Causal model search applied to economics: gains, pitfalls and challenges

Alessio Moneta

Institute of Economics
Scuola Superiore Sant'Anna, Pisa
amoneta@sssup.it

25 October 2013

Workshop: Case Studies of Causal Discovery with Model Search
Carnegie Mellon University

Preamble

- 20 years after the publication of *Causation, Prediction and Search* by Spirtes-Glymour-Scheines, 13 years after its 2$^{\text{nd}}$ ed. and *Causality* by Pearl :
  - causal model search is a fairly known approach in econometrics
  - but not many applications

Some numbers

- Scopus database (2001 - present): published articles in economics, econometrics and finance
  - containing *causal*, *causality* or *causation* in title, abstract, keywords: 4193 (about 5% of all articles)
  - same words + P. Spirtes or J. Pearl in the references: 84
  - C. Granger in the references: 1450 (1785 in text or references)
  - J.D. Angrist or J.-S. Pischke or G. Imbens: 338
  - J. Heckman: 277
  - P. Holland or D. Rubin: 193

- One field of econometrics where causal model search brings a clear contribution: Structural VAR (macroeconomic time series)
  - advantage of graphical causal models with respect to Granger-causality
- Less applications to microeconomic (panel) data
  - here the natural experiment approach is the standard

    instrumental variables (cfr. *Mostly Harmless Econometrics* approach by Angrist and Pischke)

    potential outcome, matching methods (Holland and Rubin approach)

My case study

- Exporting activity and productivity of firms

  causal relation between engagement in international trade and firm performance?

  which direction? both directions? complicated relationships?

Joint work with T. Ciarli and A. Coad (University of Sussex):
Exporting and productivity as part of the growth process: Results from a structural VAR
Working Paper

Background

- There is a substantial literature on the causes (and consequences) of international trade
  - comparative advantage
  - increasing return to scale
  - consumer love for variety
  - nature of exporting firms

(cfr. Bernard et al. *Firms in International Trade*, JEP 2007)

Background

- Trading firms are different from non-trading firms

  in terms of: size, productivity, skill and capital intensity, wages

  across a wide range of countries/industries

  differences exist even before exporting begins

- Are exporting firms more productive because of exporting or because only the most productive firms are able to enter the export market?

  possibly both

  - trade costs
  - learning by exporting

  mixed evidence about that

  (cfr. Bernard & Bradford *Exceptional exporter performance: cause, effect, or both?*, JIE 1999)

- Use of Instrumental Variables (cfr. Hansen, T. 2010) or matching methods.

Chile

- Focus on the dynamic interaction of export and productivity growth of exporting firms in Chile (2001-2006)

Our data

▷ Survey of manufacturing plants (Encuesta Nacional Insustrial Manufacturera) collected by Chilean Statistical Institute (INE):

plants with more than 10 employees, more than 6 months activity, classified in manufacturing sectors (ISIC 4 digits)

exporting in years 2001-2006

- size, proxied by employment (*employ*)
- output, proxied by (deflated) total sales (*output*) =

  domestic sales (*domsales*) +

  exports (*exp*)
- productivity, estimated by total factor productivity (*tfp*)

Measurement issues

- plants and firms
- which measure of firm performance?
- total factor productivity
- deflators
- treatment of outliers
- levels or differences

Our variables

1. *gr-domsales*$:= \Delta \log domsales_t = \log domsales_t - \log domsales_{t-1}$
2. *gr-empl*$:= \Delta \log empl_t$
3. *gr-exp*$:= \Delta \log exp_t$
4. *gr-tfp*$:= \Delta \log tfp_t$

▷ sample size: 4021 observations

1309 plants

5 time periods (not for all firms: the panel is not balanced)

10 manufacturing sectors

Causal model search

- Our strategy:
  - estimate a (pooled-panel) Vector Autoregressive (VAR) model
  - on the basis of the estimated residuals search for the underlying Structural VAR

VAR and SVAR

- Latent DGP:

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

$\mathbf{y}_t = (y_{1t}, \ldots, y_{Kt})^T$.
In our case $\mathbf{y}_t = (\textit{gr-domsales}, \textit{gr-empl}, \textit{gr-exp}, \textit{gr-tfp})^T$
B: $K \times K$ *zero diagonal* matrix
$\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_p$: $K \times K$ matrices
$\varepsilon_1, \ldots, \varepsilon_K$: structural error terms

- Equivalently (**S**tructural **V**ector **A**uto**r**egression):

$$\boldsymbol{\Gamma}_0\mathbf{y}_t = \boldsymbol{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \boldsymbol{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \quad (2)$$

$\boldsymbol{\Gamma}_0 = \mathbf{I} - \mathbf{B}$

- Estimation: endogeneity problem

VAR and SVAR

- Estimable (*reduced-form*) model (Vector Autoregression):

$$\begin{aligned} \mathbf{y}_t &= \mathbf{\Gamma}_0^{-1}\mathbf{\Gamma}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Gamma}_0^{-1}\mathbf{\Gamma}_p\mathbf{y}_{t-p} + \mathbf{\Gamma}_0^{-1}\boldsymbol{\varepsilon}_t \\ &= \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t \end{aligned} \tag{3}$$

$\mathbf{u}_t = \mathbf{\Gamma}_0^{-1}\boldsymbol{\varepsilon}_t$: vector of reduced-form error terms, white noise

- VAR (*reduced-form*) models are easy to estimate, but insufficient for causal knowledge

  NB: less parameters in VAR (eq. 3) than SVAR (eq. 2)

The problem of identification

▷ To identify the SVAR:

It is crucial (and sufficient) to find the correct $\mathbf{\Gamma}_0$ which produces the right transformation $\mathbf{\Gamma}_0\mathbf{u}_t = \boldsymbol{\varepsilon}_t$ of the VAR error terms $\mathbf{u}_t$

- NB: $\mathbf{\Gamma}_0$ incorporates information about contemporaneous causal relations among $y_{1t}, \ldots, y_{Kt}$

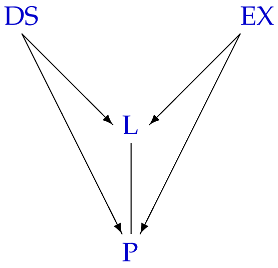
The problem of identification

- How to find Γ_0? Possible solutions:
  - Graphical Causal Models (Bessler and Lee 2002; Demiralp and Hoover 2003; Moneta 2006)

    if based on partial correlations the underlying assumption is Gaussianity

Results from GCM-SVAR

Ouput of PC algorithm applied to VAR-estimated residuals (OLS, 2 lags):



DS: $gr\text{-}domsales_t$
L: $gr\text{-}empl_t$
EX: $gr\text{-}exp_t$
P: $gr\text{-}tfp_t$

Conditional independence test

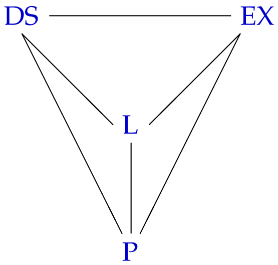
- Previous graph: based on zero partial correlation tests

  Fisher z transformation

  only one conditional independence was not rejected at 0.01 significance level: $H_0 : DS \perp\!\!\!\perp EX$ (p-value: 0.0167)

Results from GCM-SVAR

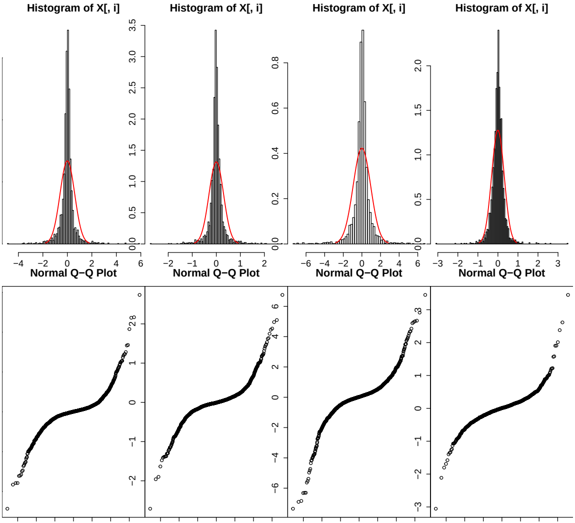
Ouput of PC algorithm applied to VAR-estimated residuals (OLS, 1 lag):



DS: $gr\text{-}domsales_t$
L: $gr\text{-}empl_t$
EX: $gr\text{-}exp_t$
P: $gr\text{-}tfp_t$

- GCM as constraint-based causal search is based on the adequateness of conditional independence tests
- Fisher's z require Gaussian data
- Wald tests on the VAR residuals: Gaussian asymptotic distribution

Non-Gaussianity



Shapiro-Wilk, Shapiro-Francia, Jarque-Bera tests strongly reject the H_0 : normality.

Alternative routes

- Testing conditional independence taking into account non-Gaussianity
  - non-parametric tests of conditional independence $X \perp\!\!\!\perp Y|Z$ based on some measures of distance between $\hat{f}(X,Y)\hat{f}(Z) = \hat{f}(X,Z)\hat{f}(Y,Z)$ (e.g. Euclidean or Hellinger or distance)
  - problems and possibilities for the problem at hand
- With these data: not much evidence for conditioned statistical dependence of any type.

Our route

- Application of Independent Component Analysis to the problem of VAR identification
- VAR-LiNGAM (cfr. Shimizu-Hoyer-Hyvärinen-Kerminen 2006; Hyvärinen-Shimizu-Hoyer 2008; Moneta-Entner-Hoyer-Coad 2013)
- Conditions:
  - Non-Gaussianity: structural shocks $\varepsilon_{1t}, \dots, \varepsilon_{Kt}$ are non-normally distributed
  - Independence: $\varepsilon_{1t}, \dots, \varepsilon_{Kt}$ are statistically independent:

$$P(\varepsilon_{1t}, \dots, \varepsilon_{Kt}) = P(\varepsilon_{1t}) \dots P(\varepsilon_{Kt})$$

  - Acyclicity: the contemporaneous causal structure among $y_{1t}, \dots, y_{Kt}$ is acyclic

VAR-LiNGAM algorithm (Hyvärinen et al. 2008)

1. Estimate the VAR model $\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t$. Check whether the residuals are non-Gaussian. Denote by $\hat{\mathbf{U}}$ the $K \times T$ matrix of estimated residuals.

2. Use FastICA to obtain $\hat{\mathbf{U}} = P\hat{\mathbf{E}}$, where $\mathbf{P}$ is $K \times K$ and $\hat{\mathbf{E}}$ is $K \times T$, such that the rows of $\hat{\mathbf{E}}$ are the independent components of $\hat{\mathbf{U}}$. Then validate non-Gaussianity and statistical independence of the components.

3. Let $\tilde{\tilde{\mathbf{\Gamma}}}_0 = \mathbf{P}^{-1}$. Find the permutation of rows of $\tilde{\tilde{\mathbf{\Gamma}}}_0$ which yields a matrix $\tilde{\mathbf{\Gamma}}_0$ without any zeros on the main diagonal. The permutation is sought which minimizes $\sum_i 1/|\tilde{\mathbf{\Gamma}}_{0_{ii}}|$.

4. Divide each row of $\tilde{\mathbf{\Gamma}}_0$ by its diagonal element, to obtain a matrix $\hat{\mathbf{\Gamma}}_0$ with all ones on the diagonal.

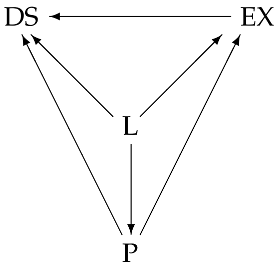
5. Let $\tilde{\mathbf{B}} = \mathbf{I} - \hat{\mathbf{\Gamma}}_0$.

6. Find the permutation matrix $\mathbf{Z}$ which makes $\mathbf{Z}\tilde{\mathbf{B}}\mathbf{Z}^T$ as close as possible to strictly lower triangular. (Minimize the sum of squares of the permuted upper-triangular elements). Set the upper-triangular elements to zero, and permute back to obtain $\hat{\mathbf{B}}$ which now contains the acyclic contemporaneous structure.

7. Calculate lagged causal effects $\hat{\mathbf{\Gamma}}_\tau = (\mathbf{I} - \hat{\mathbf{B}})\hat{\mathbf{A}}_\tau, \tau = 1, \ldots, p,$.

Results from VAR-LiNGAM

Ouput of LiNGAM applied to VAR-estimated residuals (LAD, 1 lag):



DS: *gr-domsales*$_t$
L: *gr-empl*$_t$
EX: *gr-exp*$_t$
P: *gr-tfp*$_t$

LAD (least absolute deviations) preferred to OLS in case of fat tails distributions

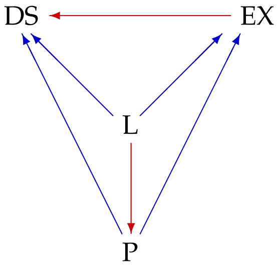
lag selection: 1 lag according to different criteria (Akaike Information, the Hannan-Quinn or the Schwarz Criterion)

Estimation

	B				**Γ_1**			
	gr_domsales	gr_empl	gr_exp	gr_tfp	l_gr_domsales	l_gr_empl	l_gr_exp	l_gr_tfp
gr_domsales	0	**0.4787**	**-0.0871**	**0.8301**	**-0.2248**	**0.1159**	-0.0148	**0.2906**
	0	0.0619	0.0127	0.0773	0.0384	0.033	0.008	0.0472
gr_empl	0	0	0	0	0.0073	-0.0229	0.0087	0.0162
	0	0	0	0	0.0068	0.0234	0.0041	0.0122
gr_exp	0	**0.4369**	0	**0.4023**	-0.0144	-0.0216	**-0.1442**	0.1088
	0	0.0889	0	0.0707	0.0157	0.0344	0.0392	0.0442
gr_tfp	0	**-0.2712**	0	0	0	**-0.0579**	**0.0121**	**-0.2699**
	0	0.0288	0	0	0.0096	0.0178	0.0044	0.02

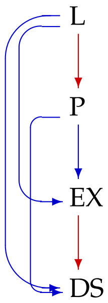
Bootstrap standard errors

Contemporaneous structure

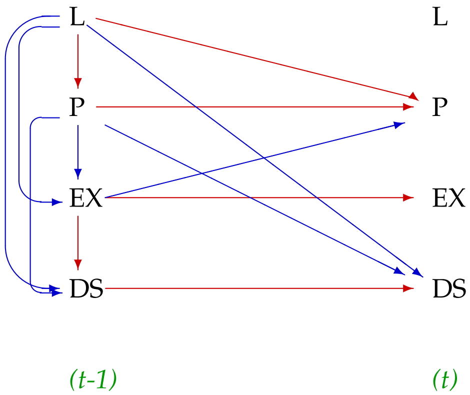


DS: $gr\text{-}domsales_t$
L: $gr\text{-}empl_t$
EX: $gr\text{-}exp_t$
P: $gr\text{-}tfp_t$
$\longrightarrow$ positive effects
$\longrightarrow$ negative effects

Contemporaneous structure



Lagged effects



L: $gr\text{-}empl_t$
P: $gr\text{-}tfp_t$
EX: $gr\text{-}exp_t$
DS: $gr\text{-}domsales_t$
$\longrightarrow$ positive effects
$\longrightarrow$ negative effects

Lagged effects displayed are only those significant at 0.05 significance level

Main causal mechanisms

Primus motor is **employment** growth, directly affecting domestic sales and exports

Employment growth negative effect on TFP growth

- downsizing firms better able to improve productivity than investing firms

TFP growth positive impacts on growth of domestic sales and of exports

- firms better off pursuing TFP growth as a prerequisite for sales growth

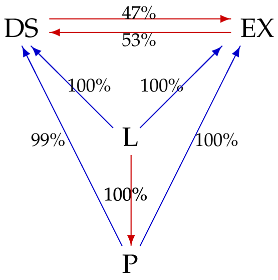
Export growth has a negative impact on growth of domestic sales.

- international firms focus on one market

TFP $\leftrightarrow$ **export**? Within the period: TFP $\rightarrow$ export and not vice versa
Exporting growth has a small positive impact on subsequent TFP growth
(much smaller effect: 0.4023 (s.e. 0.0707) vs. 0.0121 (s.e. 0.0044))

Robustness checks

Checking the robustness of the contemporaneous causal structure through a bootstrap procedure, percentage of links found:



DS: *gr-domsales*$_t$
L: *gr-empl*$_t$
EX: *gr-exp*$_t$
P: *gr-tfp*$_t$
$\longrightarrow$ positive effects
$\longrightarrow$ negative effects

Stability of causal orders

- across bootstrap samples
- control variables: dummies
- different measure of productivity
- sectors
- size
- old exporters new exporters

▷ so far: stability of the link $\text{TFP}_t \longrightarrow \text{export}_t$, absence of the link $\text{TFP}_t \longleftarrow \text{export}_t$, and weak but stable link $\text{export}_t \longrightarrow \text{TFP}_{t+1}$

▷ possibility of studying net effects of interventions through impulse response functions

What can we learn from causal-search methods?

- Gains
  - the emphasis is on the structural model (data generating process), it tries to goes beyond the reduced form model (associational model)
  - on the basis of an adequate characterization of the joint distribution of the observed variables it allows discriminating the possible causal structures
  - the automatic features of these methods helps perform a rigorous robustness analysis

- Pitfalls
  - possible sensitivity measured variables, controls, rescaling, lags

    possibility of data-mining/publication bias
  - heterogeneity (across individuals and over time)
  - feedback, latent variables and net effects
  - drawback of VAR analysis: number of shocks = number of equations

- Challenges
  - integrating causal search methods with panel data analysis

    allowing across-individual heterogeneity
  - integration with instrumental variable estimation
  - allowing for common shocks / common factors
  - meta-analysis of results coming from alternative specifications

Thank you