Independent and conditionally independent counterfactuals

Marcin Wolski

European Investment Bank

M.Wolski@eib.org

CRoNoS & MDA 2019
Limassol
April 15, 2019

Views expressed in this study are those of the author only, and do not necessarily represent the position of the European Investment Bank.

Overview

1. Introduction
   - Motivation
2. Framework
   - Unconditional dependence
   - Conditional setup
3. Numerics (unconditional)
   - Monte Carlo setup (unconditional)
4. Empirical application
   - Understanding the (Granger) dependence in the US grain market
5. Conclusions
   - The main take-aways
   - References
6. Supplementary materials

Counterfactuals

The goal of the counterfactual analysis is the comparison between what actually happened to what would have happened under an alternative scenario.

How to define alternative scenarios?

- Exogenous policy change (Rothe, 2010),
- treatment group (Chernozhukov et al., 2013),
- filter the dependence between variables (this paper).

Quick literature overview

The vast majority of impact evaluation studies focus on parametric models

- treatment effect models (Heckman, 1978),
- propensity score matching (Rosenbaum and Rubin, 1983),
- matching estimators models (Abadie and Imbens, 2002),
- OLS, diff-in-diff estimators (Gertler et al., 2010).

Non-parametric methods

- propensity score through a nonparametric regression model (Heckman, et al. (1997, 1998)),
- non-parametric/parametric method (Chernozhukov et al., 2013)
  - under an assumption called conditional exogeneity counterfactual effects can be interpreted as causal effects.
- fully nonparametric approach
  - total effects (Rothe, 2010),
  - partial effects (Rothe, 2012).

This paper

Theoretical contribution

- provide a fully non-parametric dependence filtering framework
  - unconditional dependence,
  - conditional dependence,
- consistent inference methods
  - Gaussian and bootstrap confidence bounds,
- utilize smooth estimates (improved MSE performance),
- numerical verification.

Empirical contribution

- build a link to hypothesis testing,
- understand the dependence structure in the US grain market.

Unconditional setup (I)

General assumptions

- Y outcome variable (1d) with CDF given by $F_Y(y)$,
- X covariate (vector) with CDF/PDF given by $F_X(x)$ and $f_X(x)$,
- i.i.d sample $\{(Y_i, X_i) : i = 1, ..., n\}$.

Variable dependence

- $F_{Y|X}(y|x) \neq F_Y(y)$ for some x, y.

The dependence filtering idea

- counterfactual outcome variable Y' with realizations y',
- $F_{Y'|X}(y|x) = F_Y(y)$ for all x, y,
- estimate realizations y'.

Unconditional setup (II)

Filtering through data sharpening (Hall and Minote, 2002)

- assume $Y' = \phi(Y, X = x) \equiv \phi(Y)$,
- so that $y' = \phi(y, x) \equiv \phi(y)$
- $\phi : \mathbb{R}^{d_X+1} \to \mathbb{R}$ and localy invertible.

Then the plug-in estimator of the joint density becomes

$$\hat{F}_{Y'|X}(y, x) = \frac{n^{-1} \sum_{i=1}^{n} \bar{K}_{\mathbf{H_Y}} \left(y - \phi(Y_i)\right) K_{\mathbf{H_X}} \left(x - X_i\right)}{n^{-1} \sum_{i=1}^{n} K_{\mathbf{H_X}} \left(x - X_i\right)} = F_Y(y), \quad (1)$$

where $\mathbf{H}_{\mathbf{H_Y}}$ and $\mathbf{H}_{\mathbf{H_X}}$ are bandwidth matrices, $\bar{K}_{\mathbf{H_Y}}$ is a cumulative and $K_{\mathbf{H_X}}$ is a scaled (multivariate) kernel function satisfying the standard regularity conditions (Wand and Jones, 1995).

Unconditional setup (III)

Theorem

Suppose that we have an i.i.d. sample $\{(Y_i, \mathbf{X_i}) : i = 1, ..., n\}$ *from a continuous distribution with well-defined and sufficiently smooth PDFs. Then, the counterfactual random variable* Y' *with realizations* y'*, satisfying the independence condition given by* $F_{Y'|X}(y|x) = F_Y(y)$*, follows*

$$F_Y(y') = F_{Y|\mathbf{X}}(y|\mathbf{x}), \tag{2}$$

where $F_{Y|\mathbf{X}}$ *is the conditional distribution function of* Y *given* $\mathbf{X} = \mathbf{x}$*, for any* y *and* $\mathbf{x}$ *in the support of* $(Y, \mathbf{X})$*.*

Unconditional setup (estimation)

The estimator of the independent counterfactuals

$$\hat{y}' = \hat{F}_Y^{-1}(\hat{F}_{Y|\mathbf{x}}(y|\mathbf{x})). \tag{3}$$

Theorem

Suppose that Assumptions 1-4 hold (see Appendix). Then

$$\sqrt{n}\left(\hat{y}' - y'\right) \xrightarrow{d} N(0, \sigma^2), \tag{4}$$

conditional on the data, where σ^2 is given by

$$\sigma^2 = \frac{F_{Y|X}(y|x)(1 - F_{Y|X}(y|x))}{f_Y\left(F_Y^{-1}(F_{Y|X}(y|x))\right)} + \frac{\int K(u)^2 \mathrm{d}u}{f_X(x)\Pi_{j=1}^{d_X} h_{jXY}} \frac{F_{Y|X}(y|x)(1 - F_{Y|X}(y|x))}{f_Y\left(F_Y^{-1}(F_{Y|X}(y|x))\right)}. \tag{5}$$

Unconditional setup (consistency)

Consistency of $\hat{y}'$ achieved under uniform convergence of estimators

- satisfied for 1-dimensional X,
- higher dimensions require lower estimate bias (higher order kernels)

Assumption (Bandwidths of conditional CDF)

As $n \to \infty$,

1. $n^{1/2} h_Y / (\log n)^{1/2} + n^{1/2} h_Y^r \to 0,$
2. $n^{1/2} h_X / \log n + n^{1/2} h_X^r \to 0,$

where r is the kernel order.

Unconditional setup (example)

Consider a stylized mean-dependent process

$$\begin{aligned} X &\sim N(0,1), \\ y_t &= ax_t + \sqrt{1-a^2}\varepsilon_t, \quad a \in (0,1). \end{aligned} \tag{6}$$

Independent counterfactuals are equal to the error term

$$\begin{aligned} y'_t &\equiv \phi(y_t, x_t) = F_Y^{-1}(F_{Y|X}(y_t|x_t)) \\ &= \sqrt{2}\mathrm{erf}^{-1}\left(\mathrm{erf}\left(\frac{y_t - ax_t}{\sqrt{2-2a^2}}\right)\right) = \frac{y_t - ax_t}{\sqrt{1-a^2}} = \varepsilon_t. \end{aligned} \tag{7}$$

More generally, under error exogeneity for nonseparable model

$$\begin{aligned} y_t &= m(x_t, \varepsilon_t), \\ y'_t &= F_Y^{-1}(F_\varepsilon(\varepsilon_t)). \end{aligned} \tag{8}$$

Conditional setup(I)

General assumptions

- Y outcome variable with CDF given by $F_Y(y)$,
- Q variable(s) with CDF/PDF given by $F_Q(q)$ and $f_q(q)$,
- X covariate (vector) with CDF/PDF given by $F_X(x)$ and $f_X(x)$,
- i.i.d sample $\{(Y_i, Q_i, X_i) : i = 1, ..., n\}$.

Variable dependence (CDF version of Diks and Panchenko, 2006)

- $F_{Y|Q,X}(y|q,x) \neq F_{Y|Q}(y|q)$ for some y, q, x.

The filtering idea

- counterfactual outcome variable Y'' with realizations y'',
- $F_{Y''|Q,X}(y|q,x) = F_{Y|Q}(y|q)$ for all y, q, x,
- estimate realizations y''.

Conditional setup (II)

Theorem

Suppose that we have an i.i.d. sample $\{(Y_i, \mathbf{Q_i}, \mathbf{X_i}) : i = 1, ..., n\}$ *from a continuous distribution with well-defined and sufficiently smooth PDFs. Then, the counterfactual random variable* Y''*, satisfying the conditional independence condition given by* $F_{Y''|Q,X}(y|q,x) = F_{Y|Q}(y|q)$*, follows asymptotically*

$$F_{Y|\mathbf{Q}}(y''|\mathbf{q}) = F_{Y|\mathbf{Q},\mathbf{X}}(y|\mathbf{q},\mathbf{x}). \tag{9}$$

Monte Carlo setup

Process specification (Diks and Wolski (2016))

$$\begin{aligned} X_i &\sim \mathrm{N}(0,1), \\ Y_i &\sim \mathrm{N}\left(0, c + aX_i^2\right), \end{aligned} \tag{10}$$

with $c > 0$ and $1 > a > 0$.

Filtering Mean Squared Error (MSE) is given by

$$\mathrm{MSE}(\hat{y}') = n^{-1} \sum_{i=1}^{n} \left(\hat{F}_Y^{-1}(\hat{F}_{Y|\mathbf{X}}^{-i}(y|\mathbf{x})) - F_Y^{-1}(F_{Y|\mathbf{X}}(y|\mathbf{x})) \right)^2 .$$

Technicalities

- compare step-wise and smooth kernel estimators (normal-scale, process-driven, LS-CV bandwidths)
- 1000 replications.

MSE performance

Table: Median MSE estimates of independent counterfactuals.

Bandwidth selector	n=50	n=100	n=200	n=500	n=1000
no smoothing	0.584	0.406	0.274	0.169	0.107
smoothing	0.292	0.232	0.178	0.116	0.080

Notes: Medians taken over 1000 Monte Carlo results for the ARCH process. Bandwidth selectors are chosen as: 'no smoothing' for step-wise estimators and 'smoothing' for normal-scale bandwidth selector.

Granger causality in the US grain market

Null hypothesis

$$\{X_t\} \text{ is not Granger causing } \{Y_t\}.$$

Diks & Wolski (2016)

- framework

$$Y_{t+1}|(X_t, Y_t) \sim Y_{t+1}|Y_t,$$

- test statistic (implication of the null)

$$q \equiv E\left[f_{X,Y,Z}(X,Y,Z)f_Y(Y) - f_{X,Y}(X,Y)f_{Y,Z}(Y,Z)\right] = 0.$$

Conditionally independent counterfactuals

- framework

$$F_{Y''_{t+1}|Y_t,X_t}(y_{t+1}|y_t,x_t) = F_{Y_{t+1}|Y_t}(y_{t+1}|y_t)\ ,$$

- test statistic

$$z \equiv \frac{\hat{y}''_{t+1} - y_{t+1}}{\hat{\sigma}}.$$

Granger causality in the US grain market

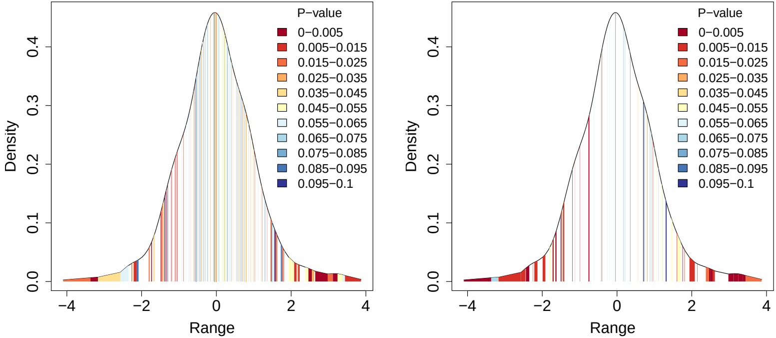
Table: US grain market results (Diks & Wolski, 2016).

Variables		Linear Granger Causality				Nonlinear Granger Causality (N)			
		Raw data		VAR residuals		Raw data		VAR residuals	
X	Y	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X
Corn	Wheat					***	***	***	***
Corn	Beans		*						
Beans	Wheat								

Notes: causality results for the pairwise relations of the log returns on the US grain market. (*), (**), (***) denote statistical significance at 10%, 5% and 1%. Period: 09/01/2010–03/06/2013. Nonlinear tests are performed on standardized data, transformed to (N)ormal marginals. The number of lags is $l_X = l_Y = 1$ *from the Bayesian Information Criterion.*

Granger causality in the US grain market

Figure: Wheat counterfactuals independent from corn (left) and beans (right).



Notes: causality results for the pairwise relations of the log returns on the corn → wheat (left panel) and beans → wheat (right panel). Period: 09/01/2010–03/06/2013. Nonlinear tests are performed on standardized data, transformed to (N)ormal marginals. The number of lags is $l_X = l_Y = 1$ *from the Bayesian Information Criterion.*

The main take-aways

Theory

- fully non-parametric dependence filtering framework,
  - unconditional + conditional dependence,
- standard + bootstrap confidence bounds,
- desired MSE performance on non-linear processes,
- good finite-sample properties.

Practice

- framework flexibility,
- hypothesis testing
  - further insights into the US grain market dependence structure.

In the future

- panel data extension,
- causal interpretation (yes, under error exogeneity).

Selected references

Gertler, P. J. and Martinez, S. and Premand, P. and Rawlings, L. B. and Vermeersch, C. M. J. (2010)
Impact Evaluation in Practice
World Bank Training

Rothe, C. (2010)
Nonparametric estimation of distributional policy effects
Journal of Econometrics 155 pp. 56 - 70

Chernozhukov, V. and Fernández-Val, I. and Melly, B. (2013)
Inference on counterfactual distributions
Econometrica 81(6) pp. 2205 - 2268

Diks, C. and Wolski, M. (2019, forthcoming)
New nonparametric measures for instantaneous and Granger-causality tail co-dependence

The End

Estimation assumptions (1)

Assumption (1)

Data $\{\mathbf{W}_i : i = 1, ..., n\}$, *where* $\mathbf{W}_i = \{W_{1i}, ..., W_{d_W i}\}$, *are i.i.d. as a* d_W*-variate smooth continuous distribution* $F_{\mathbf{W}}(\mathbf{w})$ *with well-defined PDF* $f_{\mathbf{W}}(\mathbf{w})$ *and respective derivatives, up to a finite order* r*, which are finite, continuous and uniformly bounded on the support.*

Estimation assumptions (2)

Assumption (2)

Kernel function $K : \mathbb{R}^{d_W} \to \mathbb{R}$ *behaves as*

$$\int K(\mathbf{w})\mathrm{d}\mathbf{w} = 1,$$
$$\int K(\mathbf{w})\mathbf{w}^c\mathrm{d}\mathbf{w} = 0 \quad \text{for} \quad c = 1, ..., r-1,$$
$$\int K(\mathbf{w})\mathbf{w}^c\mathrm{d}\mathbf{w} = \kappa_r I_{d_W} < \infty \quad \text{for} \quad c = r, \tag{11}$$

and $K(\mathbf{w})$ *is* r*-times differentiable, where* I_{d_W} *is a* $d_W \times d_W$ *identity matrix.*

Estimation assumptions (3/4)

Assumption (3)

As $n \to \infty$,

1. $n^{1/2} h_0 / (\log n)^{1/2} + n^{1/2} h_0^r \to 0$,
2. $n^{1/2} \det \mathbf{H}^{1/2} / \log n + n^{1/2} \max \mathbf{H}^{r/2} \to 0$.

Assumption (4)

We assume that (i) distribution functions F_Y *and* $F_{Y|\mathbf{X}}$ *are Hadamard differentiable, (ii)* F_Y^{-1} *is uniformly Lipschitz and bounded by* $[a, b] \in \mathbb{R}$*, (iii)* Y *is supported by a compact interval on* $J \in \mathbb{R}$ *for which* $F_{Y|\mathbf{X}}(y|\mathbf{x})$ *is uniformly bounded by* $[p_1, p_2] \in (0, 1)$.