Causal inference with spillover effects to evaluate the EU's Cohesion Fund

Angelos Alexopoulos[1], Ilias Kostarakos[2], Pantelis Samartsidis[3], Petros Varthalitis[1]

CRETE Conference, Ios, July 9, 2026

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

Outline

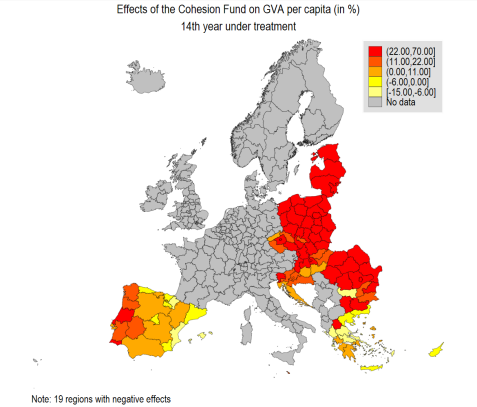
- Motivation and research question
- Data
- Empirical causal model
- Results and discussion

Cohesion policy

- **Cohesion policy** is one of the largest items of the EU budget and has disbursed almost **€960 billion** to EU regions.
- Its objective is to reduce regional disparities and support less-developed regions.
- We focus on the **Cohesion Fund** (CF), one of the main cohesion-policy instruments.
- Given its scale and redistributive purpose, its **effectiveness** remains a central empirical and policy question.

Starting point: spatially heterogeneous effects

Previous estimates (Alexopoulos et al., 2025) show that CF effects are not uniform across European regions.



Regional GVA effects from previous direct-effect analysis

- Effects vary substantially across regions.
- Some high-effect regions are geographically close to other treated regions.
- Regional policies may therefore affect outcomes beyond administrative borders.

Heterogeneous spatial patterns motivate a spillover-aware causal framework.

Research question

We ask whether Cohesion Fund transfers affect regional economic outcomes, and whether these effects extend beyond recipient regions.

Main question
What are the direct and spillover causal effects of the CF on investment and employment?

Research question

We ask whether Cohesion Fund transfers affect regional economic outcomes, and whether these effects extend beyond recipient regions.

Main question
What are the direct and spillover causal effects of the CF on investment and employment?

We estimate:

- **Direct effects**: impact of a region's own CF transfers.
- **Dynamic effects**: how effects evolve over time.
- **Heterogeneous effects**: how effects differ across regions.
- **Spillover effects**: how CF transfers affect connected regions.

Spillovers and no interference

The CF is not only a local transfer program. Funded projects may affect economic activity beyond recipient regions.

Examples include:

- transport infrastructure improving access for nearby regions;
- demand linkages through purchases from nearby firms;
- commuting and labour-market connections;
- productivity or market-access spillovers;
- possible displacement if treated regions attract resources away from others.

Spillovers and no interference

The CF is not only a local transfer program. Funded projects may affect economic activity beyond recipient regions.

Examples include:

- transport infrastructure improving access for nearby regions;
- demand linkages through purchases from nearby firms;
- commuting and labour-market connections;
- productivity or market-access spillovers;
- possible displacement if treated regions attract resources away from others.

Standard causal evaluations often assume no interference: a region's potential outcome depends only on its own treatment. With regional spillovers, outcomes may also depend on treatment exposure in connected regions.

Related literature and gap

The empirical literature on EU cohesion policy has mainly relied on:

- **RDD and IV designs** exploiting eligibility rules or institutional variation; see, e.g., Becker et al. (2013), Pellegrini et al. (2013), Becker et al. (2018), and Cerqua and Pellegrini (2018).
- **Dose-response approaches** estimating how transfer intensity relates to regional outcomes; see, e.g., Mohl and Hagen (2008), Becker et al. (2012), and Crescenzi and Giua (2020).
- **Synthetic-control and factor-model approaches** constructing counterfactual regional outcomes in settings with heterogeneous timing (Alexopoulos et al., 2025).

Related literature and gap

The empirical literature on EU cohesion policy has mainly relied on:

- **RDD and IV designs** exploiting eligibility rules or institutional variation; see, e.g., Becker et al. (2013), Pellegrini et al. (2013), Becker et al. (2018), and Cerqua and Pellegrini (2018).
- **Dose-response approaches** estimating how transfer intensity relates to regional outcomes; see, e.g., Mohl and Hagen (2008), Becker et al. (2012), and Crescenzi and Giua (2020).
- **Synthetic-control and factor-model approaches** constructing counterfactual regional outcomes in settings with heterogeneous timing (Alexopoulos et al., 2025).

Gap
Existing approaches typically focus on direct effects and often impose limited forms of no interference.

Contribution

We develop an empirical framework to evaluate the Cohesion Fund allowing for dynamic, heterogeneous, and spillover effects across European regions.

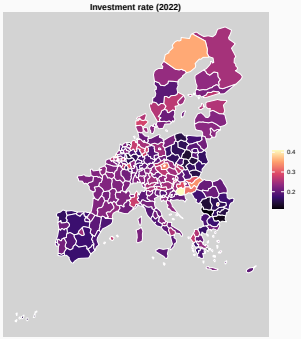
Our contribution is to:

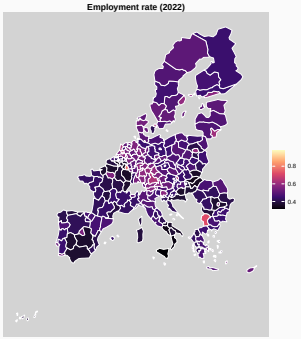
- estimate **region- and time-specific direct effects** on investment and employment.
- relax **no interference** by allowing outcomes to depend on exposure in connected regions.
- account for **unobserved time-varying confounding**.
- improve efficiency by exploiting **multiple regional outcomes**.
- characterize **uncertainty** around direct and spillover effects.

Data and outcomes

Cohesion Fund payments from Cohesion portal with regional macro data from **ARDECO**.

- **Sample**: 242 NUTS2 regions across 27 European countries, 1980–2022.
- **Main outcomes**: Gross Fixed Capital Formation and Employment Rate
- **Treatment**: Cohesion Fund exposure at the regional level.

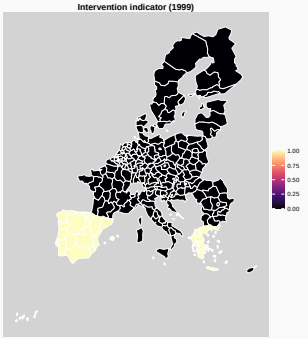


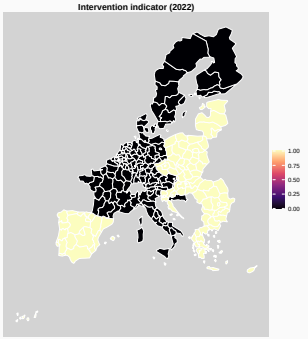


Treatment timing

The treatment is **staggered** across regions and persistent once region enters.

- **Treated regions**: 96 NUTS2 regions over 1993–2022.
- **Treatment duration**: average of 23.9 years under treatment.
- **Timing**: 18 regions treated throughout the sample; 78 regions enter later.





Empirical causal model

Estimands of interest: exposure mappings

To make spillovers empirically tractable we define

$$\mathbf{d}_{i,t} = \left(\mathbf{1}\{D_{i,t} > 0\}, \sum_{\ell=1}^{t} D_{i,\ell} \right),$$

$$\delta_{i,t} = \left(\mathbf{1}\left\{ \sum_{j \sim i} D_{j,t} > 0 \right\}, \sum_{j \sim i} \mathbf{1}\{D_{j,t} > 0\}, \frac{1}{N_i} \sum_{j \sim i} \sum_{\ell=1}^{t} D_{j,\ell} \right).$$

$\mathbf{d}_{i,t}$: region i's treatment path; $\delta_{i,t}$: its exposure to connected regions; $j \sim i$ denotes its N_i neighbours.

Potential outcomes are indexed as: $Y_{i,t}(\mathbf{d}_{i,t}, \delta_{i,t})$.

Estimands

1. **Direct effect**. For region i and t such that $\mathbf{d}_{i,t} \neq \mathbf{0}$

$$Y_{i,t}\left(\mathbf{d}_{i,t}, \delta_{i,t}\right) - Y_{i,t}\left(\mathbf{0}, \delta_{i,t}\right)$$

2. **Spillover effect**. For region i and t such that $\delta_{i,t} \neq \mathbf{0}$

$$Y_{i,t}\left(\mathbf{d}_{i,t}, \delta_{i,t}\right) - Y_{i,t}\left(\mathbf{d}_{i,t}, \mathbf{0}\right)$$

3. **Total effect**. For region i and t such that $\mathbf{d}_{i,t} \neq \mathbf{0}$ and $\delta_{i,t} \neq \mathbf{0}$

$$Y_{i,t}\left(\mathbf{d}_{i,t}, \delta_{i,t}\right) - Y_{i,t}\left(\mathbf{0}, \mathbf{0}\right)$$

Estimands

1. **Direct effect**. For region i and t such that $\mathbf{d}_{i,t} \neq \mathbf{0}$

$$Y_{i,t}\left(\mathbf{d}_{i,t}, \delta_{i,t}\right) - Y_{i,t}\left(\mathbf{0}, \delta_{i,t}\right)$$

2. **Spillover effect**. For region i and t such that $\delta_{i,t} \neq \mathbf{0}$

$$Y_{i,t}\left(\mathbf{d}_{i,t}, \delta_{i,t}\right) - Y_{i,t}\left(\mathbf{d}_{i,t}, \mathbf{0}\right)$$

3. **Total effect**. For region i and t such that $\mathbf{d}_{i,t} \neq \mathbf{0}$ and $\delta_{i,t} \neq \mathbf{0}$

$$Y_{i,t}\left(\mathbf{d}_{i,t}, \delta_{i,t}\right) - Y_{i,t}\left(\mathbf{0}, \mathbf{0}\right)$$

- **Consistency**: $Y_{i,t} = \sum_{\mathbf{d}_t} \sum_{\delta_t} \mathbf{1}\{\mathbf{D}_{i,t} = \mathbf{d}_t\}\mathbf{1}\{\mathbf{\Delta}_{i,t} = \delta_t\} Y_{i,t}\left(\mathbf{d}_t, \delta_t\right)$
- Missing potential outcomes (**counterfactuals**) need to be imputed

Counterfactual estimation

Let $\mathcal{Y}^{mis}$ denote the missing potential outcomes needed to compute direct, spillover, and total effects.

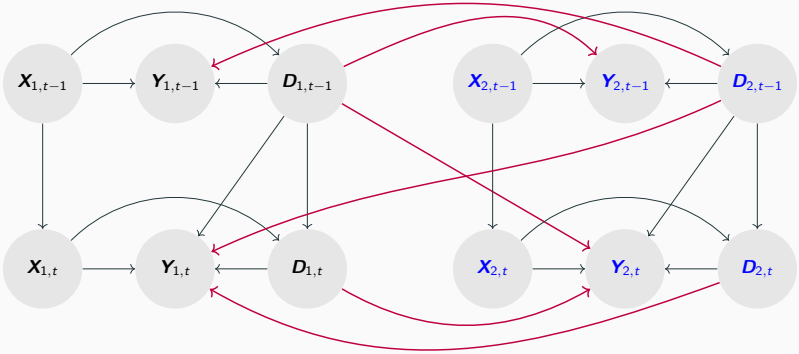
We use Bayesian posterior prediction:

$$\pi\left(\mathcal{Y}^{mis} \mid \text{data}\right) \propto \int_{\Theta} \pi\left(\mathcal{Y}^{mis} \mid \text{data}, \theta\right) \pi\left(\theta \mid \text{data}\right) d\theta.$$

1. Specify a model for observed and missing potential outcomes.
2. Estimate the posterior distribution of the model parameters.
3. Draw missing potential outcomes from the posterior predictive distribution.

Interpretation
The causal problem becomes a counterfactual-imputation problem.

Structural assumptions



We can now write that

$$P(\text{data}) = \prod_{s=1}^{2} \Big[P(\mathbf{X}_{s,1})\, P(D_{s,1} \mid \mathbf{X}_{s,1})\, P(\mathbf{X}_{s,2} \mid \mathbf{X}_{s,1})\, P(\mathbf{Y}_{s,1} \mid \mathbf{X}_{s,1}, \mathbf{D}_1) \times P(D_{s,2} \mid \mathbf{X}_{s,2}, D_{s,1})\, P\left(Y_{s,2} \mid \bar{\mathbf{D}}_1, \bar{\mathbf{D}}_2, \mathbf{X}_{s,2}\right) \Big]$$

Outcome model and latent confounding

For each outcome $k = 1, \ldots, K$, we model the potential outcome as:

$$Y_{i,k,t}\left(\mathbf{d}_{i,t}, \delta_{i,t}\right) \sim \text{Normal}\left[g_k\left(\mathbf{d}_{i,t}, \delta_{i,t}, \mathbf{x}_{i,t}, \mathbf{u}_i\right), \sigma_k^2\left(\mathbf{d}_{i,t}, \delta_{i,t}\right)\right].$$

- $\mathbf{d}_{i,t}$ captures own CF treatment history.
- $\delta_{i,t}$ captures exposure to CF treatment in connected regions.
- $\mathbf{x}_{i,t}$ are observed regional covariates.
- $\mathbf{u}_i$ are latent region-specific confounders.
- $g_k(\cdot)$ captures heterogeneous and non-linear treatment-response patterns.

Key idea
The model imputes missing counterfactual outcomes while allowing for direct effects, spillovers, latent confounding, and multiple outcomes.

Relation to standard TWFE models

A standard two-way fixed-effect model can be written as:

$$Y_{i,t} = \alpha_i + \lambda_t + \beta D_{i,t} + \varepsilon_{i,t}.$$

This imposes:

- a single average treatment effect β;
- no spillovers across regions;
- additive unit and time effects;
- limited flexibility for dynamic and heterogeneous effects.

Our model generalizes this structure:

$$Y_{i,k,t} = g_k\left(\mathbf{d}_{i,t}, \delta_{i,t}, \mathbf{x}_{i,t}, \mathbf{u}_i\right) + \varepsilon_{i,k,t}.$$

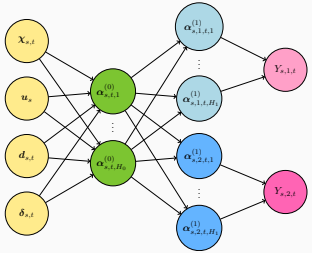
Compared to TWFE we allow treatment effects to vary over time, across regions, across outcomes, and through spillover exposure.

Neural-network mean model

We model the conditional mean flexibly using a neural network:

$$\mathbb{E}[Y_{i,k,t}\left(\mathbf{d}_{i,t}, \delta_{i,t}\right) \mid \mathbf{x}_{i,t}, \mathbf{u}_i] = g_k\left(\mathbf{d}_{i,t}, \delta_{i,t}, \mathbf{x}_{i,t}, \mathbf{u}_i\right).$$

- Captures **non-linear direct and spillover effects** of CF exposure.
- Allows **interactions** between own treatment, neighbour exposure, and regional characteristics.
- Allows **heterogeneous effects** across regions and outcomes.



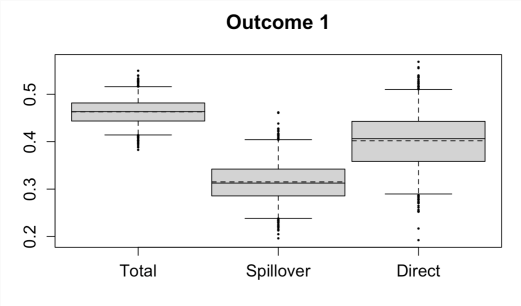
Neural networks provide flexible function approximation for non-linear prediction and treatment-effect heterogeneity.

See, e.g., Athey and Imbens (2019) and Farrell et al. (2021).

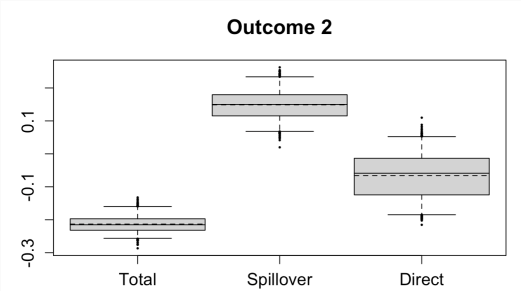
Results

Distribution of estimated effects

We first summarize the posterior distribution of **direct**, **spillover**, and **total** effects across regions.



Gross Fixed Capital Formation



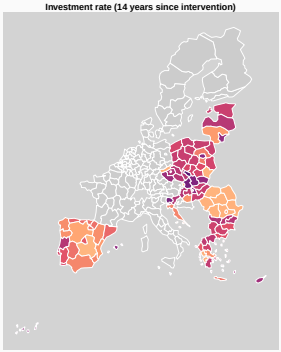
Employment rate

Investment effects are predominantly **positive**, suggesting a clear response of GFCF to CF.

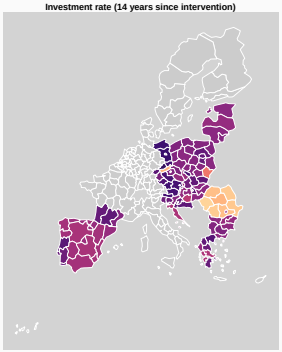
Employment effects are more mixed and tend to be **negative** for the direct and total components, while spillover effects appear more positive.

Regional effects on GFCF: 14 years after treatment

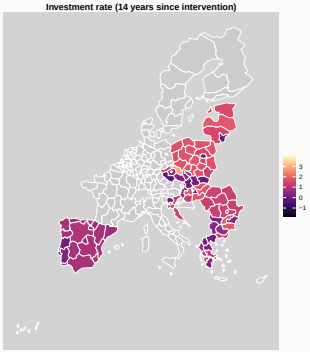
Direct, **spillover**, and **total** effects on GFCF.



Direct



Spillover

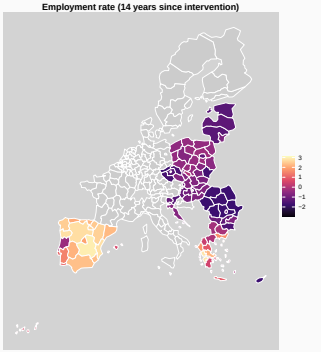


Total

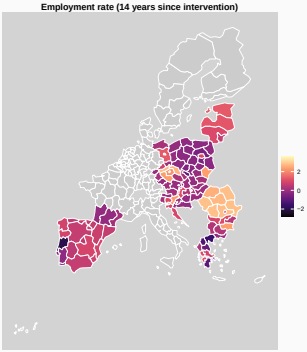
For GFCF, the maps show that both direct and spillover effects contribute to the overall total effect, with substantial spatial heterogeneity across regions.

Regional effects on employment: 14 years after treatment

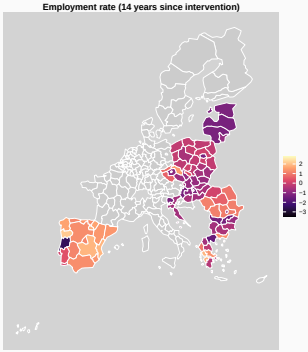
Direct, **spillover**, and **total** effects on employment.



Direct



Spillover

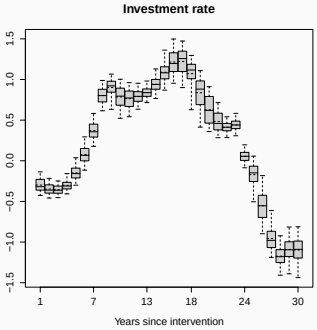


Total

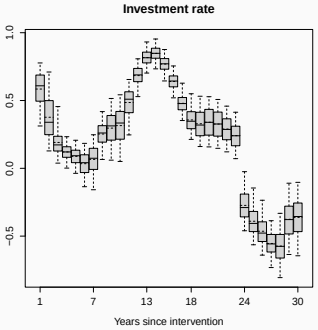
For employment, the total effect combines a negative direct component with a positive spillover component, leading to strong heterogeneity across regions.

Dynamic effects on GFCF

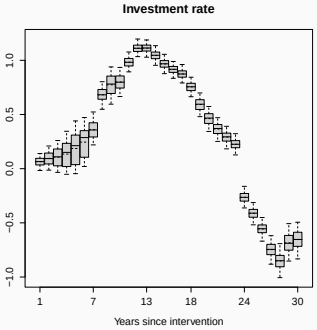
The figure shows ATT_t compare the evolution of **direct**, **spillover**, and **total** effects on investment (GFCF) over event time.



Direct



Spillover

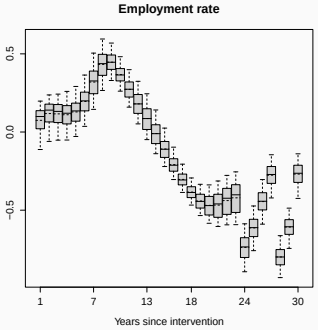


Total

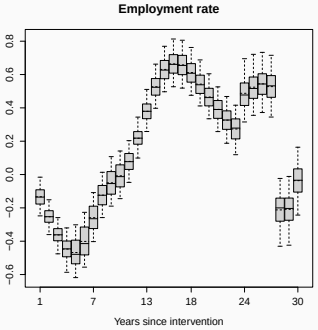
GFCF effects rise after treatment, peak in the medium run, and decline later.

Dynamic effects on employment

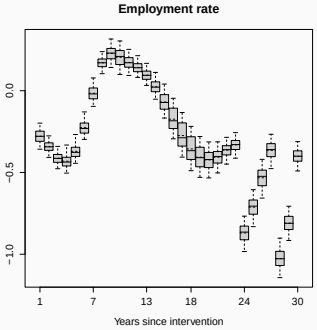
The figure compares the evolution of **direct**, **spillover**, and **total** effects on employment over event time.



Direct



Spillover



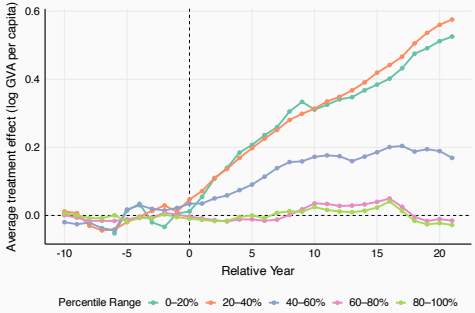
Total

Employment effects are mixed: spillovers appear positive over much of the horizon, while direct effects are weaker or negative in later years.

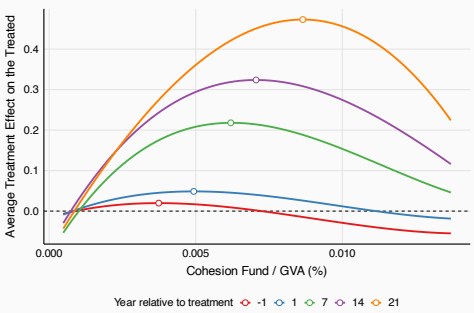
What comes next?

We have separated the effects of the CF into **direct**, **spillover**, and **total**.

A natural next step is to follow (Alexopoulos et al., 2025) and study whether these components are also **heterogeneous** and **non-linear**.



Heterogeneity by pre-treatment income group



Non-linear effects of CF intensity

Extend the spillover framework to examine whether **direct**, **spillover**, and **total** effects vary by initial development, treatment intensity, and regional connectivity.

Discussion and conclusions

- We propose a spillover-aware framework for evaluating the EU Cohesion Fund.
- The approach distinguishes between:
  - **direct effects** from a region's own CF exposure;
  - **spillover effects** from CF exposure in connected regions;
  - **total effects** combining both channels.
- The results suggest that CF effects on **investment** are positive and dynamic, consistent with a capital-accumulation channel.
- For **employment**, the evidence is more nuanced: spillover effects are positive, while direct effects are weaker or negative in later periods.
- Spatial patterns show that effects are **heterogeneous across regions**, reinforcing the need to account for regional linkages.
- **Bottom line**: spillovers are not just a nuisance; they are part of the policy effect.

Thank you!

Appendix

Identification

Using SWIGs, we can conclude that

$$P\left(Y_{s,2} \mid \bar{\mathbf{D}}_1 = \bar{\mathbf{d}}_1, \bar{\mathbf{D}}_2 = \bar{\mathbf{d}}_2, \mathbf{X}_{s,2}\right) = P\left(Y_{s,2}(\bar{\mathbf{d}}_1, \bar{\mathbf{d}}_2) \mid \bar{\mathbf{D}}_1 = \bar{\mathbf{d}}_1, \bar{\mathbf{D}}_2 = \bar{\mathbf{d}}_2, \mathbf{X}_{s,2}\right)$$
$$= P\left(Y_{s,2}(\bar{\mathbf{d}}_1, \bar{\mathbf{d}}_2) \mid \mathbf{X}_{s,2}\right)$$

In other words, we can **identify** the distribution of the potential outcome from the observed data

Need to specify

- Prior $\pi\left(\{U_s\} \mid \theta_u\right) \pi\left(\theta_u\right) \pi\left(\theta_d\right) \pi\left(\theta_y\right)$
- Potential outcomes model $\pi\left(Y_{i,t}(\mathbf{d}, \delta) \mid \mathbf{X}_{i,t}, \bar{\mathbf{D}}_t, \mathbf{U}_s, \theta_y\right)$
- **Treatment assignment** $\pi\left(D_{i,t} \mid \mathbf{X}_{i,t}, \bar{\mathbf{D}}_{s,t-1}, \mathbf{U}_s, \theta_d\right)$

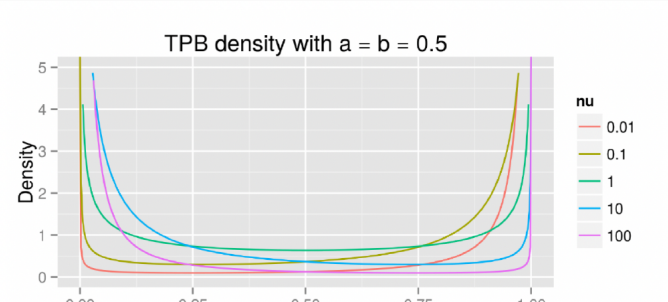
Prior on the network coefficients

Let $\{\{\beta_{i,j}\}_{i=0}^{H_0}\}_{j=1}^{H_1}$ be the first layer regression coefficients

We use a global-local style **shrinkage** prior[1]

$$\beta_{i,j} \sim \text{Normal}\left(0, \frac{1}{\phi_{i,j}} - 1\right), \qquad \phi_{i,j} \sim \text{TPB}\left(0.5, 0.5, \frac{1}{\zeta_j} - 1\right)$$

$$\zeta_j \sim \text{TPB}\left(0.5, 0.5, \frac{1}{\varrho} - 1\right), \qquad \varrho \sim \text{TPB}\left(0.5, 0.5, \nu\right)$$



Prior on the hidden confounders

For each $j = 1, \ldots, D_u$, let $\upsilon_j = (u_{1,j}, u_{2,j}, \ldots, u_{S,j})^\top$

We assign the Leroux conditional autogressive model i.e.

$$\upsilon_j \sim \text{Normal}\left(\mathbf{0}, \zeta_j \boldsymbol{Q}(\varphi_j)^{-1}\right),$$

where

$$\boldsymbol{Q}(\varphi_j) = \varphi_j(\boldsymbol{D} - \boldsymbol{W}) + (1 - \varphi_j)\boldsymbol{I},$$

- $\boldsymbol{D}$ is the $S \times S$ diagonal matrix with entries N_s (number of neighbours)
- $\boldsymbol{W}$ is the $S \times S$ adjacency matrix
- $\varphi_j \sim \text{Beta}(3, 2)$
- $\zeta_j = 1$

Intervention assignment model

$\pi\left(D_{i,t} \mid \bar{\mathbf{D}}_{s,t-1}, \mathbf{X}_{i,t}, \mathbf{U}_s, \theta_u\right)$ involves the hidden confounders $\mathbf{U}_i$: **non-ignorable**

Potential **drawbacks** of modelling the intervention assignment

- Computation
- Little information on $\mathbf{U}_s$
- Potential **misspecification**

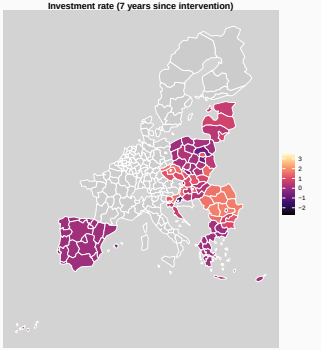
We typically ignore likelihood terms: **cut-posterior**[2] approach

Generally good asymptotic properties[3]

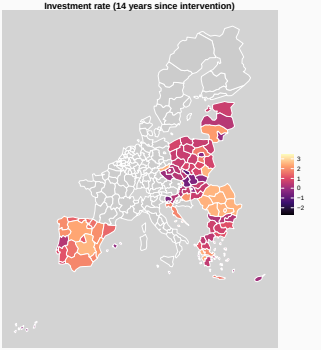
[2] Plummer, *Stat Comput* 25 (1), 37-43, 2015.
[3] Pompe & Jacob, arXiv:2110.11149, 2021.

Direct effects on GFCF

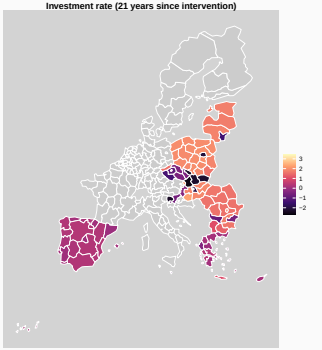
The maps show the spatial distribution of **direct effects on Gross Fixed Capital Formation** at selected horizons after treatment.



7 years since treatment



14 years since treatment

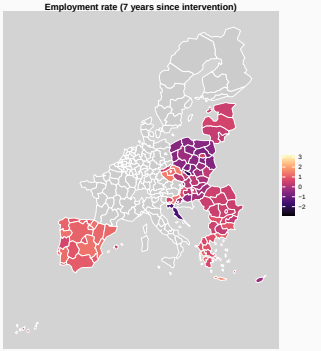


21 years since treatment

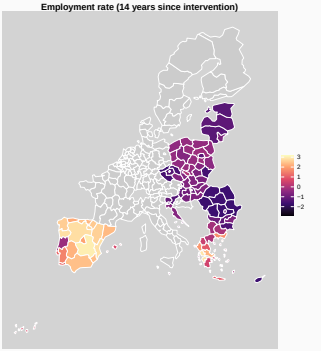
Direct effects capture the impact of a region's own Cohesion Fund exposure, holding spillover exposure fixed.

Direct effects on employment

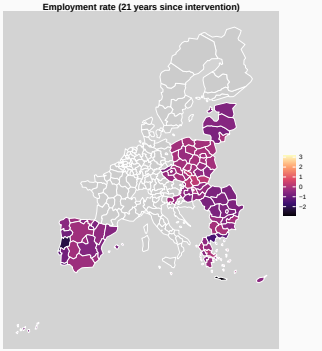
The maps show the spatial distribution of **direct effects on employment** at selected horizons after treatment.



7 years since treatment



14 years since treatment

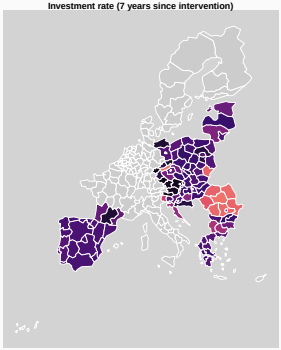


21 years since treatment

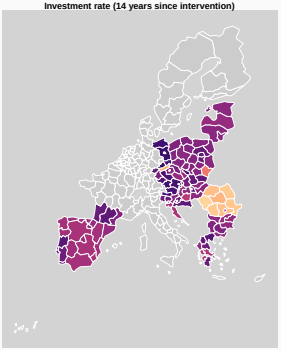
Direct effects capture the impact of a region's own Cohesion Fund exposure, holding spillover exposure fixed.

Spillover effects on GFCF

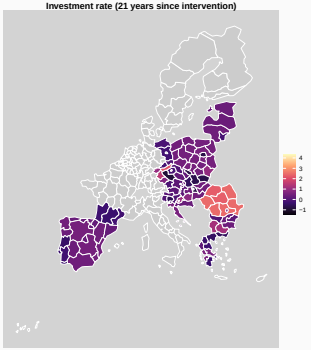
The maps show the spatial distribution of **spillover effects on Gross Fixed Capital Formation** at selected horizons after treatment.



7 years since treatment



14 years since treatment

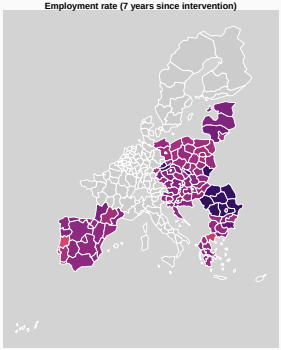


21 years since treatment

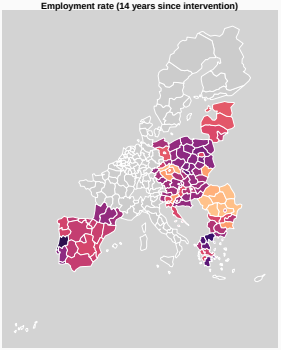
Spillover effects capture the impact of Cohesion Fund exposure in connected regions, holding own exposure fixed.

Spillover effects on employment

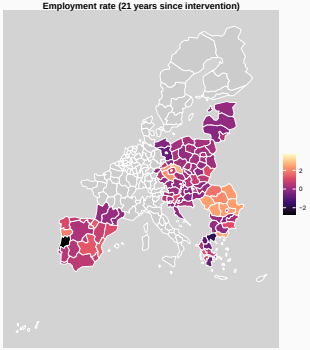
The maps show the spatial distribution of **spillover effects on employment** at selected horizons after treatment.



7 years since treatment



14 years since treatment



21 years since treatment

Spillover effects capture the impact of Cohesion Fund exposure in connected regions, holding own exposure fixed.

References

Alexopoulos, A., I. Kostarakos, C. Mylonakis, and P. Varthalitis (2025). The heterogeneous causal effects of the eu's cohesion fund. *arXiv preprint arXiv:2504.13223*.

Athey, S. and G. W. Imbens (2019). Machine learning methods that economists should know about. *Annual Review of Economics 11*(1), 685–725.

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.

Bibliography ii

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.

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.

Farrell, M. H., T. Liang, and S. Misra (2021). Deep neural networks for estimation and inference. *Econometrica 89*(1), 181–213.

Mohl, P. and T. Hagen (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*.

Pellegrini, G., F. Terribile, O. Tarola, T. Muccigrosso, and F. Busillo (2013). Measuring the effects of european regional policy on economic growth: A regression discontinuity approach. *Papers in Regional Science 92*, 217–233.