The heterogeneous causal effects of the EU's Cohesion Fund

Angelos Alexopoulos[1], Ilias Kostarakos[2], Christos Mylonakis[3], Petros Varthalitis[1]

CRETE 2025,

Tinos, July 9, 2025

[1]Athens University of Economics and Business
[2]European Commission, JRC-Seville
[3]CEMFI

Outline

- Motivation
- Data
- Empirical model
- Results

Cohesion policy

- From its inception as a sub-national transfers program, **cohesion policy**, has disbursed almost **€960 bil.** to the EU's regions; second largest item of the EU budget, accounting for almost one-third.
- Its **targets** include among others, '...reducing disparities between the levels of development of the various regions and the backwardness of the least favored regions'.
- Naturally, given its scope and size, the **effectiveness** of cohesion policy has been the subject of long debates.
- We focus on one of the cohesion policy instruments, the **Cohesion Fund**.

Our Aim

- Ex-post **evaluation of the impact** of the Cohesion Fund (CF)
- Estimation of **causal effects** on regional output and investment
- Identification of **time-varying and heterogeneous** effects across regions

Related Literature

There is a **vast and expanding** empirical literature that aims to identify the impact of cohesion policy on regional economic performance, based on different methodological approaches:

1. **RDD** or **IV** to identify the causal impact of the EU's Structural Funds on the outcome of interest (output, investment, employment) - see, e.g., Becker et al. (2010), Becker et al. (2013), Pellegrini et al. (2013), Becker et al. (2018) and Cerqua and Pellegrini (2018).
2. **Dose-response** approaches to identify the level of EU transfers intensity (GDP share of EU transfers) that will maximize the outcome effects; Hagen and Mohl (2008), Becker et al. (2012), Crescenzi and Giua (2020)

Identified methodological issues

- **Only local or homogeneous effects**
  Dose-response models assume constant effects across units/time; IV and RDD estimate LATE for compliers only
- **Less efficient estimators**
  Dose-response suffers from high variance; IV can be inefficient, especially with weak instruments.
- **Do not account for unobserved confounding**
  Dose-response requires unconfoundedness and common support; IV needs a valid instrument, often unavailable.

Our contribution

We overcome the above-mentioned issues by implementing a **generalized synthetic control** approach (Athey et al. (2021)) which relies on interactive fixed effects (Pesaran (2006) and Bai (2009)) to:

- Estimate **region- and time-specific** causal effects
- Drop the (common in DiD) **parallel trends** assumption
- Model the presence of the **unobserved confounders** directly via the interactive fixed effects.

Data

Data sources

- The **Cohesion portal** from which we source data on the **EU's Cohesion Fund expenditures** (in current prices); the entire universe of the Cohesion Fund payments is covered in the latest update.
- The **Annual Regional Database** (ARDECO) of the European Commission for macroeconomic data.

Sample

- Our dataset originally included data from **27 European countries** with $N = 242$ for a maximum of $T = 43$.
- The treated sample covers **96 EU NUTS2 regions over the 1993-2022 period** for a total of 2297 observations and an average of 23.9 years under treatment.
- These regions belong to **16 countries**, the majority of which are the newest Member States of the EU, i.e. the Central and Eastern European countries that entered the Union starting with the 2004 enlargement.
- Only 18 regions (located in Greece and Portugal) are under treatment for the whole sample period, with 78 regions being treated for up to 23 years.

Empirical framework

Potential Outcomes Framework

We work under the framework of potential outcomes (Holland, 1986):

- $y_{it}(0)$: Potential untreated outcome.
- $y_{it}(1)$: Potential treated outcome.

Only one of these outcomes is observed. Moreover, $\mathcal{N}$ denotes the set of **treated** regions and T_i the last time period before unit i received the intervention. For treated regions $i \in \mathcal{N}$ at time $t > T_i$, the causal effect of the policy intervention is:

$$\tau_{it} = y_{it}(1) - y_{it}(0) = y_{it} - y_{it}(0)$$

Treatment Indicator:

- $D_{it} = 1$ if region i received treatment at time t.
- $D_{it} = 0$ otherwise.

Causal inference using matrix completion

- We estimate unobserved potential untreated outcomes $y_{it}(0)$ for $i \in \mathcal{N}$, $t > T_i$.
- We estimate the causal effect:

$$\widehat{\tau}_{it} = y_{it} - \widehat{y}_{it}(0), \tag{1}$$

and other summaries of interest, e.g.,

$$\widehat{ATT}_t = \frac{1}{N_{tr}} \sum_{i \in \mathcal{N}} \widehat{\tau}_{it}. \tag{2}$$

The model

We use a factor model for the observed outcomes (Athey et al., 2021, Xu, 2017):

$$y_{it} = \tau_{it} D_{it} + X_{it}^{\top} \beta + \lambda_i^{\top} f_t + \epsilon_{it},$$

where: f_t: $K \times 1$ vector of unobserved factors, λ_i: $K \times 1$ vector of factor loadings, ϵ_{it}: random error, X_{it} : $p \times 1$ vector and β are the corresponding coefficients

Estimation

- For the set of **never-treated and not-yet-treated** regions ($i \notin \mathcal{N}$, or $i \in \mathcal{N}$ and $t \leq T_i$) we have that $D_{it} = 0$, i.e.,

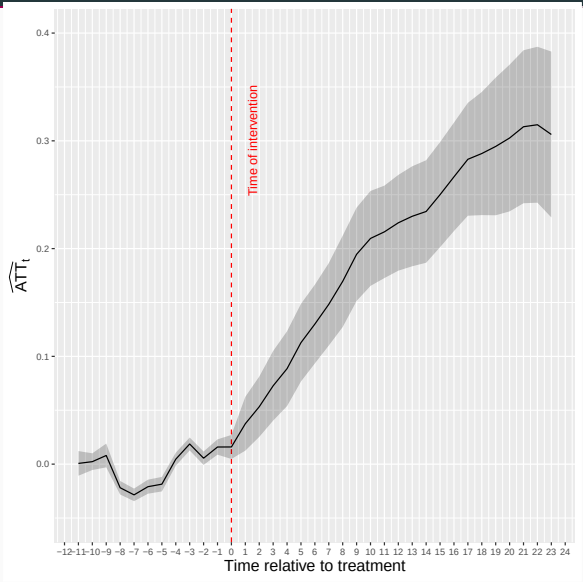
$$y_{it} = X_{it}^{\top}\beta + \lambda_i^{\top} f_t + \epsilon_{it},$$

- By using **only** observations of the never-treated and not-yet-treated regions we employ the matrix completion estimator developed by Athey et al. (2021) to obtain $\hat{\beta}$, $\{\hat{f}_t\}_{t=1}^{T}$, and $\{\hat{\lambda}_i\}_{i=N}^{T}$
- Then, we estimate $y_{it}(0)$ for **treated** regions, $i \in \mathcal{N}$ and $t > T_i$, as

$$\widehat{y}_{it}(0) = X_{it}^{\top}\hat{\beta} + \hat{\lambda}_i^{\top}\hat{f}_t.$$

Results

Average Treatment Effect: GVA

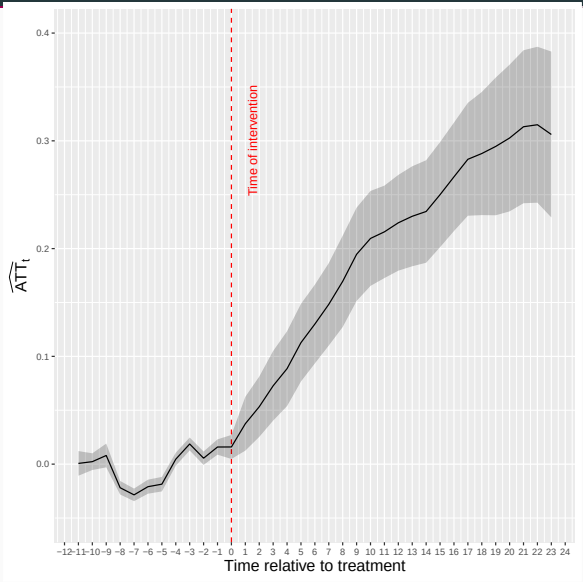


y_{it}: GVA of the i-th region at year t

$$\widehat{ATT}_t = \frac{1}{N_{tr}} \sum_{i \in \mathcal{N}} \widehat{\tau}_{it}$$

$$= \frac{1}{N_{tr}} \sum_{i \in \mathcal{N}} [y_{it} - \widehat{y}_{it}(0)]$$

Average Treatment Effect: GVA



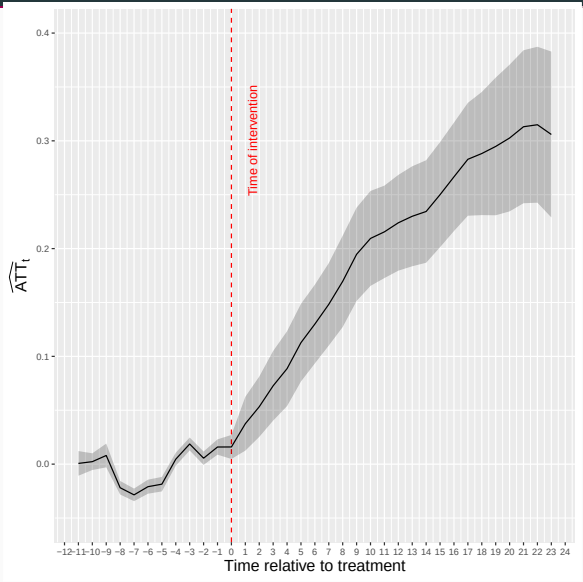
y_{it}: GVA of the i-th region at year t

$$\widehat{ATT}_t = \frac{1}{N_{tr}} \sum_{i \in \mathcal{N}} \widehat{\tau}_{it}$$

$$= \frac{1}{N_{tr}} \sum_{i \in \mathcal{N}} [y_{it} - \widehat{y}_{it}(0)]$$

- Cohesion Fund payments have a **significant and sustained positive** effect on the GVA per capita over time.
- Most of the GVA increase occurs within the first 10 years of treatment; smaller gains beyond that period.

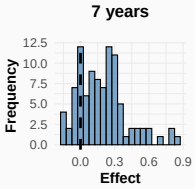
Average Treatment Effect: Investment

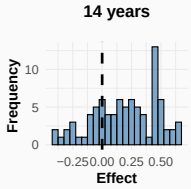


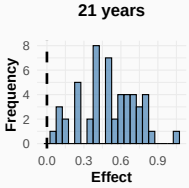
y_{it}: GFCF of the i-th region at year t

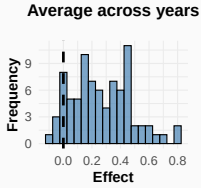
- Lasting **positive** impact.
- **Front loaded effect:** The largest increase occurs in the first 7 years.

Heterogeneous effects: Distribution of GVA effects





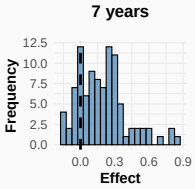


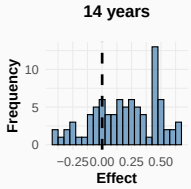


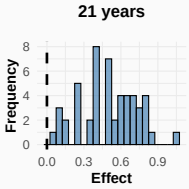
$$\widehat{\tau}_i = \frac{1}{T - T_i} \sum_{t=T_i+1}^{T} \widehat{\tau}_{it}$$

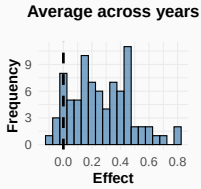
$$= \frac{1}{T - T_i} \sum_{t=T_i+1}^{T} [y_{it} - \widehat{y}_{it}(0)]$$

Heterogeneous effects: Distribution of GVA effects









$$\widehat{\tau}_i = \frac{1}{T - T_i} \sum_{t=T_i+1}^{T} \widehat{\tau}_{it}$$

$$= \frac{1}{T - T_i} \sum_{t=T_i+1}^{T} [y_{it} - \widehat{y}_{it}(0)]$$

- Significant heterogeneity
- The observed variation suggests that EU policymakers could **optimize** resource allocation by considering the intensity of treatment (CF expenditures relative to GVA)

Heterogeneous Effects: Distribution metrics

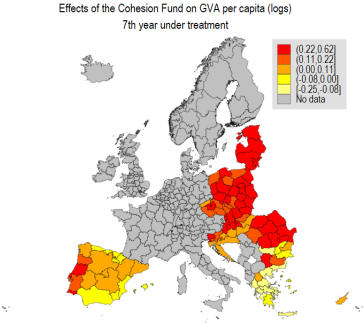
Table 1: Summary statistics for the distribution of the regional effect on GVA per capita

	7 years	14 years	21 years	Average
% of regions with $\widehat{\tau}_i > 0$	75	80	89.8	75.9
Skewness	0.169	0.184	0.06	1.067
Kurtosis	2.735	1.87	2.02	3.508

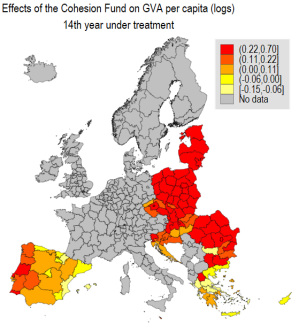
Key results:

1. **Positive effects:** Share of regions with $\widehat{\tau}_i > 0$ rises from 75% to 89.8% over time.
2. **Skewness:** Low and declining; large positive effects becoming less pronounced by year 21.
3. **Kurtosis:** Mild variation; occasional extreme positive effects.

Geographical illustration of heterogeneous effects



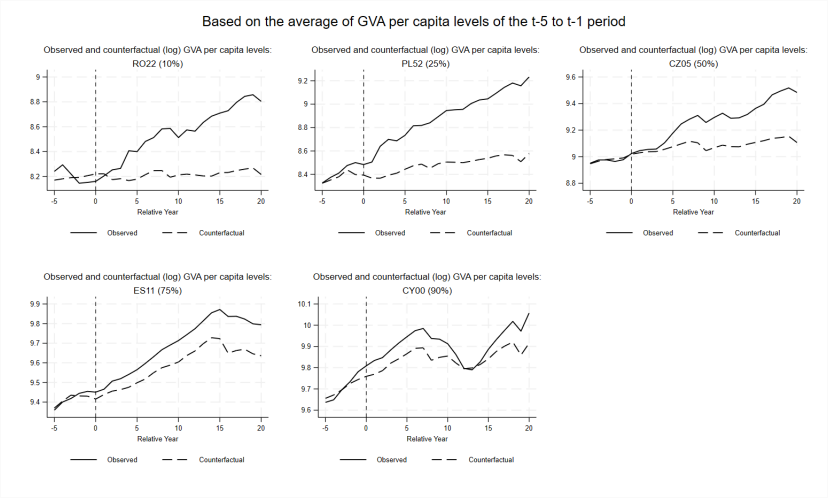
Note: 24 regions with negative effects



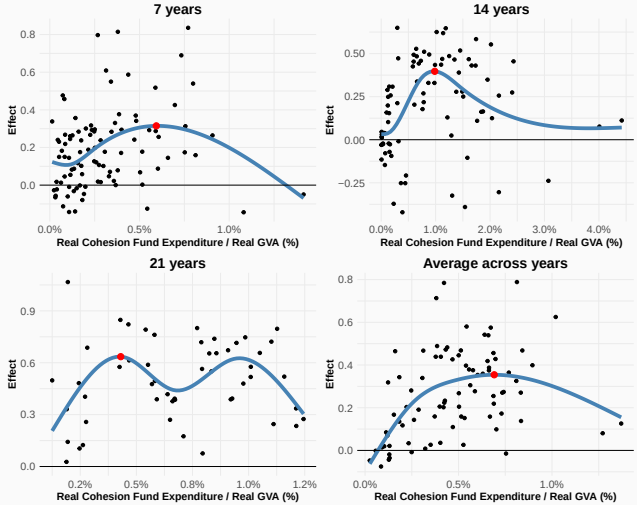
Note: 19 regions with negative effects

Geographical Dispersion of GVA effects

Heterogeneous effects: More benefit for poor regions



Nonlinear effects



Non-linear relationship between impact magnitude and treatment intensity

Concluding remarks

- We found **positive and persistent effect** of the **cohesion fund** on output (and investment)
- Our estimated **heterogeneous** causal effects show that:
  - GVA (and investment) gains are positive but **unevenly** distributed
  - **Less developed** regions benefit more from the program
  - A **non-linear** relationship between funding intensity and effect's size.
- **Policy implication**:
  - Large degree of heterogeneity maybe implies the need for more **tailored design** of the CF policy

Thank you!

Appendix

Allocation of the Cohesion Fund

Period	Countries	Regions	GDP pc $< 75\%$	CF amount	Share of CP(%)
1994-1999	4	42	16	18078.27	
2000-2006	16	100	74	30618.79	17.7
2007-2013	16	102	71	68939.32	27.1
2014-2020	15	83	76	47953.53	22.3

Notes: Values are in €millions.
The number of regions with GDP per capita lower than 75% of the EU average is calculated based on data corresponding to the first year of each programming period. Moreover, regions correspond to the 2021 NUTS2 version.
Share of CP refers to CF payments as a share of the total cohesion policy payments over the respective period.

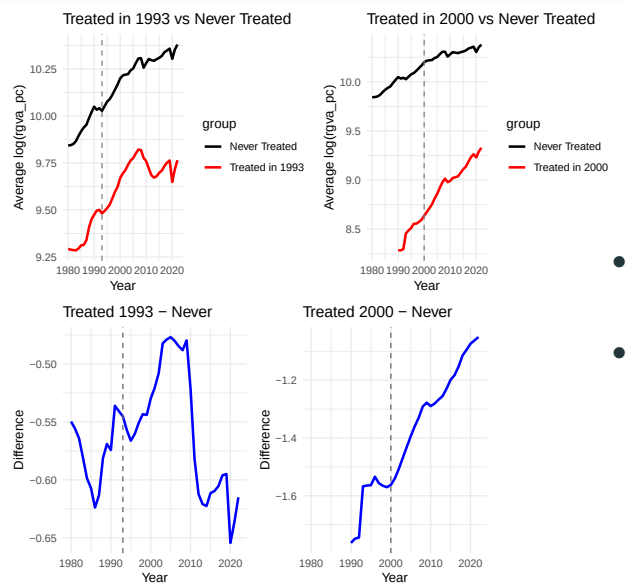
Factor Model Advantages

- Relaxes parallel trends assumption.
- Models unobserved time-varying confounders using $\lambda_i^\top f_t$.
- Accounts for heterogeneous causal effects τ_{it} across time and regions.

Average Treatment Effect on the Treated (ATT):

$$\widehat{ATT}_t = \frac{1}{N_{tr}} \sum_{i \in \mathcal{N}} \widehat{\tau}_{it}. \tag{3}$$

Average GVA for never vs not yet treated



- Need for **robustness** against parallel trends
- Need to account for **heterogeneity**

Estimation of treatment effects

We fit the factor model (Athey et al., 2021):

$$y_{it} = L_{it} + \epsilon_{it}, \quad L_{it} = \lambda_i^\top f_t, \tag{4}$$

which in matrix notation writes

$$\boldsymbol{Y} = \boldsymbol{L}^* + \varepsilon, \quad \text{where } \mathbb{E}[\varepsilon | \boldsymbol{L}^*] = \boldsymbol{0}.$$

- The goal is to estimate $\boldsymbol{L}^*$.

Key Assumptions

- **Identification relies on two main assumptions:**
  - **Strict exogeneity of treatment:** Treatment is uncorrelated with the idiosyncratic error term after controlling for latent factors and observed covariates X_{it}, ruling out endogenous selection.
  - **No interference between units:** A unit's potential outcomes do not directly depend on the treatment of other units.
- **Partial relaxation of interference:** Interactive fixed effects allow for some correlated shocks (e.g., trade linkages or coordinated policy), partially addressing cross-unit dependence.

These assumptions ensure the causal interpretation of $\widehat{\tau}_{it}$.

The minimization problem

Formally, the estimator is the following:

$$\min_{\boldsymbol{L},\boldsymbol{\Gamma},\boldsymbol{\Delta}} \left\{ \frac{1}{|\mathcal{O}|} \|\boldsymbol{P}_{\mathcal{O}}(\boldsymbol{Y} - \boldsymbol{L} - \boldsymbol{\Gamma}\mathbf{1}_T^\top - \mathbf{1}_N\boldsymbol{\Delta}^\top)\|_F^2 + \lambda\|\boldsymbol{L}\|_* \right\},$$

where $\|\boldsymbol{L}\|_* = \|\boldsymbol{L}\|_1^S = \sum_i \sigma_i(\boldsymbol{L})$.
The solution $\hat{\boldsymbol{L}}$ is combined with $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{\Delta}}$ to recover $\boldsymbol{L}^*$.

Practically, the minimization problem is the following:

$$\min_{\delta_i,\gamma_t,L} \sum_{i=1}^{N}\sum_{i=1}^{T}(1 - D_{it})(Y_{it} - \delta_i - \gamma_t)^2 + \lambda\|\boldsymbol{L}\|_*$$

References

Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association 116*(536), 1716–1730.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*, 1229–1279.

Becker, S., P. H. Egger, and M. von Ehrlich (2010). Going NUTS: The effect of EU structural funds on regional performance. *Journal of Public Economics 94(9-10)*, 578–590.

Bibliography ii

Becker, S., P. H. Egger, and M. von Ehrlich (2012). Too much of a good thing? on the growth effects of the EU's regional policy. *European Economic Review 56(4)*, 648–668.

Becker, S., P. H. Egger, and M. von Ehrlich (2013). Absorptive capacity and the growth effects of regional transfers: A regression discontinuity design with heterogeneous treatment effects. *American Economic Journal: Economic Policy 5(4)*, 29–77.

Becker, S., P. H. Egger, and M. von Ehrlich (2018). Effects of EU regional policy: 1989-2013. *Regional Science and Urban Economics 69*, 143–152.

Cerqua, A. and G. Pellegrini (2018). Are we spending too much to grow? the case of structural funds. *Journal of Regional Science 58*, 535–563.

Bibliography iii

Crescenzi, R. and M. Giua (2020). One or many cohesion policies of the european union? on the differential economic impacts of cohesion policy across member states. *Regional Studies 54(1)*, 10–20.

Hagen, T. and P. Mohl (2008). Which is the right dose of eu cohesion policy for economic growth? *ZEW Discussion Papers, ZEW - Leibniz Centre for European Economic Research 08-104*.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association 81*(396), 945–960.

Pellegrini, G., F. Terribile, O. Tarola, T. Muccigrosso, and F. Busillo (2013). Does EU regional policy enhance growth? *Papers in Regional Science 92*, 217–233.

Pesaran, M. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*, 967–1012.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis 25*(1), 57–76.