Probabilistic and Causal Knowledge in Econometric Models

Alessio Moneta

Max Planck Institute of Economics
Jena, Germany

Canterbury, 10 September 2008

Causality and Probability in the Sciences

Motivations

Is causal induction possible in economics (non-experimental setting)?

For causal inference in econometrics one needs:

- *some* background causal knowledge / assumptions
- statistical description of the data (statistical model)

Agreements and disagreements

Distinction between causal and statistical requirements (Hoover 2007)

Questions I want to face:

- Is statistical and causal information separable? Should it be?
- Is statistical knowledge causality-laden?

Motivations

Is causal induction possible in economics (non-experimental setting)?

For causal inference in econometrics one needs:

- *some* background causal knowledge / assumptions
- statistical description of the data (statistical model)

Agreements and disagreements

Distinction between causal and statistical requirements (Hoover 2007)

Questions I want to face:

- Is statistical and causal information separable? Should it be?
- Is statistical knowledge causality-laden?

Motivations

Is causal induction possible in economics (non-experimental setting)?

For causal inference in econometrics one needs:

- *some* background causal knowledge / assumptions
- statistical description of the data (statistical model)

Agreements and disagreements

Distinction between causal and statistical requirements (Hoover 2007)

Questions I want to face:

- Is statistical and causal information separable? Should it be?
- Is statistical knowledge causality-laden?

Motivations

Is causal induction possible in economics (non-experimental setting)?

For causal inference in econometrics one needs:

- *some* background causal knowledge / assumptions
- statistical description of the data (statistical model)

Agreements and disagreements

Distinction between causal and statistical requirements (Hoover 2007)

Questions I want to face:

- Is statistical and causal information separable? Should it be?
- Is statistical knowledge causality-laden?

Motivations

Is causal induction possible in economics (non-experimental setting)?

For causal inference in econometrics one needs:

- *some* background causal knowledge / assumptions
- statistical description of the data (statistical model)

Agreements and disagreements

Distinction between causal and statistical requirements (Hoover 2007)

Questions I want to face:

- Is statistical and causal information separable? Should it be?
- Is statistical knowledge causality-laden?

Motivations

Is causal induction possible in economics (non-experimental setting)?

For causal inference in econometrics one needs:

- *some* background causal knowledge / assumptions
- statistical description of the data (statistical model)

Agreements and disagreements

Distinction between causal and statistical requirements (Hoover 2007)

Questions I want to face:

- Is statistical and causal information separable? Should it be?
- Is statistical knowledge causality-laden?

Overview

- statistical model
- statistical (in-)dependence
- in which sense statistical information is causality laden
- in which sense it is not
- two important pieces of information:
  - functional form
  - level of aggregation

Overview

- statistical model
- statistical (in-)dependence
- in which sense statistical information is causality laden
- in which sense it is not
- two important pieces of information:
  - functional form
  - level of aggregation

Overview

- statistical model
- statistical (in-)dependence
- in which sense statistical information is causality laden
- in which sense it is not
- two important pieces of information:
  - functional form
  - level of aggregation

Overview

- statistical model
- statistical (in-)dependence
- in which sense statistical information is causality laden
- in which sense it is not
- two important pieces of information:
  - functional form
  - level of aggregation

Overview

- statistical model
- statistical (in-)dependence
- in which sense statistical information is causality laden
- in which sense it is not
- two important pieces of information:
  - functional form
  - level of aggregation

Overview

- ▷ statistical model
- ▷ statistical (in-)dependence
- ▷ in which sense statistical information is causality laden
- ▷ in which sense it is not
- ▷ two important pieces of information:
  - functional form
  - level of aggregation

Overview

- ▷ statistical model
- ▷ statistical (in-)dependence
- ▷ in which sense statistical information is causality laden
- ▷ in which sense it is not
- ▷ two important pieces of information:
  - functional form
  - level of aggregation

Statistical models

Statistical models (see Spanos 1999):

- ▷ probability model:
  - density function of random variables
  - space of parameters (in the parametric case)
  - support of the density

  e.g. $\Phi_\theta = \{f(x;\theta), \theta \in \Theta, x \in \mathcal{R}_x\}$
- ▷ sampling model:

  for example: $(X_1, X_2, \ldots, X_n)$ is a random sample.

Statistical models

Statistical models (see Spanos 1999):

- ▷ probability model:
  - density function of random variables
  - space of parameters (in the parametric case)
  - support of the density

  e.g. $\Phi_\theta = \{f(x;\theta), \theta \in \Theta, x \in \mathcal{R}_x\}$

- ▷ sampling model:

  for example: $(X_1, X_2, \ldots, X_n)$ is a random sample.

Statistical models

Statistical models (see Spanos 1999):

- ▷ probability model:
  - density function of random variables
  - space of parameters (in the parametric case)
  - support of the density

  e.g. $\Phi_\theta = \{f(x;\theta), \theta \in \Theta, x \in \mathcal{R}_x\}$
- ▷ sampling model:

  for example: $(X_1, X_2, \ldots, X_n)$ is a random sample.

Useful statistical information

Statistical information *useful for causal inference*:

- stochastic generating mechanism (GM) (Spanos 1999)

$$Y = E(Y|\mathcal{D}) + u \tag{1}$$

satisfying: $E(u \cdot E(Y|\mathcal{D})) = 0; \quad E(u|\mathcal{D}) = 0; \quad E(u^2|\mathcal{D}) = Var(Y|\mathcal{D}) < \infty$

  - ▷ regression function decomposition
  - ▷ structural equation models (in the Haavelmo's tradition)

- conditional independence (CI) tests
  - ▷ $H_0: \quad X \perp\!\!\!\perp Y|Z \quad \equiv \quad f(x|y,z) = f(x|z)$
  - ▷ graphical causal inference based on conditional independence tests and on inferential rules like Markov and Faithfulness conditions (cfr. Pearl 2000, Spirtes-Glymour-Scheines 2000)

Useful statistical information

Statistical information *useful for causal inference*:

- stochastic generating mechanism (GM) (Spanos 1999)

$$Y = E(Y|\mathcal{D}) + u \tag{1}$$

satisfying: $E(u \cdot E(Y|\mathcal{D})) = 0$; $E(u|\mathcal{D}) = 0$; $E(u^2|\mathcal{D}) = Var(Y|\mathcal{D}) < \infty$

  - ▷ regression function decomposition
  - ▷ structural equation models (in the Haavelmo's tradition)

- conditional independence (CI) tests
  - ▷ $H_0: \quad X \perp\!\!\!\perp Y|Z \quad \equiv \quad f(x|y,z) = f(x|z)$
  - ▷ graphical causal inference based on conditional independence tests and on inferential rules like Markov and Faithfulness conditions (cfr. Pearl 2000, Spirtes-Glymour-Scheines 2000)

Useful statistical information

Statistical information *useful for causal inference*:

- stochastic generating mechanism (GM) (Spanos 1999)

$$Y = E(Y|\mathcal{D}) + u \tag{1}$$

satisfying: $E(u \cdot E(Y|\mathcal{D})) = 0; \quad E(u|\mathcal{D}) = 0; \quad E(u^2|\mathcal{D}) = Var(Y|\mathcal{D}) < \infty$

  - ▷ regression function decomposition
  - ▷ structural equation models (in the Haavelmo's tradition)

- conditional independence (CI) tests
  - ▷ $H_0: \quad X \perp\!\!\!\perp Y|Z \quad \equiv \quad f(x|y,z) = f(x|z)$
  - ▷ graphical causal inference based on conditional independence tests and on inferential rules like Markov and Faithfulness conditions (cfr. Pearl 2000, Spirtes-Glymour-Scheines 2000)

Caveat: not *any* stochastic GM or any test of CI conveys useful information!

example:

$$Y_{it} = f(X_{it}, Z_{it}) + \epsilon_t \quad (2)$$

Y_{it}: expenditure on some goods g (e.g. food);
X_{it}: income;
Z_{it}: family characteristics

we may be led astray if:

- the conditions for which the data are obtained change at each time point;
- instability;
- if we aggregate heterogenous families.

Caveat: not *any* stochastic GM or any test of CI conveys useful information!

example:

$$Y_{it} = f(X_{it}, Z_{it}) + \epsilon_t \tag{2}$$

Y_{it}: expenditure on some goods g (e.g. food);
X_{it}: income;
Z_{it}: family characteristics

we may be led astray if:

- the conditions for which the data are obtained change at each time point;
- instability;
- if we aggregate heterogenous families.

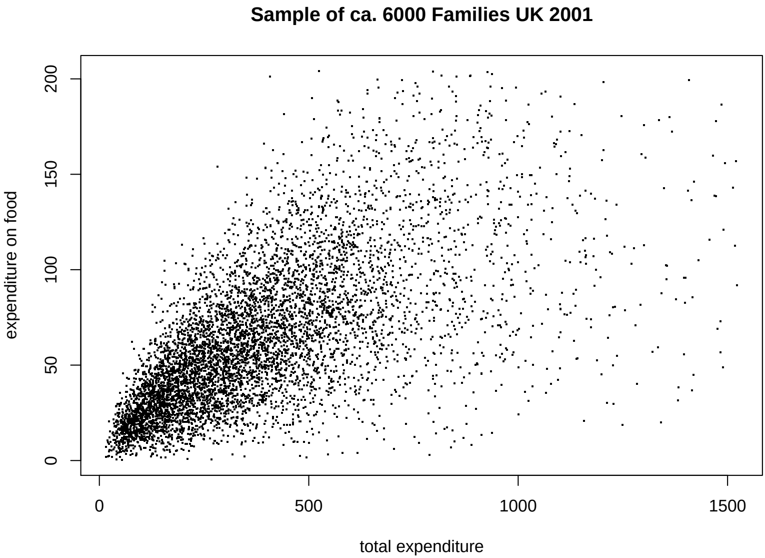
Chance set-ups

A statistical model, to be linked to the world, must represent a chance set-up that gave rise to the observed data (see Cartwright 1999).

Chance set-ups (Hacking 1965):

- device (not necessarily physical) natural / artificial
- generation of trials / observations
- each trial has a result which is a member of class of possible results

phenomena of *chance regularity*


Sample of ca. 6000 Families UK 2001
expenditure on food
0
50
100
150
200
0
500
1000
1500
total expenditure

Chance set-ups

Are there causal presuppositions in thinking about chance set-ups?
in some sense yes:

- cfr. *propensities*

but...

- we do need to have a complete knoweldge of the chance set-up
- chance set-ups with opposite causal relations may be described by the same statistical model

Causal direction

Testing conditional independence does not *in general* require causal presuppositions about:

- presence of cycles, feedbacks
- causal sufficieny (latent variables)

Why? Because statistical tests of independence do not invoke causal notions (properties about directions and sufficiency).

$$X \perp\!\!\!\perp Y|Z \quad \equiv \quad f(x|y,z) = f(x|z)$$

$\equiv$ knowing realizations of Z, information about realization of X conveys useful information about realization of$Y$

Functional form

Functional form of causal dependence

Specification of the functional form of the regression function

Linearity assumptions.
e.g. $Y_t = \beta_0 + \beta_1 x_t + u_t, \quad u_t \sim NIID(0, \sigma^2)$

Correlation coefficient ρ: measure of linear dependence.

Is it possible to test conditional independence without assumptions on the functional form?

Yes: nonparametric tests of conditional independence.

$H_0: \quad X \perp\!\!\!\perp Y|Z \quad \equiv \quad f(x|y,z) = f(x|z)$

But: *curse of dimensionality.*

Functional form

Functional form of causal dependence

Specification of the functional form of the regression function

Linearity assumptions.
e.g. $Y_t = \beta_0 + \beta_1 x_t + u_t, \;\; u_t \sim NIID(0, \sigma^2)$

Correlation coefficient ρ: measure of linear dependence.

Is it possible to test conditional independence without assumptions on the functional form?

Yes: nonparametric tests of conditional independence.

$H_0: \quad X \perp\!\!\!\perp Y|Z \quad \equiv \quad f(x|y,z) = f(x|z)$

But: *curse of dimensionality.*

Functional form

Functional form of causal dependence

Specification of the functional form of the regression function

Linearity assumptions.
e.g. $Y_t = \beta_0 + \beta_1 x_t + u_t, \quad u_t \sim NIID(0, \sigma^2)$

Correlation coefficient ρ: measure of linear dependence.

Is it possible to test conditional independence without assumptions on the functional form?

Yes: nonparametric tests of conditional independence.

$H_0: \quad X \perp\!\!\!\perp Y|Z \quad \equiv \quad f(x|y,z) = f(x|z)$

But: *curse of dimensionality.*

Functional form

Functional form of causal dependence

Specification of the functional form of the regression function

Linearity assumptions.
e.g. $Y_t = \beta_0 + \beta_1 x_t + u_t, \;\; u_t \sim NIID(0, \sigma^2)$

Correlation coefficient ρ: measure of linear dependence.

Is it possible to test conditional independence without assumptions on the functional form?

Yes: nonparametric tests of conditional independence.

$$H_0: \quad X \perp\!\!\!\perp Y|Z \quad \equiv \quad f(x|y,z) = f(x|z)$$

But: *curse of dimensionality.*

Functional form

Functional form of causal dependence

Specification of the functional form of the regression function

Linearity assumptions.
e.g. $Y_t = \beta_0 + \beta_1 x_t + u_t, \quad u_t \sim NIID(0, \sigma^2)$

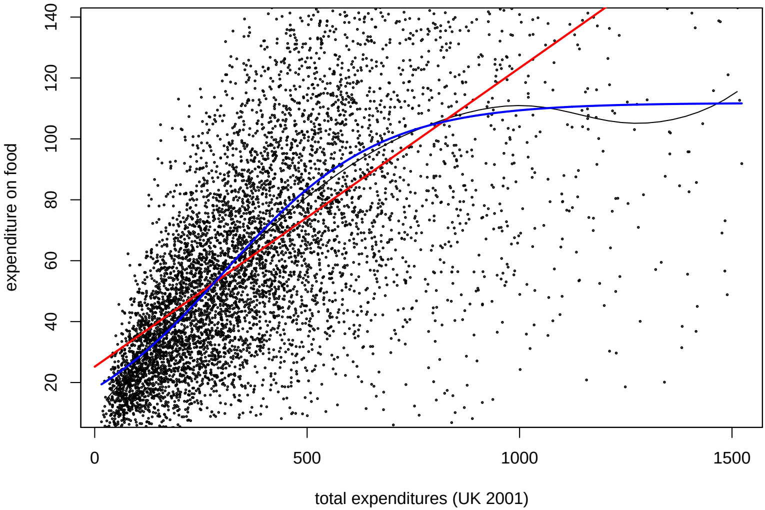
Correlation coefficient ρ: measure of linear dependence.

Is it possible to test conditional independence without assumptions on the functional form?

Yes: nonparametric tests of conditional independence.

$H_0: \quad X \perp\!\!\!\perp Y|Z \quad \equiv \quad f(x|y,z) = f(x|z)$

But: *curse of dimensionality*.


expenditure on food
140
120
100
80
60
40
20
0
500
1000
1500
total expenditures (UK 2001)

Aggregation

Choice of the variables can be partially theory or *causality-laden*.

Mixing populations:

- Simpson paradox: correlations in a sample vanish in subsamples;

Problem of heterogeneity / aggregation:

- Example: suppose you model the dependence between consumption on a particular commodity (e.g. food) and income. How is this dependence across sub-groups? sub categories of income/commodities?
- a form of dependence at micro level may be different at macro level

Aggregation

Choice of the variables can be partially theory or *causality-laden*.

Mixing populations:

- Simpson paradox: correlations in a sample vanish in subsamples;

Problem of heterogeneity / aggregation:

- Example: suppose you model the dependence between consumption on a particular commodity (e.g. food) and income. How is this dependence across sub-groups? sub categories of income/commodities?
- a form of dependence at micro level may be different at macro level

Aggregation

Choice of the variables can be partially theory or *causality-laden*.

Mixing populations:

- Simpson paradox: correlations in a sample vanish in subsamples;

Problem of heterogeneity / aggregation:

- Example: suppose you model the dependence between consumption on a particular commodity (e.g. food) and income. How is this dependence across sub-groups? sub categories of income/commodities?
- a form of dependence at micro level may be different at macro level

Conclusions

Statistical evidence for causal inference is not immediate: it needs an inferential step via the statistical model, separated from the causal (structural) model.

To what extent is the statistical evidence causality laden?

- it involves a chance set-up which is only partially causality-laden
- assumptions on causal order are not necessary to build a statistical model (although they may be useful)
- assumptions on the functional form: useful information for specification (in principle not necessary)
- level of aggregation

Conclusions

Statistical evidence for causal inference is not immediate: it needs an inferential step via the statistical model, separated from the causal (structural) model.

To what extent is the statistical evidence causality laden?

- it involves a chance set-up which is only partially causality-laden
- assumptions on causal order are not necessary to build a statistical model (although they may be useful)
- assumptions on the functional form: useful information for specification (in principle not necessary)
- level of aggregation

Conclusions

Statistical evidence for causal inference is not immediate: it needs an inferential step via the statistical model, separated from the causal (structural) model.

To what extent is the statistical evidence causality laden?

- it involves a chance set-up which is only partially causality-laden
- assumptions on causal order are not necessary to build a statistical model (although they may be useful)
- assumptions on the functional form: useful information for specification (in principle not necessary)
- level of aggregation

Conclusions

Statistical evidence for causal inference is not immediate: it needs an inferential step via the statistical model, separated from the causal (structural) model.

To what extent is the statistical evidence causality laden?

- it involves a chance set-up which is only partially causality-laden
- assumptions on causal order are not necessary to build a statistical model (although they may be useful)
- assumptions on the functional form: useful information for specification (in principle not necessary)
- level of aggregation