Independent and conditionally independent counterfactual distributions

Marcin Wolski

European Investment Bank

M.Wolski@eib.org

Society for Nonlinear Dynamics and Econometrics
Tokyo
March 19, 2018

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

Overview

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 distributions,
  - conditional distributions,
- consistent inference methods
  - Gaussian and bootstrap confidence bounds,
- utilize smooth estimates (improved MSE performance),
- numerical verification.

Empirical contribution

- filter out the sovereign risk transmission on corporate costs of borrowing in selected euro area countries.

Unconditional setup (I)

General assumptions

- Y outcome variable (1d) with CDF/PDF given by $F_Y(y)$ and $f_y(y)$,
- 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 (Skaug and Tjostheim, 1993)

- $f_{X,Y}(x, y) \neq f_X(x) f_Y(y)$ for some x, y.

The filtering idea

- counterfactual distribution of outcome variable Y',
- $f_{Y'|X}(y|x) = f_Y(y)$ for all x, y.

Unconditional setup (II)

Filtering through data sharpening (Hall and Minote, 2002)

- assume that $Y' = \phi(Y|X = x) \equiv \phi(Y)$,
- $\phi : \mathbb{R} \to \mathbb{R}$ and localy invertible.

Then the plug-in estimator of the joint density becomes

$$\hat{f}_{Y',X}(y, x) = n^{-1} \sum_{i=1}^{n} K_{\mathbf{H}} \left(y - \phi(Y_i), x - X_i\right), \quad (1)$$

where $\mathbf{H}$ is a 2×2 bandwidth matrix and $K_{\mathbf{H}}$ 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 distribution* Y', *satisfying the independence condition given by* $f_{Y'|X}(y|x) = f_Y(y)$, *follows asymptotically*

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

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 counterfactual distribution

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

Theorem

Suppose that Assumptions 2-5 hold. 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*

$$\sigma^2 = \frac{F_Y(y)(1 - F_Y(y)) + F_{Y|\mathbf{X}}(y|x)(1 - F_{Y|\mathbf{X}}(y|x))}{f_Y\left(F_Y^{-1}(F_{Y|\mathbf{X}}(y|x))\right)}. \tag{5}$$

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 $n \to \infty$,

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

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

where r is the kernel order.

Conditional setup(I)

General assumptions

- Y outcome variable with CDF/PDF given by $F_Y(y)$ and $f_y(y)$,
- 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 (Diks and Panchenko, 2006)

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

The filtering idea

- counterfactual distribution of outcome variable Y'',
- $f_{Y''|Q,X}(y|q, x) = f_{Y|Q}(y|q)$ for all y, q, x.

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 distribution* Y''*, satisfying the conditional independence condition given by* $f_{Y''|Q,X}(y|q,x) = f_{Y|Q}(y|q)$*, follows asymptotically*

$$F_{Y|\mathbf{Q}}(y''|\mathbf{q}) = F_{Y|\mathbf{Q},\mathbf{X}}(y|\mathbf{q},\mathbf{x}). \quad (6)$$

Monte Carlo setup

Process specification (Diks and Wolski (2016))

$$
\begin{aligned}
X_i &\sim \mathrm{N}(0, 1), \\
Y_i &\sim \mathrm{N}\left(0, c + aX_i^2\right),
\end{aligned}
\tag{7}
$$

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

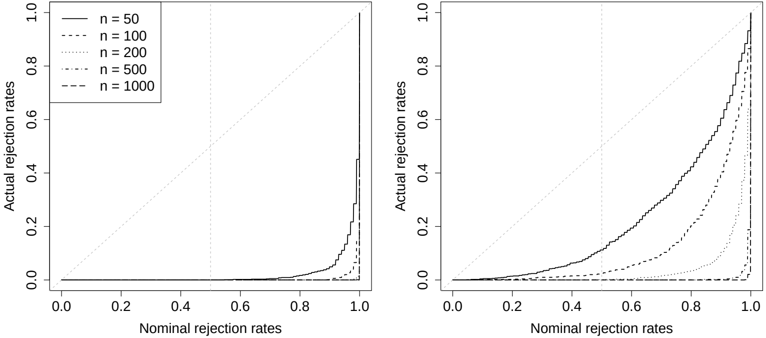
Table: Median MSE estimates of independent counterfactual distributions.

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.

Independence (Skaug and Tjostheim (1993))

Figure: Independence of counterfactual distributions.



Notes: Power-size plots show the actual rejection rates of the null hypothesis of independence for given nominal levels. Distribution under the null hypothesis is approximated with 99 bootstrap replicas. The results are aggregated over 1000 Monte Carlo simulations of ARCH process. Bandwidth selectors are chosen as 'no smoothing' (left) and 'smoothing' (right).

Sovereign spill-overs to bank lending rates

The factors which can hamper the effectiveness of monetary policy transmission to the bank lending rate include (EIB, 2016)

- high level of sovereign debt (sovereign performance),
- sluggish economic activity (macro performance),
- insufficient banks' capital positions (financial sector),
- high economic uncertainty (behavioral aspects),
- demand-side factors (corporate sector),
- and possibly other country-specific factors.

Empirical setup (I)

The basic pass-through equation

$$\Delta C_t = \sum_{k=0}^{IR} \beta_{kR} \Delta R_{t-k} + \sum_{j=1}^{IC} \beta_{jC} \Delta C_{t-k} + \sum_{m=1}^{IS} \beta_{mS} \Delta S_{t-m} + \alpha \nu_{t-1} + \varepsilon_t, \quad (8)$$

with

- C_t corporate cost of borrowing,
- R_t reference rate,
- $S_t$is the sovereign risk component,
- ν_t is the error correction factor $(C_t = \mu_0 + \mu_R R_t + \mu_S S_t + \nu_t)$,
- ε_t is the standard error term.

Empirical setup (II)

The basic pass-through equation

$$\Delta C_t = \sum_{k=0}^{IR} \beta_{kR} \Delta R_{t-k} + \sum_{j=1}^{IC} \beta_{jC} \Delta C_{t-k} + \sum_{m=1}^{IS} \beta_{mS} \Delta S_{t-m} + \alpha \nu_{t-1} + \varepsilon_t, \quad (9)$$

The filtering-equivalent equation

$$f_{\Delta C''|.}(\Delta C_t | \Delta R_t^{IR}, \Delta C_{t-1}^{IB}, \Delta S_{t-1}^{IS}, \nu_{t-1}) = f_{\Delta C|.}(\Delta C_t | \Delta R_t^{IR}, \Delta C_{t-1}^{IB}, \xi_{t-1}), \quad (10)$$

where
R_t^{IR} is vector of lags given by $R_t^{IR} = \{R_{t-IR}, ..., R_t\}$
$C_t^{IC} = \{C_{t-IC}, ..., C_t\}$,
$S_t^{IS} = \{S_{t-IS}, ..., S_t\}$.

Empirical setup (III)

The filtering pass-through equation

$$\hat{F}_{\Delta C|.}\left(\Delta C''|\Delta R_t^{IR}, \Delta C_{t-1}^{IC}, R_{t-1}, C_{t-1}\right)$$
$$= \hat{F}_{\Delta C|.}\left(\Delta C_t|\Delta R_t^{IR}, \Delta C_{t-1}^{IC}, \Delta S_{t-1}^{IS}, R_{t-1}, C_{t-1}, S_{t-1}\right), \quad (11)$$

Computational details

- lag order set to 1,
- smooth kernel PDF/CDF estimates (8th order Gaussian),
- normal-scale bandwidths.

Data

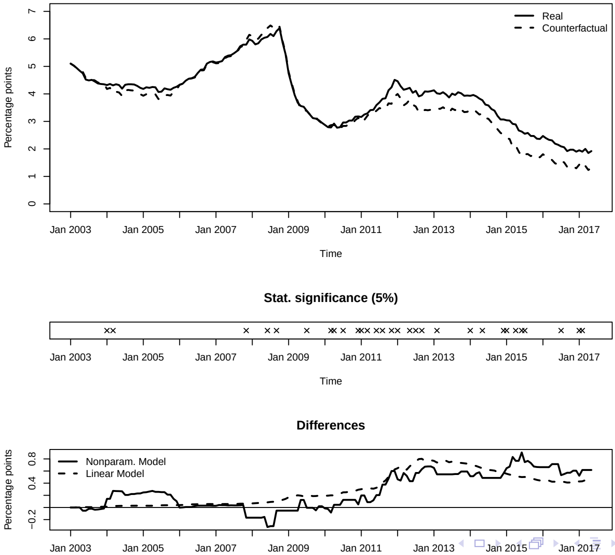
Data choice

- corporate cost of borrowing
  - bank loans + overdrafts,
  - new businesses,
- focus on Spain and Italy,
- data range: January 2003 until May 2017.

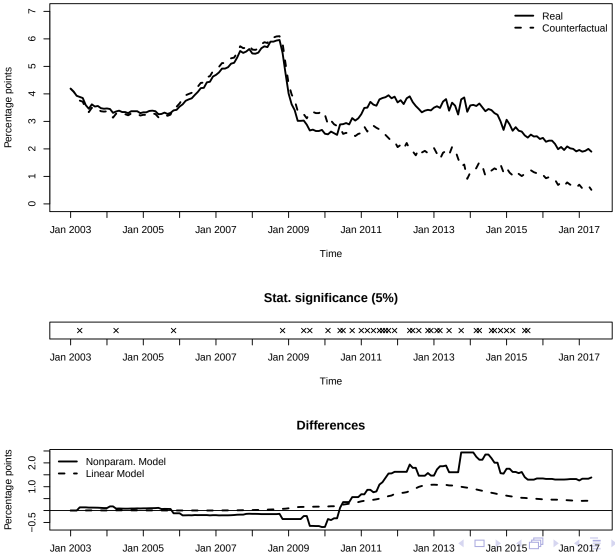
Data sources

- corporate borrowing rates of all maturities,
- 3-month EURIBOR rate as reference rate (robust to different maturities),
- sovereign risk approx. by 10-year sovereign yield spread over Germany.

Results - Italy



Results - Spain



The main take-aways

Theory

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

Practice

- framework flexibility,
- linear models can underestimate spillovers of sovereign risk distortions,
- heterogeneity in the ECB interest rate pass-through
  - heavy sovereign risk pass-through in Spain,
  - significant transmission of sovereign risk in Italy and Spain during the sovereign debt crisis.

In the future

- panel data extension,
- causal interpretation of the results.

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. (2018, forthcoming)
NCoVaR Granger causality
EIB Working Paper

The End

Filtering properties

Lemma

Suppose that Y' satisfies the conditions outlined in Theorem 1. Then,

$$F_{Y'}(y') = \delta(y, \mathbf{x}) F_Y(y'),$$

where $\delta(y, \mathbf{x}) = F_Y(y) F_{\mathbf{X}}(\mathbf{x}) / F_{Y,\mathbf{X}}(y, \mathbf{x})$.

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*

$$\int K(\mathbf{w})\mathrm{d}\mathbf{w} = 1,$$
$$\int K(\mathbf{w})\mathbf{w}^c\mathrm{d}\mathbf{w} = 0 \quad \text{for} \quad c = 1, ..., r-1, \tag{12}$$
$$\int K(\mathbf{w})\mathbf{w}^c\mathrm{d}\mathbf{w} = \kappa_r I_{d_W} < \infty \quad \text{for} \quad c = r,$$

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 $n \to \infty$,

(i) $n^{1/2} h_0 / (\log n)^{1/2} + n^{1/2} h_0^r \to 0$,

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

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* $[p_1, p_2] \in (0, 1)$.

Data description

	Obs.	Mean	St. dev.	Min	Max	ADF	ADF (Δ)
France							
Corporate borrowing cost	173	3.051	1.118	1.450	5.820	0.678	0.000
Sovereign risk	173	0.321	0.298	-0.007	1.450	0.530	0.000
EURIBOR (3 month)	173	1.570	1.542	-0.330	5.113	0.403	0.000
Italy							
Corporate borrowing cost	173	3.945	1.085	1.850	6.390	0.728	0.000
Sovereign risk	173	1.256	1.163	0.098	4.833	0.645	0.000
EURIBOR (3 month)	173	1.570	1.542	-0.330	5.113	0.403	0.000
Spain							
Corporate borrowing cost	173	3.528	0.952	1.900	5.960	0.818	0.000
Sovereign risk	173	1.195	1.307	-0.021	5.512	0.899	0.000
EURIBOR (3 month)	173	1.570	1.542	-0.330	5.113	0.403	0.000

Notes: Time span covers January 2003 - May 2017. Corporate borrowing cost is taken as the composite indicator of the cost of borrowing for non-financial corporations across maturities. Sovereign risk is taken as 10-year sovereign yield spread against German equivalent. ADF and ADF (Δ) denote the p-values from the Augmented Dickey-Fuller test on levels and first differences, respectively. Sources: ECB and Bloomberg.