Causal Inference in Empirical Economics: How to Cope with Complexity?

Alessio Moneta

Max Planck Institute of Economics
Jena, Germany

25 August 2011

MAX-PLANCK-GESELLSCHAFT

Altonaer Stiftung für philosophische Grundlagenforschung

Introduction

- How is possible learn *valid and reliable* causal relationships:
  - based on observations (empirical economics)
  - taking into account the potential influence of many interacting variables.
- Under these conditions causal inference is difficult. Possible and often practiced strategy:
  - give up empirical causal analysis, be content with empirical regularities.
  - causal analysis only in the framework of non-validated theoretical models.
- However, importance of empirically validated causal claims for policy issue, as well as for a deeper understanding of the phenomena.

Introduction

- How is possible learn *valid and reliable* causal relationships:
  - based on observations (empirical economics)
  - taking into account the potential influence of many interacting variables.
- Under these conditions causal inference is difficult. Possible and often practiced strategy:
  - give up empirical causal analysis, be content with empirical regularities.
  - causal analysis only in the framework of non-validated theoretical models.
- However, importance of empirically validated causal claims for policy issue, as well as for a deeper understanding of the phenomena.

Introduction

- How is possible learn *valid and reliable* causal relationships:
  - based on observations (empirical economics)
  - taking into account the potential influence of many interacting variables.
- Under these conditions causal inference is difficult. Possible and often practiced strategy:
  - give up empirical causal analysis, be content with empirical regularities.
  - causal analysis only in the framework of non-validated theoretical models.
- However, importance of empirically validated causal claims for policy issue, as well as for a deeper understanding of the phenomena.

Outline

- Is it possible to infer causal relationships from statistical data? Also in case of complex interactions?
- Possible methods for causal inference in case of:
  - non-Gaussianity
  - nonlinearity
  - aggregation structure
- The limits of causal inference from statistical data.
- In complex evolutionary systems: importance of singular, cumulative causation.

Outline

- ▷ Is it possible to infer causal relationships from statistical data? Also in case of complex interactions?
- ▷ Possible methods for causal inference in case of:
  - non-Gaussianity
  - nonlinearity
  - aggregation structure
- ▷ The limits of causal inference from statistical data.
- ▷ In complex evolutionary systems: importance of singular, cumulative causation.

Outline

- ▷ Is it possible to infer causal relationships from statistical data? Also in case of complex interactions?
- ▷ Possible methods for causal inference in case of:
  - non-Gaussianity
  - nonlinearity
  - aggregation structure
- ▷ The limits of causal inference from statistical data.
- ▷ In complex evolutionary systems: importance of singular, cumulative causation.

Outline

- ▷ Is it possible to infer causal relationships from statistical data? Also in case of complex interactions?
- ▷ Possible methods for causal inference in case of:
  - non-Gaussianity
  - nonlinearity
  - aggregation structure
- ▷ The limits of causal inference from statistical data.
- ▷ In complex evolutionary systems: importance of singular, cumulative causation.

Regularities

- Notwithstanding its complexity, economic world is full of regularities, as results of habits, social norms, institutional restrictions, conventions, etc.
- Interpretation of regularities:
  - Chance, uncertainty at the level of individual outcomes;
  - Order at the aggregate level.
- Underlying causal data generating mechanism.

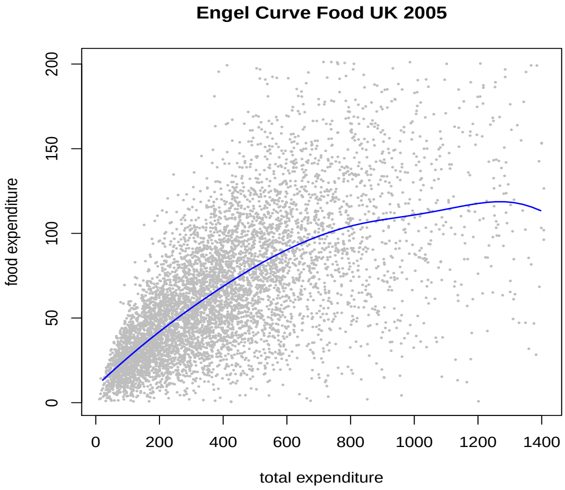
Regularities

- Notwithstanding its complexity, economic world is full of regularities, as results of habits, social norms, institutional restrictions, conventions, etc.
- Interpretation of regularities:
  - Chance, uncertainty at the level of individual outcomes;
  - Order at the aggregate level.
- Underlying causal data generating mechanism.

Regularities

- Notwithstanding its complexity, economic world is full of regularities, as results of habits, social norms, institutional restrictions, conventions, etc.
- Interpretation of regularities:
    - Chance, uncertainty at the level of individual outcomes;
    - Order at the aggregate level.
- Underlying causal data generating mechanism.

A famous empirical regularity



Empirical regularities

- The graph shows:
  - food expenditure grows less proportionally than total expenditure

    cfr. Engel law: as income increases, budget shares devoted to food decline;
  - as income increases, expenditures tend become more heterogeneous

    food expenditure variance is increasing
- Engel law robust across countries and over time
- Increasing variety robust across categories of expenditure

Empirical regularities

- The graph shows:
  - food expenditure grows less proportionally than total expenditure

    cfr. Engel law: as income increases, budget shares devoted to food decline;
  - as income increases, expenditures tend become more heterogeneous

    food expenditure variance is increasing
- Engel law robust across countries and over time
- Increasing variety robust across categories of expenditure

Empirical regularities

- The graph shows:
  - food expenditure grows less proportionally than total expenditure

    cfr. Engel law: as income increases, budget shares devoted to food decline;

  - as income increases, expenditures tend become more heterogeneous

    food expenditure variance is increasing

- Engel law robust across countries and over time
- Increasing variety robust across categories of expenditure

Interpreting Regularities

- What is the structure underlying these regularities?
- Engel law:
  - consumption is driven by wants, physiological wants are subject to saturation
- Increasing variety:
  - people have quite homogeneous physiological wants, but variety in preferences increases as the possibilities increase.

Interpreting Regularities

- What is the structure underlying these regularities?
- Engel law:
  - consumption is driven by wants, physiological wants are subject to saturation
- Increasing variety:
  - people have quite homogeneous physiological wants, but variety in preferences increases as the possibilities increase.

Interpreting Regularities

- What is the structure underlying these regularities?
- Engel law:
  - consumption is driven by wants, physiological wants are subject to saturation
- Increasing variety:
  - people have quite homogeneous physiological wants, but variety in preferences increases as the possibilities increase.

Interpreting Regularities

▷ Complexity of consumption decisions:

> *"influences other than current prices and current total expenditure must be systematically modelled if even the broad pattern of demand is to be explained" (Deaton and Muellbauer 1980)*

- Interaction of many variables with income and prices: status, location, family composition, cultural and biological determinants.
- Nonlinear interactions (e.g. demand for luxury goods)
- Dependence on the peers, imitation vs. specialization (cfr. Veblen's conspicuous consumption)

Regularity and chance mechanism

- Behind observed economic regularity we hypothesize the presence of a chance set up, a social mechanism which has generated the data.
- Social mechanism involving causal relationships.
- Problem of underdetermination: social mechanisms are often latent and several chance mechanisms may be compatible with the observed regularities.
- Nevertheless, in some cases is it possible to infer causal relationships from regularities
- The work of empirical economist: try to find out meaningful (i.e. causally interpretable) regularities
- Is this possible in a system of complex interactions?

Regularity and chance mechanism

- Behind observed economic regularity we hypothesize the presence of a chance set up, a social mechanism which has generated the data.
- Social mechanism involving causal relationships.
- Problem of underdetermination: social mechanisms are often latent and several chance mechanisms may be compatible with the observed regularities.
- Nevertheless, in some cases is it possible to infer causal relationships from regularities
- The work of empirical economist: try to find out meaningful (i.e. causally interpretable) regularities
- Is this possible in a system of complex interactions?

Regularity and chance mechanism

- Behind observed economic regularity we hypothesize the presence of a chance set up, a social mechanism which has generated the data.
- Social mechanism involving causal relationships.
- Problem of underdetermination: social mechanisms are often latent and several chance mechanisms may be compatible with the observed regularities.
- Nevertheless, in some cases is it possible to infer causal relationships from regularities
- The work of empirical economist: try to find out meaningful (i.e. causally interpretable) regularities
- Is this possible in a system of complex interactions?

Regularity and chance mechanism

- Behind observed economic regularity we hypothesize the presence of a chance set up, a social mechanism which has generated the data.
- Social mechanism involving causal relationships.
- Problem of underdetermination: social mechanisms are often latent and several chance mechanisms may be compatible with the observed regularities.
- Nevertheless, in some cases is it possible to infer causal relationships from regularities
- The work of empirical economist: try to find out meaningful (i.e. causally interpretable) regularities
- Is this possible in a system of complex interactions?

Regularity and chance mechanism

- Behind observed economic regularity we hypothesize the presence of a chance set up, a social mechanism which has generated the data.
- Social mechanism involving causal relationships.
- Problem of underdetermination: social mechanisms are often latent and several chance mechanisms may be compatible with the observed regularities.
- Nevertheless, in some cases is it possible to infer causal relationships from regularities
- The work of empirical economist: try to find out meaningful (i.e. causally interpretable) regularities
- Is this possible in a system of complex interactions?

Regularity and chance mechanism

- Behind observed economic regularity we hypothesize the presence of a chance set up, a social mechanism which has generated the data.
- Social mechanism involving causal relationships.
- Problem of underdetermination: social mechanisms are often latent and several chance mechanisms may be compatible with the observed regularities.
- Nevertheless, in some cases is it possible to infer causal relationships from regularities
- The work of empirical economist: try to find out meaningful (i.e. causally interpretable) regularities
- Is this possible in a system of complex interactions?

From regularity to causality (in general)

- Regularity in the framework of a statistical model
- Statistical model, in Fisher's tradition (cfr. Spanos 1999):
  - Probability model: $\Phi = \{f(x, y, z, \ldots; \theta)\}$
  - Sampling model:
    $\mathbf{X} := (X_1, \ldots, X_n), \mathbf{Y} := (Y_1, \ldots, Y_n), \mathbf{Z} := (Z_1, \ldots, X_n)$ is a random sample

    for random variables $X, Y, Z, \ldots$

Statistical dependence and independence

- Statistical dependence:

$$X \not\perp\!\!\!\perp Y \text{ iff } f(XY) \neq f(X)f(Y) \tag{1}$$

e.g. non-zero correlation $\rho(X, Y)$

- Statistical independence:

$$X \perp\!\!\!\perp Y \text{ iff } f(XY) = f(X)f(Y) \tag{2}$$

which implies zero correlation: $\rho(X, Y) = 0$.

- Conditional dependence:

$$X \not\perp\!\!\!\perp Y|Z \text{ iff } f(XY|Z) \neq f(X|Z)f(Y|Z) \tag{3}$$

e.g. non-zero partial correlation $\rho(X, Y|Z)$ (function of simple correlations)

- Conditional independence:

$$X \perp\!\!\!\perp Y|Z \text{ iff } f(XY|Z) \neq f(X|Z)f(Y|Z) \tag{4}$$

which implies zero partial correlation: $\rho(X, Y|Z) = 0$

Statistical dependence and independence

- Statistical dependence:

$$X \not\perp\!\!\!\perp Y \text{ iff } f(XY) \neq f(X)f(Y) \tag{1}$$

e.g. non-zero correlation $\rho(X, Y)$

- Statistical independence:

$$X \perp\!\!\!\perp Y \text{ iff } f(XY) = f(X)f(Y) \tag{2}$$

which implies zero correlation: $\rho(X, Y) = 0$.

- Conditional dependence:

$$X \not\perp\!\!\!\perp Y|Z \text{ iff } f(XY|Z) \neq f(X|Z)f(Y|Z) \tag{3}$$

e.g. non-zero partial correlation $\rho(X, Y|Z)$ (function of simple correlations)

- Conditional independence:

$$X \perp\!\!\!\perp Y|Z \text{ iff } f(XY|Z) \neq f(X|Z)f(Y|Z) \tag{4}$$

which implies zero partial correlation: $\rho(X, Y|Z) = 0$

Statistical dependence and independence

- Statistical dependence:

$$X \not\perp\!\!\!\perp Y \text{ iff } f(XY) \neq f(X)f(Y) \tag{1}$$

  e.g. non-zero correlation $\rho(X, Y)$

- Statistical independence:

$$X \perp\!\!\!\perp Y \text{ iff } f(XY) = f(X)f(Y) \tag{2}$$

  which implies zero correlation: $\rho(X, Y) = 0$.

- Conditional dependence:

$$X \not\perp\!\!\!\perp Y|Z \text{ iff } f(XY|Z) \neq f(X|Z)f(Y|Z) \tag{3}$$

  e.g. non-zero partial correlation $\rho(X, Y|Z)$ (function of simple correlations)

- Conditional independence:

$$X \perp\!\!\!\perp Y|Z \text{ iff } f(XY|Z) \neq f(X|Z)f(Y|Z) \tag{4}$$

  which implies zero partial correlation: $\rho(X, Y|Z) = 0$

Statistical dependence and independence

- Statistical dependence:

$$X \not\perp\!\!\!\perp Y \text{ iff } f(XY) \neq f(X)f(Y) \tag{1}$$

  e.g. non-zero correlation $\rho(X, Y)$

- Statistical independence:

$$X \perp\!\!\!\perp Y \text{ iff } f(XY) = f(X)f(Y) \tag{2}$$

  which implies zero correlation: $\rho(X, Y) = 0$.

- Conditional dependence:

$$X \not\perp\!\!\!\perp Y|Z \text{ iff } f(XY|Z) \neq f(X|Z)f(Y|Z) \tag{3}$$

  e.g. non-zero partial correlation $\rho(X, Y|Z)$ (function of simple correlations)

- Conditional independence:

$$X \perp\!\!\!\perp Y|Z \text{ iff } f(XY|Z) \neq f(X|Z)f(Y|Z) \tag{4}$$

  which implies zero partial correlation: $\rho(X, Y|Z) = 0$

Statistical and causal dependence

- Statistical dependence is not causal dependence
- but ...
- from conditional dependence and independence is possible to get useful inputs for causal inference
- notwithstanding the importance of background knowledge.

Statistical and causal dependence

- Statistical dependence is not causal dependence
- but ...
- from conditional dependence and independence is possible to get useful inputs for causal inference
- notwithstanding the importance of background knowledge.

Statistical and causal dependence

- Statistical dependence is not causal dependence
- but ...
- from conditional dependence and independence is possible to get useful inputs for causal inference
- notwithstanding the importance of background knowledge.

The SVAR approach to causality

- Suppose the (unobserved) Data Generating Process is:

$$\mathbf{y}_t = \mathbf{B}\mathbf{y}_t + \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \qquad (5)$$

  where $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt})^T$: ($k$ time series variables);
  $\mathbf{B}, \mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_p$: $(k \times k)$ matrices (structural coefficients);
  $\mathbf{B}$ has a zero diagonal;
  ε_t is a $k \times 1$ vector of structural error terms (usually $E(\varepsilon_t \varepsilon_t') = \mathbf{I}$).

- SVAR:

$$\mathbf{\Gamma}_0\mathbf{y}_t = \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \varepsilon_t, \qquad (6)$$

  where $\mathbf{\Gamma}_0 = (\mathbf{I} - \mathbf{B}^{-1})$

- Both equations (5) and (6) are structural-form equations: unobserved, DGP, and not directly estimable.

The SVAR approach to causality

- Suppose the (unobserved) Data Generating Process is:

$$\mathbf{y}_t = \mathbf{B}\mathbf{y}_t + \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (5)$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt})^T$: ($k$ time series variables);
$\mathbf{B}, \mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_p$: $(k \times k)$ matrices (structural coefficients);
$\mathbf{B}$ has a zero diagonal;
$\boldsymbol{\varepsilon}_t$ is a $k \times 1$ vector of structural error terms (usually $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{I}$).

- SVAR:

$$\mathbf{\Gamma}_0\mathbf{y}_t = \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (6)$$

where $\mathbf{\Gamma}_0 = (\mathbf{I} - \mathbf{B}^{-1})$

- Both equations (5) and (6) are structural-form equations: unobserved, DGP, and not directly estimable.

The SVAR approach to causality

- Suppose the (unobserved) Data Generating Process is:

$$\mathbf{y}_t = \mathbf{B}\mathbf{y}_t + \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (5)$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt})^T$: ($k$ time series variables);
$\mathbf{B}, \mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_p$: $(k \times k)$ matrices (structural coefficients);
$\mathbf{B}$ has a zero diagonal;
ε_t is a $k \times 1$ vector of structural error terms (usually $E(\varepsilon_t \varepsilon_t') = \mathbf{I}$).

- SVAR:

$$\mathbf{\Gamma}_0\mathbf{y}_t = \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \varepsilon_t, \quad (6)$$

where $\mathbf{\Gamma}_0 = (\mathbf{I} - \mathbf{B}^{-1})$

- Both equations (5) and (6) are structural-form equations: unobserved, DGP, and not directly estimable.

The SVAR approach to causality

- Suppose the (unobserved) Data Generating Process is:

$$\mathbf{y}_t = \mathbf{B}\mathbf{y}_t + \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{5}$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt})^T$: ($k$ time series variables);
$\mathbf{B}, \mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_p$: $(k \times k)$ matrices (structural coefficients);
$\mathbf{B}$ has a zero diagonal;
$\boldsymbol{\varepsilon}_t$ is a $k \times 1$ vector of structural error terms (usually $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{I}$).

- SVAR:

$$\mathbf{\Gamma}_0\mathbf{y}_t = \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{6}$$

where $\mathbf{\Gamma}_0 = (\mathbf{I} - \mathbf{B}^{-1})$

- Both equations (5) and (6) are structural-form equations: unobserved, DGP, and not directly estimable.

The SVAR approach to causality

- Suppose the (unobserved) Data Generating Process is:

$$\mathbf{y}_t = \mathbf{B}\mathbf{y}_t + \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{5}$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt})^T$: ($k$ time series variables);
$\mathbf{B}, \mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_p$: $(k \times k)$ matrices (structural coefficients);
$\mathbf{B}$ has a zero diagonal;
$\boldsymbol{\varepsilon}_t$ is a $k \times 1$ vector of structural error terms (usually $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{I}$).

- SVAR:

$$\mathbf{\Gamma}_0\mathbf{y}_t = \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{6}$$

where $\mathbf{\Gamma}_0 = (\mathbf{I} - \mathbf{B}^{-1})$

- Both equations (5) and (6) are structural-form equations: unobserved, DGP, and not directly estimable.

The SVAR approach to causality

▷ Suppose the (unobserved) Data Generating Process is:

$$\mathbf{y}_t = \mathbf{B}\mathbf{y}_t + \boldsymbol{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \boldsymbol{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (5)$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt})^T$: ($k$ time series variables);
$\mathbf{B}, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_p$: $(k \times k)$ matrices (structural coefficients);
$\mathbf{B}$ has a zero diagonal;
$\boldsymbol{\varepsilon}_t$ is a $k \times 1$ vector of structural error terms (usually $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{I}$).

▷ SVAR:

$$\boldsymbol{\Gamma}_0\mathbf{y}_t = \boldsymbol{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \boldsymbol{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (6)$$

where $\boldsymbol{\Gamma}_0 = (\mathbf{I} - \mathbf{B}^{-\mathbf{1}})$

▷ Both equations (5) and (6) are structural-form equations: unobserved, DGP, and not directly estimable.

Reduced-form VAR

▷ VAR model (reduced-form equation):

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t \quad (7)$$

Note the absence of the matrix of contemporaneous variables. On the r.h.s: only pre-determined variables.

▷ Eq. (6) is connected to eq. (7) in the following way:

$$\mathbf{y}_t = \mathbf{\Gamma}_0^{-1} \mathbf{\Gamma}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_0^{-1} \mathbf{\Gamma}_p \mathbf{y}_{t-p} + \mathbf{\Gamma}_0^{-1} \varepsilon_t \quad (8)$$

so that $\mathbf{A}_i = \mathbf{\Gamma}_0^{-1} \mathbf{\Gamma}_i$ for $i = 1, \ldots, p$ and $\mathbf{u}_t = \mathbf{\Gamma}_0^{-1} \varepsilon_t$.

▷ Notice that: $E(\mathbf{u}_t \mathbf{u}_t') \neq \mathbf{I}$

Reduced-form VAR

- VAR model (reduced-form equation):

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t \tag{7}$$

Note the absence of the matrix of contemporaneous variables.
On the r.h.s: only pre-determined variables.

- Eq. (6) is connected to eq. (7) in the following way:

$$\mathbf{y}_t = \mathbf{\Gamma}_0^{-1}\mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_0^{-1}\mathbf{\Gamma}_p\mathbf{y}_{t-p} + \mathbf{\Gamma}_0^{-1}\varepsilon_t \tag{8}$$

so that $\mathbf{A}_i = \mathbf{\Gamma}_0^{-1}\mathbf{\Gamma}_i$ for $i = 1, \ldots, p$ and $\mathbf{u}_t = \mathbf{\Gamma}_0^{-1}\varepsilon_t$.

- Notice that: $E(\mathbf{u}_t\mathbf{u}_t') \neq \mathbf{I}$

Reduced-form VAR

- VAR model (reduced-form equation):

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t \tag{7}$$

  Note the absence of the matrix of contemporaneous variables. On the r.h.s: only pre-determined variables.

- Eq. (6) is connected to eq. (7) in the following way:

$$\mathbf{y}_t = \mathbf{\Gamma}_0^{-1}\mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_0^{-1}\mathbf{\Gamma}_p\mathbf{y}_{t-p} + \mathbf{\Gamma}_0^{-1}\boldsymbol{\varepsilon}_t \tag{8}$$

  so that $\mathbf{A}_i = \mathbf{\Gamma}_0^{-1}\mathbf{\Gamma}_i$ for $i = 1, \ldots, p$ and $\mathbf{u}_t = \mathbf{\Gamma}_0^{-1}\boldsymbol{\varepsilon}_t$.

- Notice that: $E(\mathbf{u}_t\mathbf{u}_t') \neq \mathbf{I}$

Reduced-form VAR

- VAR model (reduced-form equation):

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t \quad (7)$$

Note the absence of the matrix of contemporaneous variables. On the r.h.s: only pre-determined variables.

- Eq. (6) is connected to eq. (7) in the following way:

$$\mathbf{y}_t = \mathbf{\Gamma}_0^{-1} \mathbf{\Gamma}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_0^{-1} \mathbf{\Gamma}_p \mathbf{y}_{t-p} + \mathbf{\Gamma}_0^{-1} \boldsymbol{\varepsilon}_t \quad (8)$$

so that $\mathbf{A}_i = \mathbf{\Gamma}_0^{-1} \mathbf{\Gamma}_i$ for $i = 1, \ldots, p$ and $\mathbf{u}_t = \mathbf{\Gamma}_0^{-1} \boldsymbol{\varepsilon}_t$.

- Notice that: $E(\mathbf{u}_t \mathbf{u}_t') \neq \mathbf{I}$

VAR estimation

- Estimation of the VAR model:

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t \tag{9}$$

  is quite straightforward.

- Most simple way: eq. (9) are k equations which can be estimated by OLS.

- Note: Granger causality is easily estimated in this context testing zero coefficients in $\mathbf{A}_1, \ldots, \mathbf{A}_p$.

VAR estimation

- Estimation of the VAR model:

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t \qquad (9)$$

  is quite straightforward.

- Most simple way: eq. (9) are k equations which can be estimated by OLS.

- Note: Granger causality is easily estimated in this context testing zero coefficients in $\mathbf{A}_1, \ldots, \mathbf{A}_p$.

Identification of SVAR

- Problem of identification: from the estimate of the VAR

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t \tag{10}$$

- we want to recover the SVAR:

$$\mathbf{\Gamma}_0\mathbf{y}_t = \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{11}$$

- more parameters in SVAR than in VAR.
- Notice that:

$$\mathbf{u}_t = \mathbf{\Gamma}_0^{-1}\varepsilon_t = \mathbf{B}\mathbf{u}_t + \varepsilon_t \tag{12}$$

- Possible strategy: from the statistical properties of $\mathbf{u}_t$ recover $\mathbf{B}$. From $\mathbf{B}$ (and all the parameters in 10) recover then:
  - $\mathbf{\Gamma}_0 = \mathbf{I} - \mathbf{B}^{-1}$
  - $\mathbf{\Gamma}_1 = \mathbf{\Gamma}_0\mathbf{A}_1$
  - …
  - $\mathbf{\Gamma}_p = \mathbf{\Gamma}_0\mathbf{A}_p$

Identification of SVAR

- Problem of identification: from the estimate of the VAR

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t \tag{10}$$

- we want to recover the SVAR:

$$\mathbf{\Gamma}_0\mathbf{y}_t = \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{11}$$

- more parameters in SVAR than in VAR.
- Notice that:

$$\mathbf{u}_t = \Gamma_0^{-1}\varepsilon_t = \mathbf{B}\mathbf{u}_t + \varepsilon_t \tag{12}$$

- Possible strategy: from the statistical properties of $\mathbf{u}_t$ recover $\mathbf{B}$. From $\mathbf{B}$ (and all the parameters in 10) recover then:
  - $\mathbf{\Gamma}_0 = \mathbf{I} - \mathbf{B}^{-1}$
  - $\mathbf{\Gamma}_1 = \mathbf{\Gamma}_0\mathbf{A}_1$
  - $\ldots$
  - $\mathbf{\Gamma}_p = \mathbf{\Gamma}_0\mathbf{A}_p$

Identification of SVAR

- Problem of identification: from the estimate of the VAR

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t \quad (10)$$

- we want to recover the SVAR:

$$\mathbf{\Gamma}_0 \mathbf{y}_t = \mathbf{\Gamma}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p \mathbf{y}_{t-p} + \varepsilon_t, \quad (11)$$

- more parameters in SVAR than in VAR.
- Notice that:

$$\mathbf{u}_t = \mathbf{\Gamma}_0^{-1} \varepsilon_t = \mathbf{B}\mathbf{u}_t + \varepsilon_t \quad (12)$$

- Possible strategy: from the statistical properties of $\mathbf{u}_t$ recover $\mathbf{B}$. From $\mathbf{B}$ (and all the parameters in 10) recover then:
  - $\mathbf{\Gamma}_0 = \mathbf{I} - \mathbf{B}^{-1}$
  - $\mathbf{\Gamma}_1 = \mathbf{\Gamma}_0 \mathbf{A}_1$
  - $\ldots$
  - $\mathbf{\Gamma}_p = \mathbf{\Gamma}_0 \mathbf{A}_p$

Identification of SVAR

- Problem of identification: from the estimate of the VAR

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t \quad (10)$$

- we want to recover the SVAR:

$$\mathbf{\Gamma}_0\mathbf{y}_t = \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (11)$$

- more parameters in SVAR than in VAR.
- Notice that:

$$\mathbf{u}_t = \mathbf{\Gamma}_0^{-1}\boldsymbol{\varepsilon}_t = \mathbf{B}\mathbf{u}_t + \boldsymbol{\varepsilon}_t \quad (12)$$

- Possible strategy: from the statistical properties of $\mathbf{u}_t$ recover $\mathbf{B}$. From $\mathbf{B}$ (and all the parameters in 10) recover then:
  - $\mathbf{\Gamma}_0 = \mathbf{I} - \mathbf{B}^{-1}$
  - $\mathbf{\Gamma}_1 = \mathbf{\Gamma}_0\mathbf{A}_1$
  - $\ldots$
  - $\mathbf{\Gamma}_p = \mathbf{\Gamma}_0\mathbf{A}_p$

Graphical Models for SVAR Identification

- Apply graphical causal search (Spirtes et al. 2000; Pearl 2000) to $\mathbf{u}_t : (u_{1t}, \ldots, u_{kt})$.
- The causal structure among $u_{1t}, \ldots, u_{kt}$ will deliver information on the matrix $\mathbf{B}$
- Notice that $\mathbf{B}$ determines causal relationships among $y_{1t}, \ldots, y_{kt}$

Graphical Models for SVAR Identification

- Apply graphical causal search (Spirtes et al. 2000; Pearl 2000) to $\mathbf{u}_t : (u_{1t}, \ldots, u_{kt})$.
- The causal structure among $u_{1t}, \ldots, u_{kt}$ will deliver information on the matrix $\mathbf{B}$
- Notice that $\mathbf{B}$ determines causal relationships among $y_{1t}, \ldots, y_{kt}$

Graphical Models for SVAR Identification

- Apply graphical causal search (Spirtes et al. 2000; Pearl 2000) to $\mathbf{u}_t : (u_{1t}, \ldots, u_{kt})$.
- The causal structure among $u_{1t}, \ldots, u_{kt}$ will deliver information on the matrix $\mathbf{B}$
- Notice that $\mathbf{B}$ determines causal relationships among $y_{1t}, \ldots, y_{kt}$

Graphical Models for SVAR Identification

- Graphical causal search applied to $\mathbf{u}_t : (u_{1t}, \ldots, u_{kt})$:
  - Test conditional independence among $(u_{1t}, \ldots, u_{kt})$
  - Apply search algorithm (e.g. PC algorithm, Spirtes et al. 2000):
    - Build a complete undirected graph among $(u_{1t}, \ldots, u_{kt})$;
    - Recursively eliminate edges using C.I. tests among $(u_{1t}, \ldots, u_{kt})$;
    - Identify unshielded colliders;
    - Identify chains;
    - Avoid cycles.

Graphical Models for SVAR Identification

- Graphical causal search applied to $\mathbf{u}_t : (u_{1t}, \ldots, u_{kt})$:
  - Test conditional independence among $(u_{1t}, \ldots, u_{kt})$
  - Apply search algorithm (e.g. PC algorithm, Spirtes et al. 2000):
    - Build a complete undirected graph among $(u_{1t}, \ldots, u_{kt})$;
    - Recursively eliminate edges using C.I. tests among $(u_{1t}, \ldots, u_{kt})$;
    - Identify unshielded colliders;
    - Identify chains;
    - Avoid cycles.

Graphical Models for SVAR Identification

- Graphical causal search applied to $\mathbf{u}_t : (u_{1t}, \ldots, u_{kt})$:
  - Test conditional independence among $(u_{1t}, \ldots, u_{kt})$
  - Apply search algorithm (e.g. PC algorithm, Spirtes et al. 2000):
    - Build a complete undirected graph among $(u_{1t}, \ldots, u_{kt})$;
    - Recursively eliminate edges using C.I. tests among $(u_{1t}, \ldots, u_{kt})$;
    - Identify unshielded colliders;
    - Identify chains;
    - Avoid cycles.

Graphical Models for SVAR Identification

- Graphical causal search applied to $\mathbf{u}_t : (u_{1t}, \ldots, u_{kt})$:
  - Test conditional independence among $(u_{1t}, \ldots, u_{kt})$
  - Apply search algorithm (e.g. PC algorithm, Spirtes et al. 2000):
    - Build a complete undirected graph among $(u_{1t}, \ldots, u_{kt})$;
    - Recursively eliminate edges using C.I. tests among $(u_{1t}, \ldots, u_{kt})$;
    - Identify unshielded colliders;
    - Identify chains;
    - Avoid cycles.

Graphical Models for SVAR Identification

- Graphical causal search applied to $\mathbf{u}_t : (u_{1t}, \ldots, u_{kt})$:
  - Test conditional independence among $(u_{1t}, \ldots, u_{kt})$
  - Apply search algorithm (e.g. PC algorithm, Spirtes et al. 2000):
    - Build a complete undirected graph among $(u_{1t}, \ldots, u_{kt})$;
    - Recursively eliminate edges using C.I. tests among $(u_{1t}, \ldots, u_{kt})$;
    - Identify unshielded colliders;
    - Identify chains;
    - Avoid cycles.

Graphical Models for SVAR Identification

- Graphical causal search applied to $\mathbf{u}_t : (u_{1t}, \ldots, u_{kt})$:
  - Test conditional independence among $(u_{1t}, \ldots, u_{kt})$
  - Apply search algorithm (e.g. PC algorithm, Spirtes et al. 2000):
    - Build a complete undirected graph among $(u_{1t}, \ldots, u_{kt})$;
    - Recursively eliminate edges using C.I. tests among $(u_{1t}, \ldots, u_{kt})$;
    - Identify unshielded colliders;
    - Identify chains;
    - Avoid cycles.

Graphical Models for SVAR Identification

- Graphical causal search applied to $\mathbf{u}_t : (u_{1t}, \ldots, u_{kt})$:
  - Test conditional independence among $(u_{1t}, \ldots, u_{kt})$
  - Apply search algorithm (e.g. PC algorithm, Spirtes et al. 2000):
    - Build a complete undirected graph among $(u_{1t}, \ldots, u_{kt})$;
    - Recursively eliminate edges using C.I. tests among $(u_{1t}, \ldots, u_{kt})$;
    - Identify unshielded colliders;
    - Identify chains;
    - Avoid cycles.

Graphical Models for SVAR Identification

- Graphical causal search applied to $\mathbf{u}_t : (u_{1t}, \ldots, u_{kt})$:
  - Test conditional independence among $(u_{1t}, \ldots, u_{kt})$
  - Apply search algorithm (e.g. PC algorithm, Spirtes et al. 2000):
    - Build a complete undirected graph among $(u_{1t}, \ldots, u_{kt})$;
    - Recursively eliminate edges using C.I. tests among $(u_{1t}, \ldots, u_{kt})$;
    - Identify unshielded colliders;
    - Identify chains;
    - Avoid cycles.

Standard setting

- Standard statistical setting
  - Gaussianity
  - Linearity
- Under these assumptions is easier to test conditional independence
  - $X \perp\!\!\!\perp Y \equiv \text{corr}(X, Y) = 0$
  - Fisher's z (Spirtes et al. 2000)

Standard setting

- Standard statistical setting
  - Gaussianity
  - Linearity
- Under these assumptions is easier to test conditional independence
  - $X \perp\!\!\!\perp Y \equiv \text{corr}(X, Y) = 0$
  - Fisher's z (Spirtes et al. 2000)

Justification of the standard setting

- Linearity:
  - Mill's argument:

    "In social phenomena the Composition of Causes is the universal law." (J.S. Mill 1843)

  - fits our intuition of proportional effect
- Gaussianity: central limit theorem.
  - fits our intuition that a random shocks is the sum of unobserved independent effects
- Empirical evidence shows:
  - Non-linearity in economic time-series (see e.g. Stock and Watson 1999; Granger 2008)
  - Non-Gaussian residuals (Lanne and Lütkepohl 2010)

Justification of the standard setting

- Linearity:
  - Mill's argument:

    > *"In social phenomena the Composition of Causes is the universal law." (J.S. Mill 1843)*

  - fits our intuition of proportional effect
- Gaussianity: central limit theorem.
  - fits our intuition that a random shocks is the sum of unobserved independent effects
- Empirical evidence shows:
  - Non-linearity in economic time-series (see e.g. Stock and Watson 1999; Granger 2008)
  - Non-Gaussian residuals (Lanne and Lütkepohl 2010)

Justification of the standard setting

- Linearity:
  - Mill's argument:

    > *"In social phenomena the Composition of Causes is the universal law." (J.S. Mill 1843)*

  - fits our intuition of proportional effect
- Gaussianity: central limit theorem.
  - fits our intuition that a random shocks is the sum of unobserved independent effects
- Empirical evidence shows:
  - Non-linearity in economic time-series (see e.g. Stock and Watson 1999; Granger 2008)
  - Non-Gaussian residuals (Lanne and Lütkepohl 2010)

Nonstandard settings

- Difficult cases for the SVAR / GM approach:
  - Non-Gaussianity
  - Non-linearity
  - Aggregation
  - Non-stability

Semi-parametric approach

- Assumptions:
  - linearity
  - non-Gaussianity
- Model:
  - VAR with non-Gaussian residuals

Reduced form (what we estimate):

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + u_t. \tag{13}$$

Structural form (what we try to get):

$$\Gamma_0 Y_t = \Gamma_1 Y_{t-1} + \ldots + \Gamma_p Y_{t-p} + \varepsilon_t, \tag{14}$$

- Application of Independent Component Analysis, based on the non-Gaussianity of ε_t (u_t), and assuming $\varepsilon_{t,1} \perp \varepsilon_{t,2} \perp \varepsilon_{t,2}, \ldots$

Semi-parametric approach

- Assumptions:
  - linearity
  - non-Gaussianity
- Model:
  - VAR with non-Gaussian residuals

  Reduced form (what we estimate):

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + u_t. \tag{13}$$

  Structural form (what we try to get):

$$\Gamma_0 Y_t = \Gamma_1 Y_{t-1} + \ldots + \Gamma_p Y_{t-p} + \varepsilon_t, \tag{14}$$

- Application of Independent Component Analysis, based on the non-Gaussianity of ε_t (u_t), and assuming $\varepsilon_{t,1} \perp\!\!\!\perp \varepsilon_{t,2} \perp\!\!\!\perp \varepsilon_{t,2}, \ldots$.

Semi-parametric approach

- Assumptions:
  - linearity
  - non-Gaussianity
- Model:
  - VAR with non-Gaussian residuals

  Reduced form (what we estimate):

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + u_t. \tag{13}$$

  Structural form (what we try to get):

$$\Gamma_0 Y_t = \Gamma_1 Y_{t-1} + \ldots + \Gamma_p Y_{t-p} + \varepsilon_t, \tag{14}$$

- Application of Independent Component Analysis, based on the non-Gaussianity of ε_t (u_t), and assuming $\varepsilon_{t,1} \perp\!\!\!\perp \varepsilon_{t,2} \perp\!\!\!\perp \varepsilon_{t,2}, \ldots$.

Semi-parametric approach

- Assumptions:
  - linearity
  - non-Gaussianity
- Model:
  - VAR with non-Gaussian residuals

  Reduced form (what we estimate):

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + u_t. \tag{13}$$

  Structural form (what we try to get):

$$\Gamma_0 Y_t = \Gamma_1 Y_{t-1} + \ldots + \Gamma_p Y_{t-p} + \varepsilon_t, \tag{14}$$

- Application of Independent Component Analysis, based on the non-Gaussianity of ε_t (u_t), and assuming $\varepsilon_{t,1} \perp\!\!\!\perp \varepsilon_{t,2} \perp\!\!\!\perp \varepsilon_{t,2}, \ldots$.

Independent Component Analysis

- Blind Source Separation method (cfr. Principal Component and Factor Analysis)
- Much used in neuroscience (brain imaging), and engineering (sources of signals recognition), only recently introduced in econometrics
- Difference with PCA and FA:
  - PCA, FA: find a set of signals which are *uncorrelated* (not identified)
  - ICA: extracts a set of independent sources
- Exploiting non-Gaussianity: identification of the sources (shocks).

Independent Component Analysis

- Blind Source Separation method (cfr. Principal Component and Factor Analysis)
- Much used in neuroscience (brain imaging), and engineering (sources of signals recognition), only recently introduced in econometrics
- Difference with PCA and FA:
  - PCA, FA: find a set of signals which are *uncorrelated* (not identified)
  - ICA: extracts a set of independent sources
- Exploiting non-Gaussianity: identification of the sources (shocks).

Independent Component Analysis

- Blind Source Separation method (cfr. Principal Component and Factor Analysis)
- Much used in neuroscience (brain imaging), and engineering (sources of signals recognition), only recently introduced in econometrics
- Difference with PCA and FA:
    - PCA, FA: find a set of signals which are *uncorrelated* (not identified)
    - ICA: extracts a set of independent sources
- Exploiting non-Gaussianity: identification of the sources (shocks).

Independent Component Analysis

- Blind Source Separation method (cfr. Principal Component and Factor Analysis)
- Much used in neuroscience (brain imaging), and engineering (sources of signals recognition), only recently introduced in econometrics
- Difference with PCA and FA:
    - PCA, FA: find a set of signals which are *uncorrelated* (not identified)
    - ICA: extracts a set of independent sources
- Exploiting non-Gaussianity: identification of the sources (shocks).

VAR-LiNGAM

Moneta, Entner, Hoyer and Coad (WP 2010)

▷ VAR-LiNGAM algorithm (Hyvärinen, Shimizu and Hoyer 2008)

- Estimate the reduced-form VAR
- Check that the residuals are non-Gaussian
- Use an ICA algorithm to decompose the residuals matrix
- Order the variables (residuals) so as to obtain a matrix Γ_0 that is close to lower triangular
- Once the instantaneous effects are identified, identify lagged effects
- Assumption of acyclicity but no need of "Faithfulness".

VAR-LiNGAM

Moneta, Entner, Hoyer and Coad (WP 2010)

▷ VAR-LiNGAM algorithm (Hyvärinen, Shimizu and Hoyer 2008)

- Estimate the reduced-form VAR
- Check that the residuals are non-Gaussian
- Use an ICA algorithm to decompose the residuals matrix
- Order the variables (residuals) so as to obtain a matrix Γ_0 that is close to lower triangular
- Once the instantaneous effects are identified, identify lagged effects
- Assumption of acyclicity but no need of "Faithfulness".

VAR-LiNGAM

Moneta, Entner, Hoyer and Coad (WP 2010)

▷ VAR-LiNGAM algorithm (Hyvärinen, Shimizu and Hoyer 2008)

- Estimate the reduced-form VAR
- Check that the residuals are non-Gaussian
- Use an ICA algorithm to decompose the residuals matrix
- Order the variables (residuals) so as to obtain a matrix Γ_0 that is close to lower triangular
- Once the instantaneous effects are identified, identify lagged effects
- Assumption of acyclicity but no need of "Faithfulness".

VAR-LiNGAM

Moneta, Entner, Hoyer and Coad (WP 2010)

▷ VAR-LiNGAM algorithm (Hyvärinen, Shimizu and Hoyer 2008)

- Estimate the reduced-form VAR
- Check that the residuals are non-Gaussian
- Use an ICA algorithm to decompose the residuals matrix
- Order the variables (residuals) so as to obtain a matrix Γ_0 that is close to lower triangular
- Once the instantaneous effects are identified, identify lagged effects
- Assumption of acyclicity but no need of “Faithfulness”.

VAR-LiNGAM

Moneta, Entner, Hoyer and Coad (WP 2010)

▷ VAR-LiNGAM algorithm (Hyvärinen, Shimizu and Hoyer 2008)

- Estimate the reduced-form VAR
- Check that the residuals are non-Gaussian
- Use an ICA algorithm to decompose the residuals matrix
- Order the variables (residuals) so as to obtain a matrix Γ_0 that is close to lower triangular
- Once the instantaneous effects are identified, identify lagged effects
- Assumption of acyclicity but no need of "Faithfulness".

VAR-LiNGAM

Moneta, Entner, Hoyer and Coad (WP 2010)

▷ VAR-LiNGAM algorithm (Hyvärinen, Shimizu and Hoyer 2008)

- Estimate the reduced-form VAR
- Check that the residuals are non-Gaussian
- Use an ICA algorithm to decompose the residuals matrix
- Order the variables (residuals) so as to obtain a matrix Γ_0 that is close to lower triangular
- Once the instantaneous effects are identified, identify lagged effects
- Assumption of acyclicity but no need of "Faithfulness".

Nonparametric approach

- Nonparametric test of conditional independence:
  - no functional form (or probability distribution) imposed
- Direct test (Chlaß and Moneta 2010):
  - $X \perp\!\!\!\perp Y \mid Z$ iff $f(X|Y,Z) = f(X|Z)$ iff $f(X,Y,Z)f(Z) = f(X,Z)f(Y,Z)$.
  - Estimate $\hat{f}(\cdot,\cdot,\cdot), \hat{f}(\cdot,\cdot), \hat{f}(\cdot)$. through kernel method.
  - Check $\hat{f}(X,Y,Z)\hat{f}(Z) \approx \hat{f}(X,Z)\hat{f}(Y,Z)$.
  - Distance measures between kernel density functions:
    - Euclidean distance (Szekely and Rizzo 2004)
    - Weighted Hellinger distance (Su and White 2008)

Nonparametric approach

- Nonparametric test of conditional independence:
  - no functional form (or probability distribution) imposed
- Direct test (Chlaß and Moneta 2010):
  - $X \perp\!\!\!\perp Y \mid Z$ iff $f(X|Y,Z) = f(X|Z)$ iff $f(X,Y,Z)f(Z) = f(X,Z)f(Y,Z)$.
  - Estimate $\hat{f}(\cdot,\cdot,\cdot)$,$\hat{f}(\cdot,\cdot)$,$\hat{f}(\cdot)$. through kernel method.
  - Check $\hat{f}(X,Y,Z)\hat{f}(Z) \approx \hat{f}(X,Z)\hat{f}(Y,Z)$.
  - Distance measures between kernel density functions:
    - Euclidean distance (Szekely and Rizzo 2004)
    - Weighted Hellinger distance (Su and White 2008)

Nonparametric approach

- Nonparametric test of conditional independence:
  - no functional form (or probability distribution) imposed
- Direct test (Chlaß and Moneta 2010):
  - $X \perp\!\!\!\perp Y \mid Z$ iff $f(X|Y,Z) = f(X|Z)$ iff $f(X,Y,Z)f(Z) = f(X,Z)f(Y,Z)$.
  - Estimate $\hat{f}(\cdot,\cdot,\cdot)$,$\hat{f}(\cdot,\cdot)$,$\hat{f}(\cdot)$. through kernel method.
  - Check $\hat{f}(X,Y,Z)\hat{f}(Z) \approx \hat{f}(X,Z)\hat{f}(Y,Z)$.
  - Distance measures between kernel density functions:
    - Euclidean distance (Szekely and Rizzo 2004)
    - Weighted Hellinger distance (Su and White 2008)

Nonparametric approach

- Nonparametric test of conditional independence:
  - no functional form (or probability distribution) imposed
- Direct test (Chlaß and Moneta 2010):
  - $X \perp\!\!\!\perp Y \mid Z \quad \text{iff} \quad f(X|Y,Z) = f(X|Z)$ iff $f(X,Y,Z)f(Z) = f(X,Z)f(Y,Z)$.
  - Estimate $\hat{f}(\cdot,\cdot,\cdot), \hat{f}(\cdot,\cdot), \hat{f}(\cdot)$. through kernel method.
  - Check $\hat{f}(X,Y,Z)\hat{f}(Z) \approx \hat{f}(X,Z)\hat{f}(Y,Z)$.
  - Distance measures between kernel density functions:
    - Euclidean distance (Szekely and Rizzo 2004)
    - Weighted Hellinger distance (Su and White 2008)

Nonparametric approach

- Curse of dimensionality
  - nonparametric estimates converge very slowly to their true values when there are many variables
- Curse of dimensionality tamed (not solved!) by bootstrap procedure
  - computational intensive
- Simulation results:
  - we are able to test conditional independence without imposing functional forms when the number of conditioned variables is limited
  - recover the causal structures for very few variables using the standard graphical-models procedure
  - in some cases sufficient for VAR identification.

Nonparametric approach

- Curse of dimensionality
  - nonparametric estimates converge very slowly to their true values when there are many variables
- Curse of dimensionality tamed (not solved!) by bootstrap procedure
  - computational intensive
- Simulation results:
  - we are able to test conditional independence without imposing functional forms when the number of conditioned variables is limited
  - recover the causal structures for very few variables using the standard graphical-models procedure
  - in some cases sufficient for VAR identification.

Nonparametric approach

- Curse of dimensionality
  - nonparametric estimates converge very slowly to their true values when there are many variables
- Curse of dimensionality tamed (not solved!) by bootstrap procedure
  - computational intensive
- Simulation results:
  - we are able to test conditional independence without imposing functional forms when the number of conditioned variables is limited
  - recover the causal structures for very few variables using the standard graphical-models procedure
  - in some cases sufficient for VAR identification.

Other issues

- Aggregation
  - Possible solution: Dynamic Factor Models
- Non-stability
  - Possible solutions: cointegration analysis, bootstrapping procedures

Is it possible to cope with complexity?

- Is it possible causal inference from statistical data in complex systems?
- Possible features of complexity:
  - Nonlinearities
  - Non-Gaussianity
  - Uncertainty about functional forms
  - Multi-variate system (cfr. curse of dimensionality)
  - Open system
  - Non-stationary time series
  - Structural changes

Is it possible to cope with complexity?

- Conjecture: from a certain level of complexity causal inference based on conditional independence is impossible.
- Argument: given the available statistical tools, impossibility of reliable inference about the data chance mechanism if all the complexity features apply.
- Should we renounce causal inference from a certain level of complexity?
  - Pessimistic views about empirical causal inference in economics: J.S. Mill, N. Cartwright, T. Lawson.
  - Availability of statistical tools.

Is it possible to cope with complexity?

- Conjecture: from a certain level of complexity causal inference based on conditional independence is impossible.
- Argument: given the available statistical tools, impossibility of reliable inference about the data chance mechanism if all the complexity features apply.
- Should we renounce causal inference from a certain level of complexity?
  - Pessimistic views about empirical causal inference in economics: J.S. Mill, N. Cartwright, T. Lawson.
  - Availability of statistical tools.

General vs Singular Causality

- Causation may be ontologically a unique concept, but it is necessary an epistemological distinction:
- General causality: notion based on regularity or stable variation.
- Singular causality: notion based on unique, singular events that change the behavior of a system.
- Are explanation and prediction possible only in the analysis of general causality?
- Narrative approach and beyond.

General vs Singular Causality

- Causation may be ontologically a unique concept, but it is necessary an epistemological distinction:
- General causality: notion based on regularity or stable variation.
- Singular causality: notion based on unique, singular events that change the behavior of a system.
- Are explanation and prediction possible only in the analysis of general causality?
- Narrative approach and beyond.

Cumulative Causation

- Notion introduced by Veblen (1898) to study the causal mechanisms of economic evolution, whose outcome is usually not in equilibrium.

"The modern scientist is unwilling to depart from the test of causal relation or quantitative sequence. When he asks the question, Why? he insists on an answer in terms of cause and effect. He wants to reduce his solution of all problems to terms of the conservation of energy and persistence of quantity...this...has in our time been made available for the handling of schemes of development and theories of a comprehensive process by the notion of cumulative causation" (Veblen 1898: 378).

Cumulative Causation

- How to empirically analyze cumulative causation?
- Simulation modelling: it "consists in replicating a portion of a real system ... to investigate causal relationships in simulated events and transferring these causes from the simulated to the real systems" (Valente 2005: 13).
- Problem of empirical validation.

  empirical calibration

Cumulative Causation

- How to empirically analyze cumulative causation?
- Simulation modelling: it “consists in replicating a portion of a real system ... to investigate causal relationships in simulated events and transferring these causes from the simulated to the real systems” (Valente 2005: 13).
- Problem of empirical validation.

  empirical calibration

Summary and Conclusions

- ▷ Complexity features pose a threat to the possibility of inferring causal relations from statistical models.
- ▷ This is possible as soon as phenomena exhibit chance regularity patterns.
  - Some degree of complexity are OK: chance presupposes uncertainty relating to the occurrence
  - as soon as regularity emerges in relation to the occurrence of many outcomes

Summary and Conclusions

- The possibility of recovering causal features of the data generating process depends on the available statistical tools + background knowledge.
- Some statistical tools for the case of causal inference in the SVAR / GM framework:
  - Independent Component Analysis
  - Nonparametric tests of conditional independence
  - Factor Models

Summary and Conclusions

- ▷ However, complexity poses serious threat on the possibility of inferring causal relations from statistical model.
  - many situations in which causal inference is difficult but still possible
- ▷ Distinction between general and singular causation.
- ▷ In a non-equilibrium (evolutionary) framework: importance of the study of process or cumulative causation (Veblen).
- ▷ Simulation models and their empirical validation.