, X. A somewhat extended formulation is required if effects also depend on further variables. Assume that outcomes depend not only on X, but also on values of a variable Z (possibly consisting of several components). Instead of (10), one has to consider the effect definition

$$\mathrm{M}(Y_1|Z=z) - \mathrm{M}(Y_0|Z=z) \tag{11}$$

As before, values of Y_j can only be observed if $X=j$; the observable conditional mean values are $\mathrm{M}(Y_j|X=j, Z=z)$. They provide unbiased estimates of $\mathrm{M}(Y_j|Z=z)$ if

$$Y_j \perp\!\!\!\perp X \,|\, Z = z \quad \text{(for } j = 0, 1) \tag{12}$$

This shows that it would suffice to perform the randomization (the random assignment of values of X to the members of Ω) separately for each value of Z.

6.4 Balanced effects and kinds of models

Neither the rule-based nor the descriptive approach requires that the explicitly represented variables, X and Z, are independent.

Relationships between these variables become important, however, if one aims to define causal effects of just one variable, say X.

Whether this is possible depends first of all on whether X and Z interact in the generation of outcomes.

If they interact, effects cannot be attributed solely to X.

6.4 Balanced effects and kinds of models (cont'd)

It is nevertheless possible to define average effects.

Following the rule-based approach, one can use the definition

$$\sum\nolimits_z (\mathrm{E}(\dot{Y}|\dot{X}=x'', \dot{Z}=z) - \mathrm{E}(\dot{Y}|\dot{X}=x', \dot{Z}=z))\, \Pr(\dot{Z}=z) \tag{13}$$

where $\Pr(\dot{Z}=z)$ refers to an (arbitrarily) specified distribution of $\dot{Z}$.

This is a *balanced effect*, meaning that the distribution of $\dot{Z}$ is identical for x' and x''.

Of course, if $\dot{X}$ and $\dot{Z}$ interact, the effect still depends on the assumed distribution of $\dot{Z}$.

6.4 Balanced effects and kinds of models (cont'd)

How to proceed when following the descriptive approach depends on the given data to which the causal statements relate.

If the data result from a process which entails a randomization of X w.r.t. Z, the distribution of Z already is approximately independent of X, and one can interpret (10) as a balanced average effect of X. Again, if X and Z interact, this effect also depends on the distribution of Z in the reference set of units.

If X and Z are not independent, one can construct a balanced effect. This is analogous to the procedure in the rule-based approach. In the descriptive approach, one starts from (11) and (arbitrarily) specifies a distribution of Z. Formally analogous to (13), an average effect can then be defined by

$$\sum\nolimits_z (\mathrm{M}(Y_1|Z=z) - \mathrm{M}(Y_0|Z=z))\, \mathrm{P}(Z=z) \tag{14}$$

6.4 Balanced effects and kinds of models (cont'd)

One might ask whether balanced effects are particularly useful. This depends on the kind of model.

With treatment models, one is normally interested in finding an effect that can be attributed solely to the treatment, given that all other possibly relevant conditions are in some sense fixed. This interest suggests to construct balanced effects.

The situation is different with explanatory models. In social research, explanatory models most often relate to situations where at least some of the causally relevant conditions are generated by actions of primary agents.

Effects of single variables are then never balanced w.r.t. all causally relevant conditions. It would be possible, of course, to construct balanced effects w.r.t. observed variables; but I think that a primary interest concerns how the real effects, which are unbalanced, come into being.

6.5 Contrasting the two approaches

1) A first difference concerns the notion of potential outcomes.

a) The rule-based approach conforms to the understanding that potential outcomes are outcomes which, under specified conditions, possibly will come into existence.

   Correspondingly, potential outcomes are defined by a rule (a linguistic if-then construction).

b) The descriptive approach, in contrast, presupposes that potential outcomes (= values of Y_0 and Y_1) already exist before values of X, and other causally relevant variables, are fixed.

   To speak of 'potential outcomes' is therefore somewhat misleading. Actually, what is potentially realized is an observation of an already existing fact (value of Y_j). So it would be less confusing to speak of 'potential observations'.

6.5 Contrasting the two approaches (cont'd)

2) It might be helpful to remember the distinction between fact-generating and data-generating processes.

The rule-based approach aims to formulate causal rules for fact-generating processes.

The descriptive approach, as it is theoretically formulated, is concerned with data-generating processes which provide partial information about hypothetically presupposed facts (values of Y_0 and Y_1).

This approach therefore seems to allow one to think of 'causal inference' in parallel to a missing data problem.

6.5 Contrasting the two approaches (cont'd)

3) It is important to understand that the variables Y_0 and Y_1 can only be defined by referring to a set of existing units.

For each particular unit, say $\omega \in \Omega$, one can posit values, $Y_0(\omega)$ and $Y_1(\omega)$, representing the outcomes corresponding to $X(\omega)=0$ and $X(\omega)=1$, respectively.

This is possible because, and insofar, one can assume that all further conditions on which outcomes depend are implicitly fixed by the reference to ω, a particular unit existing in particular circumstances.

The descriptive approach is therefore essentially static and not well suited for causal interpretations of temporally extended processes.

6.5 Contrasting the two approaches (cont'd)

The rule-based approach, in contrast, is not based on a reference to a set of already existing units, but relates to generic units which are only defined by values of variables.

It is therefore not possible to define variables corresponding to Y_0 and Y_1.

Instead, there is a single outcome variable, $\dot{Y}$, having possible values which only become realized when, and after, $\dot{X}$, and any further variables which define the generic unit, have taken specific values.

There are no restrictions for thinking of a temporally extended process connecting $\dot{X}$ and the final outcome, $\dot{Y}$.

6.5 Contrasting the two approaches (cont'd)

4) As a consequence, the independence requirement (12) cannot be formulated in the conceptual framework of the rule-based approach.

Based on the mentioned understanding of $\dot{Y}$, one can define variables $\dot{Y}_j$ having distributions defined by $\Pr[\dot{Y}_j|\dot{Z}=z] = \Pr[\dot{Y}|\dot{X}=j, \dot{Z}=z]$.

An independence condition paralleling (12) is then trivially true:

$$\dot{Y}_j \perp\!\!\!\perp \dot{X} \,|\, \dot{X}=j, \dot{Z}=z \quad (\text{for } j=0,1) \tag{15}$$

But this condition has not the same interpretation.

(12) can be interpreted as the requirement that X, conditional on values of Z, is approximately independent of all further circumstances which are fixed by the implicit reference to particular units.

(15), in contrast, only says that the outcome variable, given $\dot{X}=j$ and $\dot{Z}=z$, is independent of any other outcome variable for which $\dot{X} \neq j$ and $\dot{Z}=z$.

6.5 Contrasting the two approaches (cont'd)

5) As mentioned, in order to satisfy the independence condition (12), there ideally should be a randomized assignment of values of X to the members of Ω (conditional on values of Z).

(15) does not require any randomization procedure.

The important point is, however, that formulating a causal rule like (8) does not entail the claim that there are no further variables on which the outcome variable depends.

Consequently, also definitions of causal effects which are derived from (8) do not entail anything about further variables on which the outcome variable depends.

Consider the causal effect defined in (5). This definition compares two generic units, one with $\dot{X}=x''$ and the other one with $\dot{X}=x'$. Both units have identical values of $\dot{Z}$; but they can differ in all other respects.

6.6 Omitted causally relevant conditions

If a causal rule does not relate to an artificial random generator, one can almost always think that the rule misses one or more causally relevant conditions.

Note that this is true even if the data used to estimate the rule result from a randomized experiment.

The point simply is that there probably are causally relevant conditions not explicitly referred to in the rule's formulation.

It is therefore not reasonable to require that a causal rule entails the claim that one has taken into account all causally relevant conditions.

6.6 Omitted causally relevant conditions (cont'd)

Moreover, except when dealing with artificial random generators, already the assumption that one can 'theoretically' refer to a complete set of variables which are causally relevant for an outcome variable seems obscure.

The descriptive approach to potential outcomes avoids this assumption and instead requires the conditional independence (12).

This independence is viewed as a precondition for thinking of a causal effect of X.

(12) cannot be formulated in a rule-based approach. In a rule-based approach one would need to refer to explicitly defined variables which, in addition to Z, are causally relevant for Y.

If one could refer to a list of such variables, say $(U_1, U_2, \ldots)$, one could use the formulation $X \perp\!\!\!\perp (U_1, U_2, \ldots) \,|\, Z = z$. However, the formulation is not useful because one cannot define, not even clearly think of, such a list of variables.

6.6 Omitted causally relevant conditions (cont'd)

Of course, it is often quite possible to think of a particular variable, say U, which is left out in the formulation of a causal rule, but should be taken into account in order to get a better understanding of the causal mechanism.

The original model, that was used to derive the causal rule, must then be enlarged by incorporating U; and this also demands to specify U's relationship with the other variables in the model.

How to do this depends on the intended use of the model.

If the model is intended to represent a randomized experiment, one can assume in the formulation of the enlarged model that $\dot{X} \perp\!\!\!\perp \dot{U} \,|\, \dot{Z} = z$.

Conditional on $\dot{Z} = z$, effects of $\dot{X}$ are then balanced w.r.t. $\dot{U}$.

6.6 Omitted causally relevant conditions (cont'd)

In social research an explanatory model can almost never be formulated as a model representing a randomized experiment.

It then depends on the details of the model how to think of $\dot{U}$'s role in the mechanism generating values of the outcome variable.

Remember the distinction between mediating and confounding variables.

In any case, one would need observations of $\dot{U}$ in order to quantify its causal role.

7. Selection and choice
7.1 Distinguishing selection problems

I distinguish three situations where Y is a variable of interest, and there is a further variable S which in some sense involves a selection.

a) S is a binary variable, and values of Y can be observed if $S = 1$, and cannot be observed if $S = 0$. This can properly be called a 'sample selection problem' because S only concerns the observability of Y but is not a causally relevant condition for Y. Here it is presupposed that Y has a distribution that exists independently of S.

b) S is a binary variable, and $S = 1$ is a necessary precondition for Y to have a distribution. For example, being employed ($S = 1$) is a necessary precondition for receiving a wage (Y).

c) S can take two or more different values, and it is assumed that the distribution of Y causally depends on the value of S. For example, values of S represent school types, and Y is a measure of educational attainment.

7.2 Selection in data-generating processes

Sample selection problems can be conceptualized in two ways.

First, Y and S are statistical variables, say

$$(Y, S) : \Omega \longrightarrow \mathcal{Y} \times \{0, 1\}$$

One knows the conditional distribution $\mathrm{P}[Y|S\!=\!1]$, but is interested in the unconditional distribution $\mathrm{P}[Y]$.

As an example, one can think that S is a response indicator in a survey: $S\!=\!1$ if a sampled unit provides a value of Y, and $S\!=\!0$ otherwise.

7.2 Selection in data-generating processes (cont'd)

Another framework uses random variables, say $\dot{Y}$ and $\dot{S}$. $\dot{S}$ is again a binary variable and records whether a value of $\dot{Y}$ can be observed. So it is assumed that one knows the conditional distribution $\Pr[\dot{Y}|\dot{S}\!=\!1]$, but is interested in the unconditional distribution $\Pr[\dot{Y}]$.

If $\dot{S}$ is independent of $\dot{Y}$, one can use $\Pr[\dot{Y}|\dot{S}\!=\!1]$ to estimate $\Pr[\dot{Y}]$. Problems occur if, and because, $\dot{S}$ and $\dot{Y}$ are correlated. This suggests to find another variable, say $\ddot{V}$ (often consisting of several components), such that

$$\dot{S} \perp\!\!\!\perp \dot{Y} \mid \ddot{V}\!=\!v$$

is approximately true. This would allow one to use $\Pr[\dot{Y}|\ddot{V}\!=\!v, \dot{S}\!=\!1]$ to estimate $\Pr[\dot{Y}|\ddot{V}\!=\!v]$.

7.3 Selection as a necessary precondition

I now consider the case where the selection variable represents a necessary precondition for values of the variable of interest.

As an example, I take $\dot{Y}$ to represent the outcome of a university education, and $\dot{S}\!=\!1$ if a person is admitted to begin with university studies, and $\dot{S}\!=\!0$ otherwise. It is further assumed that $\dot{Y}$ depends on a variable $\ddot{X}$ representing the level of academic performance the person has reached just before the selection variable gets a particular value.

A functional model could then be graphically depicted as follows:

Model 7.1

$$\begin{tikzcd}
\ddot{X} \arrow[r] \arrow[d] & \dot{Y} \\
\dot{S} \arrow[ur] &
\end{tikzcd}$$

7.3 Selection as a necessary precondition (cont'd)

The picture is possibly misleading, however, because it suggests that $\dot{Y}$ has a distribution even if $\dot{S}\!=\!0$.

But, obviously, a process generating a value of $\dot{Y}$ can only take place if $\dot{S}\!=\!1$.

This entails that there are two rules:

$$x \longrightarrow \Pr(\dot{Y}\!=\!y|\ddot{X}\!=\!x, \dot{S}\!=\!0) = 0 \tag{16}$$

and

$$x \longrightarrow \Pr(\dot{Y}\!=\!y|\ddot{X}\!=\!x, \dot{S}\!=\!1) \tag{17}$$

7.3 Selection as a necessary precondition (cont'd)

Nevertheless, both $\ddot{X}$ and $\dot{S}$ are causally relevant for $\dot{Y}$.

Referring to expectations of $\dot{Y}$, $\dot{S}$ can be viewed as an event variable having the effect $\mathrm{E}(\dot{Y}|\ddot{X}\!=\!x, \dot{S}\!=\!1)$.

Effects of $\ddot{X}$ can be defined, of course, only conditional on $\dot{S}\!=\!1$:

$$\mathrm{E}(\dot{Y}|\ddot{X}\!=\!x'', \dot{S}\!=\!1) - \mathrm{E}(\dot{Y}|\ddot{X}\!=\!x', \dot{S}\!=\!1) \tag{18}$$

Note that in this model there can be no interaction of $\ddot{X}$ and $\dot{S}$ w.r.t. $\dot{Y}$.

7.4 Counterfactual and modal questions

Although $\dot{S}\!=\!1$ is a necessary precondition for values of $\dot{Y}$, one can ask hypothetical questions. Two forms of such questions must be distinguished:

a) *Counterfactual questions* presuppose that $\dot{S}$ already has taken a particular value. For example: Given $\ddot{X}\!=\!x$ and $\dot{S}\!=\!0$, what value of $\dot{Y}$ might be expected if $\dot{S}$ had taken the value 1 instead of 0? (The complementary question obviously has a trivial answer.)

b) *Modal questions* presuppose a situation where $\dot{S}$ has not already taken a particular value, and a process that might generate a value of $\dot{Y}$ has not yet started.

I will not discuss the counterfactual questions. To answer the modal questions, one can use the rules (16) and (17), respectively. The selection variable $\dot{S}$ then only serves to distinguish the two modal questions and does not provide any further information.

7.5 Is there a sample selection problem?

The rule (16) can be established by referring to the institutional framework without reference to sampled data.

Numerically specified versions of the rule (17) must be estimated from sampled data.

However, as suggested by the rule's formulation, one can simply use the data for those students who actually began a university education ($\dot{S}=1$).

Note that selection on $\dot{S}=1$ has nothing to do with a sample selection problem. The condition $\dot{S}=1$ simply determines the scope of the rule to be estimated.

It is possible, of course, that the available data are selective; for example, response rates can depend on $\dot{Y}$. There is then a sample selection problem that leads to biased estimates of $\Pr(\dot{Y}=y|\ddot{X}=x,\dot{S}=1)$.

But this bias is not due to conditioning on $\dot{S}=1$.

7.6 Consideration of omitted confounders

There is, however, another problem that must be considered: the omission of possibly relevant confounding variables. This requires to refer to an enlarged model that explicitly contains a further confounding variable. I use the following

Model 7.2

$$\begin{tikzcd} \ddot{X} \arrow[r] \arrow[d] & \dot{Y} \\ \dot{S} \arrow[ur] & \ddot{Z} \arrow[u] \arrow[l] \end{tikzcd}$$

where $\ddot{Z}$ is a further variable on which $\dot{S}$ depends, e.g., an indicator of a person's educational aspiration. Moreover, it is assumed that also $\dot{Y}$ depends on $\ddot{Z}$ entailing that $\ddot{Z}$ is a confounding variable.

7.6 Consideration of omitted confounders (cont'd)

There is no problem if data are available for both $\ddot{X}$ and $\ddot{Z}$.

But now assume that data on $\ddot{Z}$ are not available so that one can only estimate a reduced version of the model; in terms of expectations:

$$\begin{aligned} &\mathrm{E}(\dot{Y}|\ddot{X}=x,\dot{S}=1) = \\ &\quad \sum\nolimits_z \mathrm{E}(\dot{Y}|\ddot{X}=x,\dot{Z}=z,\dot{S}=1)\,\Pr(\dot{Z}=z|\ddot{X}=x,\dot{S}=1) \end{aligned} \tag{19}$$

where $\dot{Z}$ is used instead of $\ddot{Z}$ in order to allow thinking of conditional distributions.

This shows that effects of $\ddot{X}$, as defined in (18), do not have a balanced formulation in the reduced model. Of course, this is the general problem resulting from the omission of confounding variables.

7.6 Consideration of omitted confounders (cont'd)

Here it is important, however, that the problem does not result from conditioning on $\dot{S}$.

It is true that $\dot{S}$ is a collider, entailing that conditioning on $\dot{S}$ changes the correlation between values of $\ddot{X}$ and $\ddot{Z}$ in a reference set (sample).

But this is not relevant here because effect definitions are conditional on $\dot{S}=1$; in the example, they only concern students who actually begin with university studies.

Problems resulting from the omission of confounding variables should therefore be clearly distinguished from 'selection problems'.

7.7 Illustration with a numerical example

To illustrate the argument, I use these fictitious data.

x	z	s	$y=0$	$y=1$	cases
0	0	0	–	–	800
0	0	1	100	100	200
0	1	0	–	–	600
0	1	1	120	280	400
1	0	0	–	–	400
1	0	1	120	480	600
1	1	0	–	–	200
1	1	1	80	720	800

7.7 Illustration with a numerical example (cont'd)

Based on these data, one can estimate the rule

$$(x,z) \longrightarrow \mathrm{E}(\dot{Y}|\ddot{X}=x,\dot{Z}=z,\dot{S}=1) \tag{20}$$

and finds the values (21)

x	z	$\mathrm{E}(\dot{Y}|\ddot{X}=x,\dot{Z}=z,\dot{S}=1)$
0	0	0.5
0	1	0.7
1	0	0.8
1	1	0.9

Conditioning on $\dot{S}=1$ changes the relationship between $\ddot{X}$ and $\dot{Z}$: (22)

x	$\Pr(\dot{Z}=1|\ddot{X}=x,\dot{S}=0)$	$\Pr(\dot{Z}=1|\ddot{X}=x,\dot{S}=1)$
0	0.429	0.667
1	0.333	0.571

This is not relevant, however, because the rule (20) only applies to situations where $\dot{S}=1$.

7.7 Illustration with a numerical example (cont'd)

Now assume that values of $\dot{Z}$ are not available.

The data then lead to the following estimate of an effect of $\ddot{X}$:

$$\mathrm{E}(\dot{Y}|\ddot{X}=1,\dot{S}=1) - \mathrm{E}(\dot{Y}|\ddot{X}=0,\dot{S}=1) = 0.857 - 0.633 = 0.224 \tag{23}$$

This effect is not balanced. As can be seen from (22), the distribution of the omitted variable $\dot{Z}$ is different for $\ddot{X}=0$ and $\ddot{X}=1$.

This means that the effect cannot be attributed solely to a difference in $\ddot{X}$.

Without observations on $\dot{Z}$ one cannot assess the contribution of this variable.

8. Choice variables and models
8.1 A notion of choice variables

As I will use the term, a *choice variable*, say C, has the following features:

a) The domain of C is a set of $m \geq 2$ alternatives, numerically represented by $\mathcal{C} = \{1,\ldots,m\}$.
b) Referring to a choice variable entails that there is an individual or collective agent who can choose, or already has chosen, a particular value of the variable. The agent associated with a choice variable C will be denoted by $A[C]$.
c) It is presupposed that the agent has the power to select one of the alternatives. In other words, $\mathcal{C}$ must only contain states which can be realized by the agent.
d) The agent is assumed to have considered the alternatives before one of them is actually chosen. I do not assume that the agent is 'rational' in any particular sense.

Following this definition, choice is a specific kind of selection, namely a selection which is reflexively generated by an actor. This is meant by the expression 'choice-based selection' ($\neq$ 'choice-based sampling').

8.2 Two contexts for using choice variables

Choice variables can be used in two different contexts.

In one context, one conceives of a choice variable as a kind of event variable.

$C = c$ then means that the agent, $A[C]$, has chosen the alternative $c \in \mathcal{C}$. In addition, $C = 0$ means that such an event has not yet occurred.

In this understanding, choice variables can be used in explanatory models, both as explanatory variables and as dependent variables.

8.2 Two contexts for using choice variables (cont'd)

A model that attempts to explain choices, considered as events, will be called an *explanatory choice model*.

Such a model can be depicted as

$$\ddot{X} \twoheadrightarrow \dot{C}$$

where $\dot{C}$ is the choice variable, and $\ddot{X}$ denotes the explanatory variable (possibly consisting of several components).

For example, one can think that the model is intended to represent the choice of a child's school type; $A[\dot{C}]$ refers to the child's parents, and $\ddot{X}$ denotes the parents' educational level.

Note that this understanding entails that the parents have the power to choose a school type for their child.

8.2 Two contexts for using choice variables (cont'd)

In a quite different context, choice variables serve to consider possible consequences of hypothetically chosen alternatives.

To hypothetically give a choice variable a value is obviously different from an actual choice, it simply means to consider that alternative.

Thus, in this understanding, a choice variable is not an event variable; and in a functional model it can only be used as an exogenous variable.

8.2 Two contexts for using choice variables (cont'd)

As an example, think again of the parents' choice of a school type for their child. One can consider the following model:

Model 8.1

$$\begin{tikzcd} \ddot{X} \arrow[r] & \dot{Y}^* \\ \ddot{C} \arrow[ur] & \end{tikzcd}$$

The model relates to a generic $A[\ddot{C}]$, the parents who are assumed to consider possible values of $\ddot{C}$ (school types). As before, $\ddot{X}$ denotes the parents' educational level. $\dot{Y}^*$ is used to assess possible outcomes: the child's educational success that can be expected if, given $\ddot{X}$, the parents would choose $\ddot{C} = c$. The corresponding function is

$$(x, c) \longrightarrow \mathrm{E}(\dot{Y}^* | \ddot{X} = x, \ddot{C} = c) \tag{24}$$

8.2 Two contexts for using choice variables (cont'd)

A model of this kind will be called an *evaluative choice model*.

Its aim is not to predict realized choices. Instead, it is intended to serve thinking about modal questions.

Consequently, also $\dot{Y}^*$ cannot be understood as representing realized outcomes, and must be distinguished from an outcome variable in an explanatory model.

8.2 Two contexts for using choice variables (cont'd)

There are obviously similarities between an evaluative choice model and a treatment model.

But also note the different tasks.

A treatment model serves to formulate a generic rule about causal effects of treatments (= events which can be deliberately generated).

An evaluative choice model serves an agent to consider the consequences of available alternatives.

While randomization might be used for a treatment model, randomization w.r.t. the choice variable would contradict the idea of a choice.

8.3 Primary and secondary actors of choice models

Remember the distinction between primary and secondary actors.

The distinction can easily be applied to explanatory choice models. These models are concerned with choices made by primary actors. The secondary actors, in contrast, are those who construct and use these models for one reason or another.

Now consider an evaluative choice model. Since the model serves an agent to think about available alternatives and possible consequences of choosing one of them, the agent is a secondary actor w.r.t. to the model.

On the other hand, the model is concerned with the agent's choice, and so the agent can also be considered as a primary actor.

However, an evaluative model has not the task to predict the agent's choice.

'To choose' and 'to predict the outcome of a choice' are obviously different activities.

8.4 Consideration of modal questions

Let me stress the connection between modal questions and evaluative choice models. An explanatory choice model cannot be used to consider modal questions (or must be reinterpreted in some way).

As an illustration, I refer to Model 8.1. This model is intended to show how expectations of an outcome variable, $\dot{Y}^*$, depend on exogenously given values of $\ddot{X}$ and hypothetically chosen values of $\ddot{C}$.

In order to become useful, one needs a quantification of the function (24). Data might come from a sample that relates to the following model where it is assumed that $\dot{Y}$ has the same meaning as $\dot{Y}^*$.

Model 8.2

$$\begin{tikzcd} \ddot{X} \arrow[r] \arrow[d] & \dot{Y} \\ \dot{C} \arrow[ur] & \end{tikzcd}$$

8.4 Consideration of modal questions (cont'd)

In this model, $\dot{C}$ is an event variable, representing alternatives chosen by the primary agents to which the model relates. Since $\dot{C}$ depends on $\ddot{X}$, part of the model consists in an explanatory choice model.

Only the evaluative model 8.1, not the explanatory model 8.2, can be used for modal questions.

In the explanatory model, values of $\dot{C}$ come into being according to a function, $\ddot{X} \twoheadrightarrow \dot{C}$, that predicts the choices actually made.

This model can therefore not be used for a situation where a choice variable can hypothetically be given different values because a choice event has not yet occurred.

Nevertheless, data corresponding to the explanatory model 8.2 can also be used to estimate the function (24).

8.4 Consideration of modal questions (cont'd)

Such data would allow one to estimate $\mathrm{E}(\dot{Y}|\ddot{X}=x, \dot{C}=c)$, and then use the rule

$$\text{Estimate } \mathrm{E}(\dot{Y}^*|\ddot{X}=x, \ddot{C}=c) \text{ by } \mathrm{E}(\dot{Y}|\ddot{X}=x, \dot{C}=c) \tag{25}$$

As assumed by Model 8.2, the data result from 'self selection' in the sense of choices, made by primary actors, which depend on a variable $\ddot{X}$; $\dot{C}$ stochastically depends on $\ddot{X}$.

But this does not entail a sample selection problem that might create a bias when using the rule (25).

It is quite possible that Model 8.2 misses a relevant confounder and could be replaced by a better model (if the necessary data would be available). But this has nothing to do with the fact that the model contains an event variable that gets its values by choices of primary actors.

8.5 Modal questions w.r.t. necessary preconditions

I now consider a situation where the choice concerns a necessary precondition for a possible outcome.

To illustrate, I continue with the example that was introduced above: beginning with university studies and consideration of possible outcomes. Details depend on the set-up of the choice situation.

I consider two situations.

8.5 Modal questions w.r.t. necessary preconditions (cont'd)

(a) I begin with a situation where an agent has the power to decide whether a person (the agent herself or someone else) will begin with university studies.

There is then a choice variable, $\ddot{C}$, having the domain $\mathcal{C} = \{1, 2\}$; and $\ddot{C}=1$ means 'beginning' and $\ddot{C}=2$ means 'not beginning' with university studies.

The outcome assumed to be relevant for the choice is the success of the university education; it will be denoted by $\dot{Y}^*$.

I further assume that expectations about $\dot{Y}^*$ depend on an exogenous variable, $\ddot{X}$, representing properties of the person who possibly begins university studies that are known in the choice situation.

8.5 Modal questions w.r.t. necessary preconditions (cont'd)

With this notation, one can again use Model 8.1 as an evaluative model and Model 8.2 as a corresponding explanatory model.

The evaluative model is to be used for the two modal questions. First, what would be the outcome if $\ddot{C}=2$?

Obviously, without beginning university studies there could be no successful outcome ($\dot{Y}^*$ is then either undefined or has the value zero).

Now consider the other question, What is the expected value of $\dot{Y}^*$ if $\ddot{C}=1$, given $\ddot{X}=x$?

To answer this question, one needs an estimate of $\mathrm{E}(\dot{Y}^*|\ddot{X}=x, \ddot{C}=1)$.

This can be derived from the explanatory model 8.2 by using the rule (25). One only needs $\mathrm{E}(\dot{Y}|\ddot{X}=x, \dot{C}=1)$.

Whether also $\mathrm{E}(\dot{Y}|\ddot{X}=x, \dot{C}=2)$ can be given a sensible interpretation is irrelevant.

8.5 Modal questions w.r.t. necessary preconditions (cont'd)

(b) The argument in (a) presupposes that the choice variables in the two models, $\ddot{C}$ and $\dot{C}$, have the same meaning.

I now consider a situation where an agent cannot immediately decide whether to begin, or not begin, with university studies, but only whether to apply for admission.

The choice variable, $\ddot{C}$, has again the domain $\mathcal{C} = \{1, 2\}$, but now $\ddot{C}=1$ means 'to apply' and $\ddot{C}=2$ means 'not to apply'.

8.5 Modal questions w.r.t. necessary preconditions (cont'd)

An evaluative choice model can now be depicted as follows.

Model 8.3

$$\begin{tikzcd}
 & \ddot{X} \arrow[r, two heads] \arrow[d, two heads] & \dot{Y}^* \\
\ddot{C} \arrow[r, two heads] & \dot{S}^* \arrow[ur, two heads] &
\end{tikzcd}$$

$\ddot{X}$ and $\dot{Y}^*$ (and correspondingly $\dot{Y}$) have the same meaning as before. In addition, there is now the variable $\dot{S}^*$ representing the decision of the admission committee: $\dot{S}^*=1$ if the applicant is admitted, and $\dot{S}^*=0$ otherwise.

Both $\dot{Y}^*$ and $\dot{S}^*$ are starred in order to distinguish these variables from corresponding variables in an explanatory model.

8.5 Modal questions w.r.t. necessary preconditions (cont'd)

The model can be used for thinking about modal questions in two steps.

In a first step, the agent, $A[\ddot{C}]$, can consider $\mathrm{E}(\dot{S}^*|\ddot{X}=x, \ddot{C}=1)$. Given information from a corresponding explanatory model, one can use $\mathrm{E}(\dot{S}|\ddot{X}=x, \dot{C}=1)$ and a suitably modified version of rule (25).

In a second step, one can think about the final outcome, $\dot{Y}^*$, conditional on $\dot{S}^*=1$. In this step, $\mathrm{E}(\dot{Y}|\ddot{X}=x, \dot{S}=1)$ can be used as an estimate of $\mathrm{E}(\dot{Y}^*|\ddot{X}=x, \dot{S}^*=1)$. Finally, both steps can be combined:

$$\text{Estimate } \mathrm{E}(\dot{Y}^*|\ddot{X}=x, \ddot{C}=1) \text{ by } \mathrm{E}(\dot{Y}|\ddot{X}=x, \dot{S}=1)\,\mathrm{E}(\dot{S}|\ddot{X}=x) \tag{26}$$

Notice that the evaluative model 8.3 relates to a choice situation where the committee has not yet decided about $A[\ddot{C}]$'s application (simply because $A[\ddot{C}]$ has not yet decided whether to apply).

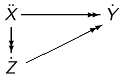
Possibly useful information can therefore only result from estimates of $\mathrm{E}(\dot{S}|\ddot{X}=x)$.

8.6 Choosing a model's perspective

The example discussed in the previous section shows that a model can simultaneously refer to two or more agents of a different kind.

Nevertheless, one has to choose a perspective. In the example, it is the perspective of a student who (possibly or actually) applies for a university study.

Now consider our standard example where a pupil's educational success (Y) depends on parents' education (X) and school type (Z). Graphically (Model 3.1):



So far, my interpretations implicitly assumed the parents' (or pupil's) perspective. However, one can also think of a situation where schools can select pupils, and do so by taking into accout the educational level of the parents. How to model this idea?

8.6 Choosing a model's perspective (cont'd)

One would need to change the model's perspective.

But note that this cannot be done by simply changing the direction of the arrow which (so far) leads from X to Z. (In fact, an arrow leading from Z to Z cannot be given a causal interpretation. Variables are not agents.)

Instead, one would need to change the object whose behavior is to be described by the model.

A) As depicted above, the model concerns the behavior of a generic pupil associated by his or her parents. This is made clear by the variables which represent properties of a generic pupil.

B) Alternatively, one can set up a model that is concerned with the behavior of schools. The model's objective then is a generic school, and the model's variables relate to this school.

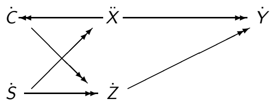
In both cases, there can be further agents. In case (A) one can additionally refer to schools, and in case (B) one can aditionally refer to pupils and parents.

8.6 Choosing a model's perspective (cont'd)

Here I briefly sketch a model taking the perspective (A), but assumes that parents do not have the power to choose a school type for the pupil.

The model is then similar to the example discussed in Section 8.5:

Model 8.4



$\dot{C}$ represents the school type intended by the parents, $\dot{S}$ represents the decision of the school, and $\dot{Z}$ is the realized school type.

The reduced model is the same as Model 3.1 with a single arrow leading from $\ddot{X}$ to $\dot{Z}$. But then it is no longer possible to consider dynamic effects.

9. Causal effects of choices
9.1 Three different choice situations

How to think of causal effects of choice variables depends on the kind of choice situation. I propose to distinguish three situations where $\dot{C}$ always is a binary choice variable with domain $\mathcal{C} = \{0, 1\}$.

(a) The choice concerns a necessary precondition for a process generating values of an outcome variable, say $\dot{Y}$, to take place. So the causal effect of $\ddot{C}$ can simply be stated: $\ddot{C}=1$ is a necessary precondition for $\dot{Y}$.

In order to quantify the effect, one can consider $\ddot{C}$ as an event variable and use

$$\mathrm{E}(\dot{Y}|\ddot{X}=x, \ddot{C}=1) \tag{27}$$

where $\ddot{X}$ represents conditions on which the process generating values of $\dot{Y}$ depends.

9.1 Three different choice situations (cont'd)

(b) Associated with the two alternatives are two qualitatively different outcome variables, say $\dot{Y}_0$ and $\dot{Y}_1$.

This implies that one has to consider two qualitatively different effects, one effect of $\ddot{C}=0$ and another one of $\ddot{C}=1$.

Both should be considered separately as indicated in (a).

(c) The choice concerns two different ways to generate values of a single outcome variable, $\dot{Y}$. So one can compare the alternatives w.r.t. expectations of $\dot{Y}$, and use the effect definition

$$\Delta^s(\dot{Y}; \ddot{C}[0,1], \ddot{X}=x) := \mathrm{E}(\dot{Y}|\ddot{X}=x, \ddot{C}=1) - \mathrm{E}(\dot{Y}|\ddot{X}=x, \ddot{C}=0) \tag{28}$$

9.1 Three different choice situations (cont'd)

In the following, I only discuss the situation (c), in particular, how to think about unobserved confounders.

Note that this problem is of no particular relevance in the situations (a) and (b).

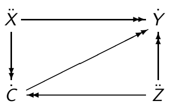
The conditional expectation referred to in (27) might well depend on further variables. But the effect of $\ddot{C}$ can always be interpreted as an average effect w.r.t. to distributions of these variables.

So this is different from the situation discussed in Chapter 7 where one is interested in effects of $\ddot{X}$ (e.g. education of married women).

9.2 The problem: omitted confounders

To focus the discussion, I consider

Model 9.1



The structure is the same as in Model 7.2, but it is now assumed that $\dot{Y}$ has a distribution for both alternatives.

As an example, one can think that the choice is between two learning frames, σ_0 and σ_1, where a person can acquire competencies represented by $\dot{Y}$, and it is assumed that the processes generating values of $\dot{Y}$ also depend on two further variables, $\ddot{X}$ and $\ddot{Z}$, having values already fixed in the choice situation.

9.2 The problem: omitted confounders (cont'd)

If data for both variables are available, one can use $\ddot{Z}=z$ as an additional condition in the effect definition (28). But if data on $\ddot{Z}$ are not available, this variable is an omitted confounder, and the effect definition (28) relates to a reduced model.

Substituting $\ddot{Z}$ by a variable $\dot{Z}$ having a distribution, the observable effect is then given by

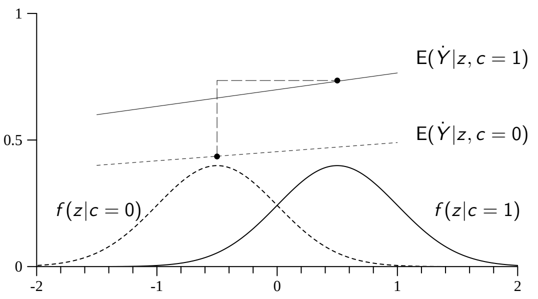
$$\begin{aligned} &\mathrm{E}(\dot{Y}|\dot{C}[0,1], \ddot{X}=x) = \\ &\quad \sum\nolimits_z [\mathrm{E}(\dot{Y}|\dot{C}=1, \ddot{X}=x, \dot{Z}=z)\,\Pr(\dot{Z}=z|\dot{C}=1, \ddot{X}=x) - \\ &\qquad \mathrm{E}(\dot{Y}|\dot{C}=0, \ddot{X}=x, \dot{Z}=z)\,\Pr(\dot{Z}=z|\dot{C}=0, \ddot{X}=x)] \end{aligned} \tag{29}$$

This effect is no longer balanced because

$$\Pr[\dot{Z}|\dot{C}=1, \ddot{X}=x] \neq \Pr[\dot{Z}|\dot{C}=0, \ddot{X}=x]$$

and it is therefore unclear how to think of a causal effect of $\dot{C}$.

9.2 The problem: omitted confounders (cont'd)



9.3 The perspective of evaluative choice models

A further motive for an interest in balanced effect formulations comes from evaluative choice models where an agent is interested in potential effects of the available alternatives.

Consider the evaluative choice model that corresponds to Model 9.1:

Model 9.2

$$\begin{tikzcd} \ddot{X} \arrow[rr] & & \dot{Y}^* \\ \ddot{C} \arrow[urr] & & \ddot{Z} \arrow[u] \end{tikzcd}$$

The agent, $A[\ddot{C}]$, knows the value of $\ddot{X}$, say x^*.

9.3 The perspective of evaluative choice models (cont'd)

Assuming that only the rule (29) is available, expectations are evaluated by comparing

$$\Pr(\dot{Y}^*|\ddot{C}{=}0, \ddot{X}{=}x^*) \quad \text{and} \quad \Pr(\dot{Y}^*|\ddot{C}{=}1, \ddot{X}{=}x^*)$$

As shown by (29), both are averages w.r.t. different distributions of $\dot{Z}$.

However, the agent can assume that there is a particular value of this variable, say z^*, that is identical for both alternatives in the given choice situation.

Even if this value is not known, it would probably preferable to use a balanced effect formulation.

9.4 Constructions of balanced effects

So one should ask, Can balanced effect formulations be constructed? There are three approaches.

a) Randomization w.r.t. all possible confounders. As I have argued in Chapter 5, this approach is problematic in social research because it would change the processes that one aims to investigate.

b) One uses a fixed distribution of the confounding variables. For example, instead of (29) one can consider

$$\sum\nolimits_z \big[\mathrm{E}(\dot{Y}|\dot{C}{=}1, \ddot{X}{=}x, \dot{Z}{=}z)\Pr(\dot{Z}{=}z) - \mathrm{E}(\dot{Y}|\dot{C}{=}0, \ddot{X}{=}x, \dot{Z}{=}z)\Pr(\dot{Z}{=}z\big] \tag{30}$$

where $\Pr[\dot{Z}]$ is an arbitrarily defined distribution (e.g., the distribution of a corresponding statistical variable in a sample).

c) One uses specific kinds of parametric models which allow one to construct balanced effects.

9.5 Parametric assumptions about confounders

I use Heckman's probit selection model (Heckman 1979) to illustrate the parametric approach to the construction of balanced effects.

Note: I discuss this model as a proposal for coping with unobserved confounders ('endogeneity bias').

As proposed by Heckman, the model is also used for 'sample selection problems' in the sense defined in Section 7.1, that is, in situations where observations are only available if $\dot{C}{=}1$.

The basic idea is to assume that omitted confounders can be implicitly taken into account by a joint parametric distribution for the variables $\dot{Y}$ and $\dot{C}$.

9.5 Parametric assumptions about confounders (cont')

To explain the argument, I start from Model 9.1. If $\ddot{Z}$ is observed, one can begin with a linear model

$$\dot{Y} = g_x(x) + g_z(z) + \dot{C}\gamma + \epsilon' \tag{31}$$

where g_x and g_z are deterministic functions of values of $\ddot{X}$ and $\ddot{Z}$, respectively, and ϵ' is a residual random variable.

Assuming that the distribution of ϵ' is independent of $\dot{C}$, one can interpret γ as a balanced effect.

9.5 Parametric assumptions about confounders (cont')

If $\ddot{Z}$ is not observed, one can consider the reduced model that is based on assuming, instead of $\ddot{Z}$, a variable $\dot{Z}$ with some unknown but exogenously given distribution.

Instead of (31), one gets the reduced model

$$\dot{Y} = g_x(x) + \dot{C}\gamma + \epsilon \tag{32}$$

where $\epsilon := g_z(\dot{Z}) + \epsilon'$.

Since $\dot{C}$ depends on $\dot{Z}$, the distribution of ϵ depends on $\dot{C}$, and the effect of $\dot{C}$ is not balanced w.r.t. ϵ. But from (32) one can derive

$$\mathrm{E}(\dot{Y}|x, c) = g_x(x) + c\gamma + \mathrm{E}(\epsilon|x, c)$$

This shows that, in order to estimate γ, one would like to know values of $\mathrm{E}(\epsilon|x, c)$.

These values cannot be observed; but given enough parametric assumptions, they can be constructed.

9.5 Parametric assumptions about confounders (cont')

There are two steps.

In the first step one assumes a model for $\dot{C}$.

This is done by employing a latent variable, η', as follows:

$$\dot{C} = I[\eta' > h_x(x) + h_z(z)] \tag{33}$$

where h_x and h_z are deterministic functions of values of $\ddot{X}$ and $\ddot{Z}$, respectively.

Again, one can define $\eta := -h_z(z) + \eta'$, and rewrite (33) as

$\dot{C} = I[\eta > h_x(x)]$.

9.5 Parametric assumptions about confounders (cont')

In the second step, the distributional assumptions come into play. These concern ϵ', η', and $\dot{Z}$:

a) $\epsilon' \sim \mathcal{N}(0, \sigma^2_{\epsilon'})$
b) $\eta' \sim \mathcal{N}(0, 1)$
c) $g_z(\dot{Z}) \sim \mathcal{N}(\mu_{g_z}, \sigma^2_{g_z})$
d) $h_z(\dot{Z}) \sim \mathcal{N}(\mu_{h_z}, \sigma^2_{h_z})$
e) ϵ' and $g_z(\dot{Z})$ are independent.
f) η' and $h_z(\dot{Z})$ are independent.

These assumptions entail: $\epsilon \sim \mathcal{N}(\mu_{g_z}, \sigma^2_\epsilon)$ with $\sigma^2_\epsilon = \sigma^2_{g_z} + \sigma^2_{\epsilon'}$, and $\eta \sim \mathcal{N}(\mu_{h_z}, \sigma^2_\eta)$ with $\sigma^2_\eta = \sigma^2_{h_z} + 1$.

Moreover, if g_z and h_z are linear functions, the joint distribution of ϵ and η is bivariate normal with a correlation ρ.

9.5 Parametric assumptions about confounders (cont')

From these assumptions and their implications, one can finally derive a method to construct values of $\mathrm{E}(\epsilon|x, c)$:

$$\mathrm{E}(\epsilon|x, c{=}1) = \mathrm{E}(\epsilon|x, \eta > h_x(x)) = \rho\,\sigma_\epsilon \frac{\phi(h_x(x))}{1 - \Phi(h_x(x))}$$

and correspondingly

$$\mathrm{E}(\epsilon|x, c{=}0) = \mathrm{E}(\epsilon|x, \eta \leq h_x(x)) = \rho\,\sigma_\epsilon \frac{\phi(h_x(x))}{1 - \Phi(-h_x(x))}$$

where ϕ and Φ denote, respectively, the density and distribution function of the standard normal distribution. Both can be combined as

$$\lambda(x, c) := c\,\frac{\phi(h_x(x))}{1 - \Phi(h_x(x))} + (1 - c)\,\frac{\phi(h_x(x))}{1 - \Phi(-h_x(x))}$$

9.5 Parametric assumptions about confounders (cont')

It can now be seen how this is an approach to the construction of balanced effects. By defining a new residual variable,

$$\epsilon^* := \epsilon - \rho\,\sigma_\epsilon\,\lambda(x, c)$$

one can rewrite (32) as

$$\dot{Y} = g(x) + \dot{C}\gamma + \rho\,\sigma_\epsilon\,\lambda(x, c) + \epsilon^* \qquad (34)$$

containing a further regressor, $\lambda(x, c)$, and the new residual, ϵ^*. Now one can derive

$$\mathrm{E}(\epsilon^*|x, c) = 0 \quad \text{and} \quad \mathrm{Cov}(\epsilon^*, \dot{C}|x) = 0$$

showing that the effect of $\dot{C}$ is balanced w.r.t. ϵ^*.

9.6 Confounding and mediating variables

It is obvious that Heckman's approach relies on very particular assumptions about the distributions of unobserved variables which are difficult to justify in applications.

The most important assumption concerns the distribution of the unobserved confounder, $\dot{Z}$.

If it is not normally distributed, e.g. if it is a binary variable, the argument depicted in the previous section will not work.

9.6 Confounding and mediating variables (cont')

A further point is noteworthy. Heckman's probit selection model deals with 'endogeneity bias' resulting from omitted variables which show up in correlations between explanatory variables and the residual variable posited in a regression model.

The model cannot, therefore, distinguish between omitted confounders and omitted mediating variables. However, also omitted mediating variables can result in correlations between explanatory and residual variables.

Consider Model 9.1. If the arrow from $\ddot{Z}$ to $\dot{C}$ is reversed, $\ddot{Z}$ changes into an endogenous mediating variable, $\dot{Z}$.

Now, if $\dot{Z}$ is omitted from a regression model for $\dot{Y}$, $\dot{C}$ is again correlated with its residual.

The argument of the previous section will go through without any formal changes. But it would be difficult to think that the model removes an 'endogeneity bias'.

10. Further examples
10.1 Sequential transitions

Here I refer to an example discussed by Buis (2009). The model can be depicted as follows:

Model 10.1

$$\begin{tikzcd}
\ddot{X} \arrow[r] \arrow[d] & \dot{Y}_2 \\
\dot{Y}_1 \arrow[ur] & \ddot{Z} \arrow[l] \arrow[u]
\end{tikzcd}$$

There are two situations. In the first situation ($S = 1$), a process takes place which leads to a value of the binary variable $\dot{Y}_1$. If $\dot{Y}_1 = 1$, there is a transition into a second situation ($S = 2$) in which another process takes place which leads to a value of the binary variable $\dot{Y}_2$ ($\dot{Y}_2 = 1$ can indicate a further transition or any other event). If $\dot{Y}_1 = 0$ then $\dot{Y}_2 = 0$.

$\ddot{X}$ and $\ddot{Z}$ are exogenous explanatory variables, presumably relevant for both transitions.

10.1 Sequential transitions (cont'd)

Buis uses the following numerical example:

	z	x	$Y_1 = 0$	$Y_1 = 1$	cases	$\mathrm{P}(Y_1 = 1\|x, z)$
$S = 1$	0	0	300	100	400	0.25
	0	1	200	200	400	0.50
	1	0	200	200	400	0.50
	1	1	100	300	400	0.75
	z	x	$Y_2 = 0$	$Y_2 = 1$	cases	$\mathrm{P}(Y_2 = 1\|Y_1 = 1, x, z)$
$S = 2$	0	0	75	25	100	0.25
	0	1	100	100	200	0.50
	1	0	100	100	200	0.50
	1	1	75	225	300	0.75

The example is very special: X and Z (the statistical variables corresponding to $\ddot{X}$ and $\ddot{Z}$) are not correlated, and they do not interact in both situations. One should therefore also consider other examples.

10.1 Sequential transitions (cont'd)

If values of Z are not available, the data look as follows:

	x	$Y_1 = 0$	$Y_1 = 1$	cases	$\mathrm{P}(Y_1 = 1\|x, z)$
$S = 1$	0	500	300	800	0.375
	1	300	500	800	0.625
	x	$Y_2 = 0$	$Y_2 = 1$	cases	$\mathrm{P}(Y_2 = 1\|Y_1 = 1, x, z)$
$S = 2$	0	175	125	300	0.417
	1	175	325	500	0.650

10.1 Sequential transitions (cont'd)

With the definition used throughout this presentation, one can distinguish the following effects:

$$\mathrm{E}(\dot{Y}_1|\ddot{X}=1,\ddot{Z}=0) - \mathrm{E}(\dot{Y}_1|\ddot{X}=0,\ddot{Z}=0) = 0.50 - 0.25 = 0.25$$

$$\mathrm{E}(\dot{Y}_1|\ddot{X}=1,\ddot{Z}=1) - \mathrm{E}(\dot{Y}_1|\ddot{X}=0,\ddot{Z}=1) = 0.75 - 0.50 = 0.25$$

$$\mathrm{E}(\dot{Y}_2|\dot{Y}_1=1,\ddot{X}=1,\ddot{Z}=0) - \mathrm{E}(\dot{Y}_2|\dot{Y}_1=1,\ddot{X}=0,\ddot{Z}=0) = 0.50 - 0.25 = 0.25$$
$$\mathrm{E}(\dot{Y}_2|\dot{Y}_1=1,\ddot{X}=1,\ddot{Z}=1) - \mathrm{E}(\dot{Y}_2|\dot{Y}_1=1,\ddot{X}=0,\ddot{Z}=1) = 0.75 - 0.50 = 0.25$$

Instead, Buis uses odds ratios:

$$\frac{\frac{\Pr(\dot{Y}_1=1\mid X=1,Z=0)}{\Pr(\dot{Y}_1=0\mid X=1,Z=0)}}{\frac{\Pr(\dot{Y}_1=1\mid X=0,Z=0)}{\Pr(\dot{Y}_1=0\mid X=0,Z=0)}} = \frac{\frac{0.5}{0.5}}{\frac{0.25}{0.75}} = 3, \quad \frac{\frac{\Pr(\dot{Y}_1=1\mid X=1,Z=1)}{\Pr(\dot{Y}_1=0\mid X=1,Z=1)}}{\frac{\Pr(\dot{Y}_1=1\mid X=0,Z=1)}{\Pr(\dot{Y}_1=0\mid X=0,Z=1)}} = \frac{\frac{0.75}{0.25}}{\frac{0.5}{0.5}} = 3$$

$$\frac{\frac{\Pr(\dot{Y}_2=1\mid \dot{Y}_1=1,X=1,Z=0)}{\Pr(\dot{Y}_2=0\mid \dot{Y}_1=1,X=1,Z=0)}}{\frac{\Pr(\dot{Y}_2=1\mid \dot{Y}_1=1,X=0,Z=0)}{\Pr(\dot{Y}_2=0\mid \dot{Y}_1=1,X=0,Z=0)}} = \frac{\frac{0.5}{0.5}}{\frac{0.25}{0.75}} = 3, \quad \frac{\frac{\Pr(\dot{Y}_2=1\mid \dot{Y}_1=1,X=1,Z=1)}{\Pr(\dot{Y}_2=0\mid \dot{Y}_1=1,X=1,Z=1)}}{\frac{\Pr(\dot{Y}_2=1\mid \dot{Y}_1=1,X=0,Z=1)}{\Pr(\dot{Y}_2=0\mid \dot{Y}_1=1X=0,Z=1)}} = \frac{\frac{0.75}{0.25}}{\frac{0.5}{0.5}} = 3$$

10.1 Sequential transitions (cont'd)

Now, what happens if $\ddot{Z}$ is not observed? I first consider the situation $S=1$.

Buis says that one 'underestimates' the effect of $\ddot{X}$. This is true if one uses odds ratios:

$$\frac{\frac{\Pr(\dot{Y}_1=1\mid X=1)}{\Pr(\dot{Y}_1=0\mid X=1)}}{\frac{\Pr(\dot{Y}_1=1\mid X=0)}{\Pr(\dot{Y}_1=0\mid X=0)}} = \frac{\frac{0.625}{0.375}}{\frac{0.375}{0.625}} = 2.778$$

It is not true, however, when using our effect definition:

$$\mathrm{E}(\dot{Y}_1|\ddot{X}=1) - \mathrm{E}(\dot{Y}_1|\ddot{X}=1) = 0.625 - 0.375 = 0.25$$

This shows the importance of the effect definition used.

10.1 Sequential transitions (cont'd)

In any case, without data on $\ddot{Z}$ one cannot estimate the conditional effects defined above, but only the 'marginal' effect

$$\mathrm{E}(\dot{Y}_1|\ddot{X}=1) - \mathrm{E}(\dot{Y}_1|\ddot{X}=0) \tag{35}$$

In Buis' example, this effect is equal to the two conditional effects (0.25). But this is simply a consequence of his very special data set. In general, this effect can be quite different.

In order to get an understanding of the effect (35) one has to consider a reduced model in which the not observed variable is eliminated. This requires to substitute $\ddot{Z}$ by a random variable $\dot{Z}$ and to assume a distribution for this variable.

In general, this distribution can well depend on values of $\ddot{X}$, and I therefore use a conditional distribution denoted by $\Pr(\dot{Z}=z|\ddot{X}=x)$.

10.1 Sequential transitions (cont'd)

Assuming as given the conditional distribution $\Pr(\dot{Z}=z|\ddot{X}=x)$, one can derive

$$\mathrm{E}(\dot{Y}_1|\ddot{X}=x) = \sum_z \mathrm{E}(\dot{Y}_1|\ddot{X}=x,\dot{Z}=z)\Pr(\dot{Z}=z|\ddot{X}=x) \tag{36}$$

This shows that, in general, (35) is not a balanced effect, meaning that the distribution of $\dot{Z}$ is different for the values of $\ddot{X}$ to be compared.

The effect (35) can nevertheless be interpreted by assuming

$$\text{The distribution of } \dot{Z} \text{ depends on } \ddot{X} \tag{37}$$

On can then use Model 10.2:

$$\begin{tikzcd}
\ddot{X} \arrow[r] \arrow[d] & \dot{Y}_1 \\
\dot{Z} \arrow[ur] &
\end{tikzcd}$$

This allows one to interpret (35) as the total effect of $\ddot{X}$ on $\dot{Y}_1$.

10.1 Sequential transitions (cont'd)

This interpretation is always possible in a purely formal sense.

Another question is whether the total effect can be given a causal interpretation. This is only possible if one can justify that $\dot{Z}$ depends on $\ddot{X}$ in a causal sense (or is independent of $\ddot{X}$).

If, on the contrary, values of $\ddot{X}$ causally depend on $\dot{Z}$, then any interpretation of the total effect is highly questionable. In fact, $\dot{Z}$ is then a confounder w.r.t. the relationship between $\ddot{X}$ and $\dot{Y}_1$, and the effect (35) has no clear meaning.

Obviously, in order to justify an interpretation of the total effect, one must refer to the meaning of $\dot{Z}$. If one simply assumes some unspecified 'unobserved heterogeneity' it seems not possible to give the effect (35) any clear meaning.

10.1 Sequential transitions (cont'd)

Given that the total effect has a sensible interpretation, one can further ask whether one can conceptually separate a direct and an indirect effect of $\ddot{X}$ on $\dot{Y}_1$.

This is possible if $\ddot{X}$ and $\dot{Z}$ do not interact w.r.t. $\dot{Y}_1$ as it is the case in Buis' example. In this example, both conditional effects are identical and can therefore be identified with the direct effect.

In general, one has to assume that $\ddot{X}$ and $\dot{Z}$ interact, and then a unique effect of $\ddot{X}$ on $\dot{Y}_1$ cannot be defined and could not be estimated even if data on $\dot{Z}$ were available.

In any case, it seems obscure to think of the total effect as being in some sense 'biased'.

10.1 Sequential transitions (cont'd)

I now consider the second situation $S=2$. Without data on $\ddot{Z}$, one gets the odds ratio

$$\frac{\frac{\Pr(\dot{Y}_2=1\mid \dot{Y}_1=1,X=1)}{\Pr(\dot{Y}_2=0\mid \dot{Y}_1=1,X=1)}}{\frac{\Pr(\dot{Y}_2=1\mid \dot{Y}_1=1,X=0)}{\Pr(\dot{Y}_2=0\mid \dot{Y}_1=1,X=0)}} = \frac{\frac{0.650}{0.350}}{\frac{0.417}{0.583}} = 2.596$$

With our effect definition one gets:

$$\mathrm{E}(\dot{Y}_2|\dot{Y}_1=1,\ddot{X}=1) - \mathrm{E}(\dot{Y}_2|\dot{Y}_1=1,\ddot{X}=0) = 0.650 - 0.417 = 0.233$$

In the following, I consider this 'marginal' effect:

$$\mathrm{E}(\dot{Y}_2|\dot{Y}_1=1,\ddot{X}=1) - \mathrm{E}(\dot{Y}_2|\dot{Y}_1=1,\ddot{X}=0) \tag{38}$$

My interpretation will parallel the one given above for the first situation.

One has to take into account, however, that the conditioning on $\dot{Y}_1=1$ (which is a 'collider' in Model 10.1) changes the relationship between the statistical variables corresponding to $\ddot{X}$ and $\ddot{Z}$.

10.1 Sequential transitions (cont'd)

Note that, in general, conditioning on $\dot{Y}_1=1$ does not *create* a correlation. This is only true in Buis' example in which X and Z (the statistical variables corresponding to $\ddot{X}$ and $\ddot{Y}$) are uncorrelated in the first situation.

In general, in social research, explanatory variables are always more or less correlated. And one should say, then, that conditioning on $\dot{Y}_1$ will *change* an already existent correlation.

In fact, it is also possible that, due to the conditioning, correlated variables become uncorrelated. This is shown by the following example.

10.1 Sequential transitions (cont'd)

	z	x	$Y_1 = 0$	$Y_1 = 1$	cases	$\mathrm{P}(Y_1 = 1 \mid x, z)$
$S = 1$	0	0	300	100	400	0.25
	0	1	150	150	300	0.50
	1	0	200	200	400	0.50
	1	1	100	300	400	0.75
	z	x	$Y_2 = 0$	$Y_2 = 1$	cases	$\mathrm{P}(Y_1 = 2 \mid Y_1 = 1, x, z)$
$S = 2$	0	0	75	25	100	0.25
	0	1	75	75	150	0.50
	1	0	100	100	200	0.50
	1	1	75	225	300	0.75

In the first situation:
$\mathrm{P}(Z = 1 \mid X = 0) = 400/800$, $\mathrm{P}(Z = 1 \mid X = 1) = 400/700$
In the second situation: $\mathrm{P}(Z = 1 \mid Y_1 = 1, X = 0) = 200/300$
$\mathrm{P}(Z = 1 \mid Y_1 = 1, X = 1) = 300/450$

10.1 Sequential transitions (cont'd)

A second preliminary remark: That conditioning on $\dot{Y}_1 = 1$ leads to a change in the correlation of X and Z is irrelevant for the conditional effects

$$\mathrm{E}(\dot{Y}_2 \mid \dot{Y}_1 = 1, \ddot{X} = 1, \ddot{Z} = z) - \mathrm{E}(\dot{Y}_2 \mid \dot{Y}_1 = 1, \ddot{X} = 0, \ddot{Z} = z) \tag{39}$$

It becomes relevant, however, if Z is not observed and one can only estimate the 'marginal' effect (38).

In order to interpret this effect, one has to consider the reduced model (as was done above for the first situation). In the reduced model:

$$\mathrm{E}(\dot{Y}_2 \mid \dot{Y}_1 = 1, \ddot{X} = x) = \sum_z \mathrm{E}(\dot{Y}_2 \mid \dot{Y}_1 = 1, \ddot{X} = x, \dot{Z} = z)\,\Pr(\dot{Z} = z \mid \dot{Y}_1 = 1, \ddot{X} = x) \tag{40}$$

showing that, in general, also the 'marginal' effect (38) is not balanced.

10.1 Sequential transitions (cont'd)

Moreover, in the second situation, the distribution of $\dot{Z}$ not only depends on $\ddot{X}$ in the first situation (if this is assumed), but in any case also on $\ddot{X}$ via the outcome of the first situation.

This can be made explicit by formally deriving

$$\Pr(\dot{Z} = z \mid \dot{Y}_1 = 1, \ddot{X} = x) = \frac{\mathrm{E}(\dot{Y}_1 \mid \ddot{X} = x, \dot{Z} = z)\,\Pr(\dot{Z} = z \mid \ddot{X} = x)}{\sum_{z'} \mathrm{E}(\dot{Y}_1 \mid \ddot{X} = x, \dot{Z} = z')\,\Pr(\dot{Z} = z' \mid \ddot{X} = x)} \tag{41}$$

Again, the important question is whether the 'marginal' effect (38) can be interpreted as a total effect, now: a total effect of $\ddot{X}$ *and* $\dot{Y}_1 = 1$.

I propose that this is possible if, in the first situation, $\dot{Z}$ either is independent of $\ddot{X}$ or can sensibly be interpreted as being dependent on $\ddot{X}$.

10.1 Sequential transitions (cont'd)

Given the stated assumption, one can interpret the right-hand side of (41) as follows:

a) The $\Pr(\dot{Z} = z \mid \ddot{X} = x)$ terms directly represent a dependence of the distribution of $\dot{Z}$ on $\ddot{X}$.
b) The $\mathrm{E}(\dot{Y}_1 \mid \ddot{X} = x, \dot{Z} = z)$ terms represent the institutional framework which is presupposed by the Model 10.1.

One therefore can use for $S = 2$ the following Model 10.3:

$$\begin{tikzcd} \ddot{X} \arrow[r] \arrow[d] & \dot{Y}_2 \\ \dot{Z} \arrow[ur] & \end{tikzcd}$$

And this allows one to interpret (38) as the total effect of $\ddot{X}$ on $\dot{Y}_2$.

10.1 Sequential transitions (cont'd)

Given that the total effect has a sensible interpretation, one can again ask, now for the situation $S = 2$, whether one can conceptually separate a direct and an indirect effect of $\ddot{X}$ on $\dot{Y}_2$.

This is possible if $\ddot{X}$ and $\dot{Z}$ do not interact w.r.t. $\dot{Y}_2$ as it is the case in Buis' example.

Of course, without data on Z, one cannot actually distinguish the direct and the indirect effect.

But knowing Buis's data, these effects are already known. The total effect is 0.233, the direct effect is 0.25, and the indirect effect is -0.017.

In general, one has to assume that $\ddot{X}$ and $\dot{Z}$ interact, and then a unique effect of $\ddot{X}$ on $\dot{Y}_1$ cannot be defined and could not be estimated even if data on $\dot{Z}$ were available.

In any case, it seems obscure to think of the total effect as being in some sense 'biased'.

10.1 Sequential transitions (cont'd)

Conclusion: If one can cope with an unobserved variable in the first situation, by interpreting the estimable 'marginal' effect as a total effect covering the unobserved variable, then this is also possible in the second situation.

Of course, the total effect in $S = 2$ cannot be attributed solely to $\ddot{X}$. It also depends on the distribution of $\dot{Z}$ in $S = 1$, possibly depending on $\ddot{X}$; and furthermore on the institutional framework given by the conditional distributions of $\dot{Y}_1$ and $\dot{Y}_2$ as presupposed by a model that explicitly represents the unobserved variable.

It seems not possible to consider the total effect as a 'biased' estimate of a well-defined effect. In general, one also has no reason for believing that X and Z are not correlated in $S = 1$ and do not interact in $S = 1$ as well as in $S = 2$.

10.2 Education and divorce

I now consider another example: How does the risk of divorce depends on the women's level of education?

Researchers often found that education has a positive impact on the risk of divorce, but there also are other findings.

In a recent study, Bernardi and Martinez-Pastor (2011) discuss the hypothesis that the observed relationships between education and risk of divorce might result from 'selection effects'.

However, so formulated, the hypothesis can create confusion.

10.2 Selection and omitted confounders (cont'd)

To see this, one can use again Model 7.2, reproduced here:

Model 10.4

$$\begin{tikzcd} \ddot{X} \arrow[r] \arrow[d] & \dot{Y} \\ \dot{S} \arrow[ur] & \ddot{Z} \arrow[l] \arrow[u] \end{tikzcd}$$

with the following interpretation. $\dot{S} = 1$ if a woman is married, and $\dot{S} = 0$ otherwise. $\dot{Y}$ is an indicator variable for becoming divorced; $\ddot{X}$ records the level of education, and $\ddot{Z}$ is an unobserved confounder.

Obviously, without being married there can be no risk of divorce.

Consequently, $\dot{S} = 1$ is also a necessary precondition for the variables', $\ddot{X}$ and $\ddot{Z}$, having an impact on the risk of divorce.

10.2 Selection and omitted confounders (cont'd)

Is there a 'selection effect'?

One can assume that distributions of values of $\ddot{X}$ and $\dot{Z}$ exist for married and unmarried women in some specified population.

So one can think that marriage changes these distributions and their correlation, and consider this as a 'selection effect'.

But this selection effect cannot change relationships between these variables and the risk of divorce; simply because these relationships are only defined conditional on $\dot{S}=1$.

10.2 Selection and omitted confounders (cont'd)

However, an important part of the research question concerns the observation of historically changing relationships between women's education and the risk of divorce.

One has then to consider at least two periods, say

$$(\ddot{X}_1, \dot{Z}_1, \dot{S}_1, \dot{Y}_1) \longrightarrow (\ddot{X}_2, \dot{Z}_2, \dot{S}_2, \dot{Y}_2) \tag{42}$$

$\dot{Z}_t$ represents the distribution of the unobserved confounder in period t.

10.2 Selection and omitted confounders (cont'd)

This allows one to think of the observed relationships in the following way:

$$\begin{aligned} &\mathrm{E}(\dot{Y}_t|\ddot{X}_t=x, \dot{S}_t=1) = \\ &\quad \sum\nolimits_z \mathrm{E}(\dot{Y}_t|\ddot{X}_t=x, \dot{Z}_t=z, \dot{S}_t=1)\,\Pr(\dot{Z}_t=z|\ddot{X}_t=x, \dot{S}_t=1) \end{aligned} \tag{43}$$

So it is quite possible that differences between the observed relationships can be due to both

a) changes in the conditional expectation $\mathrm{E}(\dot{Y}_t|\ddot{X}_t=x, \dot{Z}_t=z, \dot{S}_t=1)$ which, presumably, has a causal interpretation, and

b) changes in the conditional distributions of the unobserved confounder, $\Pr[\dot{Z}_t|\ddot{X}_t=x, \dot{S}_t=1]$.

Most probably, they are due to both kinds of changes so that one would like to learn about the quantitative relevance of unobserved confounders.

But this would require to observe the confounder.

10.2 Selection and omitted confounders (cont'd)

The question remains whether one can think of (b) as a 'selection effect'. Obviously not in the static sense referred to above. One would need to consider the function

$$(x, z) \longrightarrow \Pr(\dot{S}_t=1|\ddot{X}_t=x, \dot{Z}_t=z) \tag{44}$$

Differences in these functions would describe a historically changing selection into marriage. However, the changes referred to in (b) do not only result from a change in the function (44). As seen from

$$\begin{aligned} &\Pr(\dot{Z}_t=z|\ddot{X}_t=x, \dot{S}_t=1) = \\ &\quad \frac{\Pr(\dot{S}_t=1|\ddot{X}_t=x, \dot{Z}_t=z)\,\Pr(\dot{Z}_t=z|\ddot{X}_t=x)}{\Pr(\dot{S}_t=1|\ddot{X}_t=x)} \end{aligned} \tag{45}$$

they can also result from a change in the conditional distributions of the unobserved confounder.

10.2 Selection and omitted confounders (cont'd)

Consequently, without observing the confounder (whose supposed existence motivates the discussion) one cannot draw any clear conclusions.

On the other hand, if one could observe $\dot{Z}_t$, one could immediately consider the relationship

$$(x, z) \longrightarrow \mathrm{E}(\dot{Y}_t|\ddot{X}_t=x, \dot{Z}_t=z, \dot{S}_t=1) \tag{46}$$

and recognize that the risk of divorce not only depends on the women's level of education, but also on another identifiable variable, $\dot{Z}$.

References

Bernardi, F., Martinez-Pastor, J.-I. (2011). Female Education and Marriage Dissolution: Is it a Selection Effect? *European Sociological Review* 27, 693–707.

Blossfeld, H.-P. (2009). Causation as a Generative Process. In: H. Engelhardt, H.-P. Kohler, A. Prskawetz (eds.), *Causal Analysis in Population Studies*, 83–109. Berlin: Springer.

Buis, M. L. (2009). The Consequences of Unobserved Heterogeneity in a Sequential Logit Model.

Goldthorpe, J. H. (2001). Causation, Statistics, and Sociology. *European Sociological Review* 17, 1–20.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica* 47, 153–161.

Rohwer, G. (2010). *Models in Statistical Social Research*. London: Routledge.

Rohwer, G. (2012). Estimating Effects with Logit Models. *NEPS Working Paper* No. 10. Bamberg: Otto-Friedrich Universität, Nationales Bildungspanel.

Rohwer, G. (2014). Factual and Modal Notions in Social Research. *Quality & Quantity* 48, 547–561.