Case Studies in Microeconomic Evaluation
Data and methods for learning what works

Case study 1

Competence Centre on Microeconomic Evaluation (CC-ME)
Joint Research Centre
16th Nov 2022

A Brief Introduction to causality and impact evaluation

Evidence-based policy making

- **Evidence-based policy making**: focus on outcomes and results rather than inputs
- Results are increasingly being used and required to enhance accountability and to guide policy decisions
- **Monitoring and impact evaluation are at the heart of evidence-based policy making** $\rightarrow$ they provide proper tools to assess the quality, efficiency and effectiveness of public policies
- This is crucial in a context in which policy makers and civil society are demanding accountability of public programs

Impact evaluation

- **Impact evaluation**: particular type of evaluation that strives to assess specific cause-and-effect questions
  $\rightarrow$ **What is the causal effect (impact) of the program on the outcome of interest?**
- Focus on the **impact**
- Changes directly attributable to the program
- Focus on **causality**

Evaluation and Impact evaluation

- Impact evaluation informs on whether a program has achieved its desired outcomes $\Rightarrow$ it deals with the **effectiveness** of the program

- In particular, it assesses whether the changes can be attributed to the policy under scrutiny
  $\rightarrow$ the central challenge in carrying out effective impact evaluations is to identify **the causal relationship between the policy and the outcomes of interest**

Causality

- Causality is what impact evaluation strives for
  $\rightarrow$ **all impact evaluation methods address some form of cause-and-effect question!**

- Cause-and-effect relationship examples:

  Does teacher training improve students' test scores?
  Do conditional cash transfer programs cause better health outcomes in children?
  Do vocational training programs increase trainees' incomes?

Causality

- Causal effects can only be retrieved by estimating the so-called **counterfactual** $\rightarrow$ **what the outcome would have been for program participants had they not participated in the program?**
- In practice, the point of all impact evaluation methods is to find a suitable **comparison group** to estimate what would have happened to the program participants without the program, and then to make comparisons with the treatment group that has received the program
- **Why?**
  Because we need to rule out the possibility that any factors other than the program of interest explain the observed impact!

The counterfactual

Consider the basic impact evaluation formula:

$$\Delta = (Y^{P=1}) - (Y^{P=0})$$

where Δ:causal impact; Y: outcome; P: program

- $(Y^{P=1})$ is the outcome obtained with the program
- $(Y^{P=0})$ is the outcome obtained without the program

Impact = (a person's outcome after participation in a program) - (the same person's outcome had she not participated in the program)

Problem: *we cannot observe the same unit under two different status*!!
While we can observe $(Y^{P=1})$, it is impossible to observe $(Y^{P=0})$ and there are no data that would enable us to measure it!!

$\rightarrow (Y^{P=0})$ represents the **counterfactual**

Estimating the counterfactual

- Estimate the counterfactual: generate two groups of units that are ***statistically*** indistinguishable from each other at the group level
- Given a ***treatment group***, the ***comparison group*** (control group) is the group of units that are ***statistically identical***, ***on average***, to the treatment group ***but for having participated into the program***
- **The counterfactual outcome** ($Y^{P=0}$) **would then be the outcome observed for this group of units**
  $\Rightarrow$ The challenge in any counterfactual impact evaluation is to identify a proper control group. If the two groups are identical with the sole exception that one group participates in the program and the other does not, *we can attribute any difference in outcome to the program participation*
- If the two groups differ for other characteristics rather than the participation, the estimation of the impact of the program through difference in outcomes would be **biased**

Choosing the right method

How to choose the right evaluation method?

The choice of the right method depends on the **operational characteristics of the program being evaluated**

Specifically:

1. Its available resources
2. Eligibility criteria for selecting beneficiaries
3. Timing of the implementation

Method 1: Difference-in-Differences

Evaluating the effects of an intervention

What would have happened to treated units in absence of the treatment?

- Our goal is to find the **Average Treatment on the Treated (ATT)**
- Ideally, we would like to observe *two parallel worlds*

Problem: We can observe only one of the two parallel worlds!

Evaluating the effects of an intervention

- Consider outcome Y

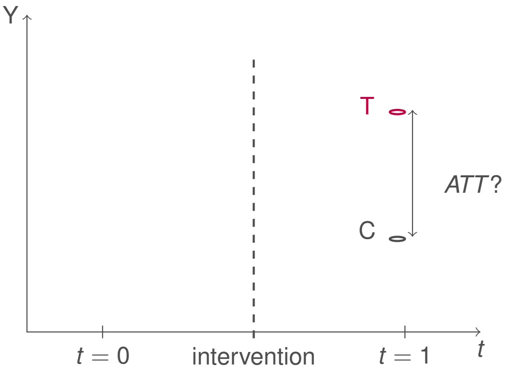
- We have 2 time periods
  - Time $t = 0$: **before** the intervention
  - Time $t = 1$: **after** the intervention

- We can identify 2 groups
  - Treatment group T: receives the intervention
  - Control group C: does not receive the intervention

Feasible but problematic solutions (1)

“Simple differences” estimator

Compares Treated units T and Non-Treated units C in post-intervention period ($t = 1$)

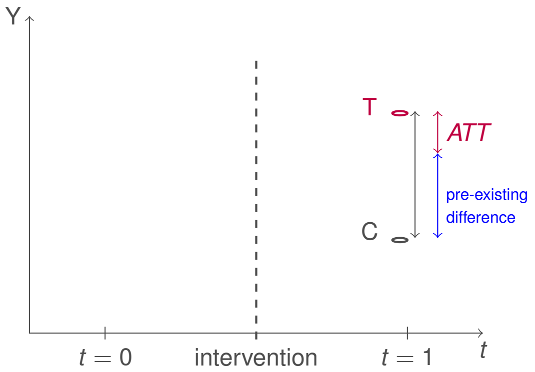


Feasible but problematic solutions (1)

“Simple differences” estimator
Compares Treated units T and Non-Treated units C in post-intervention period ($t = 1$)

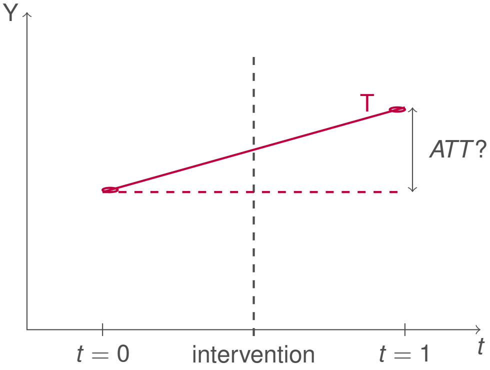
Problem: unobserved differences between treated and non-treated units that are correlated with outcomes influence the estimation of the effect

- Simple difference ignores pre-existing heterogeneity between T and C groups



Feasible but problematic solutions (2)

“Before-After” estimator
compares outcomes of treated units T before and after intervention,
i.e. $t = 0$ vs $t = 1$



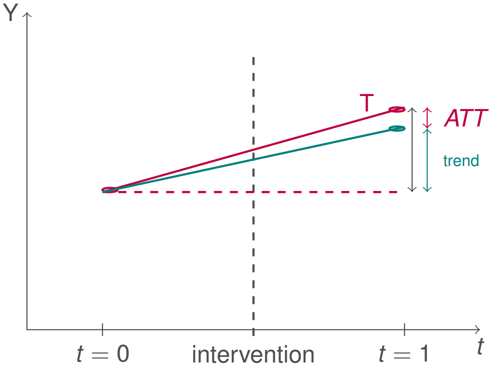
Feasible but problematic solutions (2)

"Before-After" estimator
compares outcomes of treated units T before and after intervention, i.e. $t = 0$ vs $t = 1$

Increasing time-trend

causes the effect of the intervention to be overestimated

- Before-after comparison ignores time-varying factors



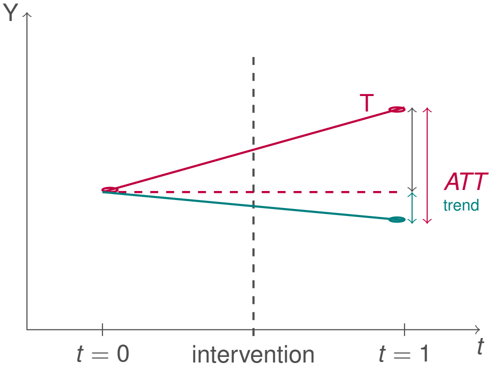
Feasible but problematic solutions (2)

"Before-After" estimator
compares outcomes of treated units T before and after intervention, i.e. $t = 0$ vs $t = 1$

Decreasing time-trend

causes the effect of the intervention to be underestimated

- Before-after comparison ignores time-varying factors



What then?

Combine the two: Difference-in-Differences (DiD)

- Take the mean value of each group's outcome **before** and **after** the intervention
- Compute the 'difference-in-differences' of the means

	Treatment Group (T)	Control Group (C)	Δ
Pre ($t = 0$)	T_0	C_0	
Post ($t = 1$)	T_1	C_1	
Change over time	$T_1 - T_0$	$C_1 - C_0$	$T_1 - T_0 - (C_1 - C_0)$ or, equivalently $T_1 - C_1 - (T_0 - C_0)$

Difference-in-Differences

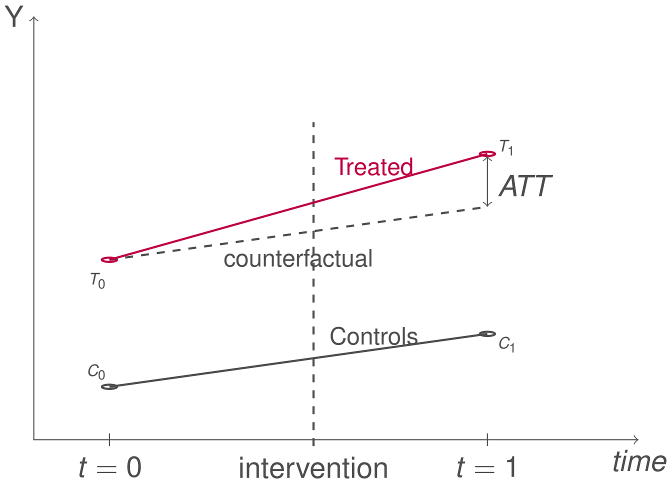
Definition
Difference-in-differences compares the changes in outcomes over time between units that are subject to the intervention (the **treatment group**) and units that are not (the **comparison or control group**).

This allows to correct for:

- pre-existing time-invariant differences across groups, and
- common time-trends

Can also include covariates, i.e. the effect can be netted out of other factors

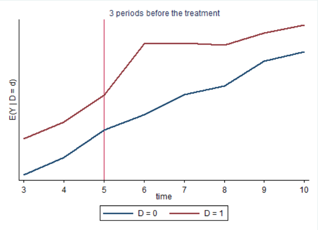
Difference-in-Differences



It's all about assumptions

DiD is a wonderful tool,
but crucially depends on the **credibility of the assumptions** in the specific case

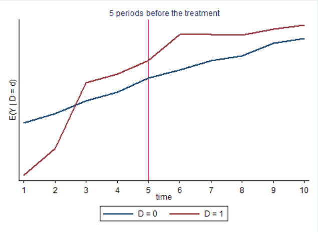
- We are "creating" a parallel world, it needs to make sense!
- The fundamental assumption is the **common trend**
  - Visual inspection of the evolution of Y in the two groups over time helps
  - Relatedly: the more periods you have (especially in the "pre" period) ... the better!



It’s all about assumptions

DiD is a wonderful tool,
but crucially depends on the **credibility of the assumptions** in the specific case

- We are “creating” a parallel world, it needs to make sense!
- The fundamental assumption is the **common trend**
  - Visual inspection of the evolution of Y in the two groups over time helps
  - Relatedly: the more periods you have (especially in the “pre” period) ... the better!



It’s all about assumptions

DiD is a wonderful tool,
but crucially depends on the **credibility of the assumptions** in the specific case

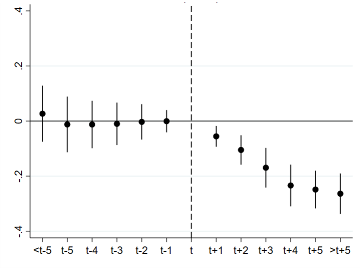
- We are “creating” a parallel world, it needs to make sense!
- The fundamental assumption is the **common trend**
  - Visual inspection of the evolution of Y in the two groups over time helps
  - Relatedly: the more periods you have (especially in the “pre” period) ... the better!

- Additionally, **no other change should occur** that systematically affects either group (treated or control)

It's all about assumptions

Unfortunately these assumptions cannot be formally tested, but...

- **Event-study analysis** can provide some informal testing
  - Tells you whether treated units behave differently from control units at each point in time (especially before the treatment)
  - Evidence of significant differences before the treatment are bad news for the common trend assumption



It’s all about assumptions

Unfortunately these assumptions cannot be formally tested, but...

- **Event-study analysis** can provide some informal testing
  - Tells you whether treated units behave differently from control units at each point in time (especially before the treatment)
  - Evidence of significant differences before the treatment are bad news for the common trend assumption
- **Unit-specific trend**
  - More time periods are required
  - We estimate a unit-specific trend (linear quadratic etc.)
  - Similar results (with vs without) are reassuring
- **Placebo tests**
  - “Move” artificially the intervention in time
  - Check the effect on similar but unaffected outcomes
  - Check the effect on a fake treatment group

Standard DiD

- A relatively identifiable group (T) receives the intervention (“treatment”) at time t
- Need to find a reasonable control group (C)
  - Control units expected to behave similarly to treated units in the absence of the treatment
  - Control units not subject to any type of intervention in the same period
- Need to gather data on <u>both T and C units, before and after the treatment</u>

Matching DiD

Sometimes you can select the control group (*C*) using a "matching procedure"

- Matching methods allow identifying the set of non-treated units that look more similar to the treated ones, based on the available observable characteristics
- The matched non-treated units become the control group
- A good match for each treated requires a large and complete set of data

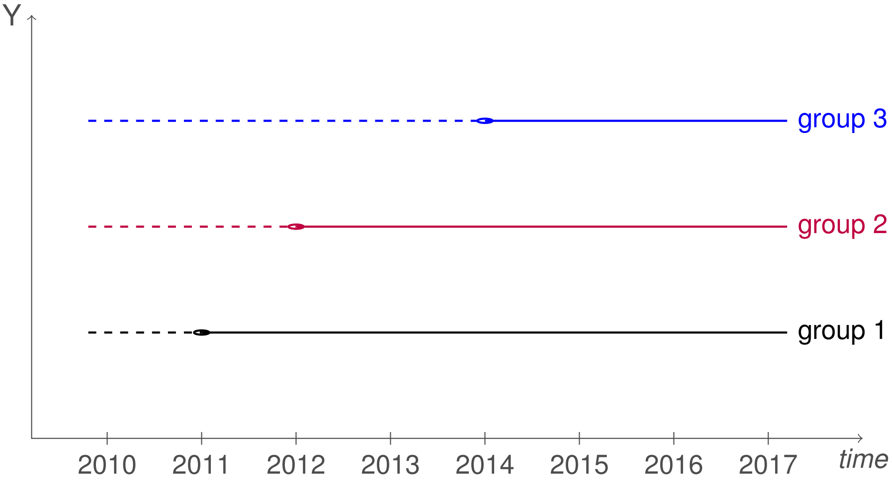
Treated units			
Age	Gender	Months unemployed	Secondary diploma
19	1	3	0
35	1	12	1
41	0	17	1
23	1	6	0
55	0	21	1
27	0	4	1
24	1	8	1
46	0	3	0
33	0	12	1
40	1	2	0

Untreated units			
Age	Gender	Months unemployed	Secondary diploma
24	1	8	1
38	0	1	0
58	1	7	1
21	0	2	1
34	1	20	0
41	0	17	1
46	0	9	0
41	0	11	1
19	1	3	0
27	0	4	0

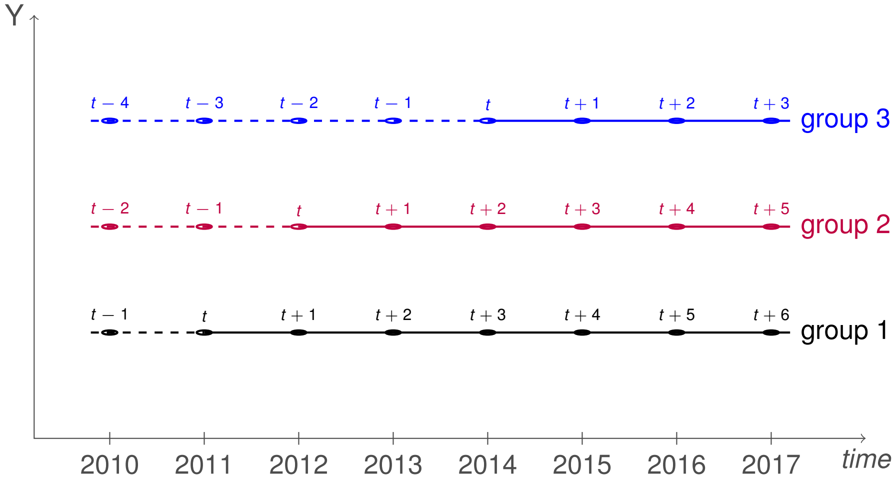
Staggered DiD

- You may have that everyone is eventually treated (T), i.e. receives the intervention ("treatment")
- As long as the treatment is staggered over time, you can identify the control group (C)
  - (Groups of) units are treated at different points in time
  - When a unit becomes treated, their control will be the units who are not yet treated
- Here, you may have "always treated", "never treated" and "switchers"
- Need to gather data on a sufficient number of "switches"

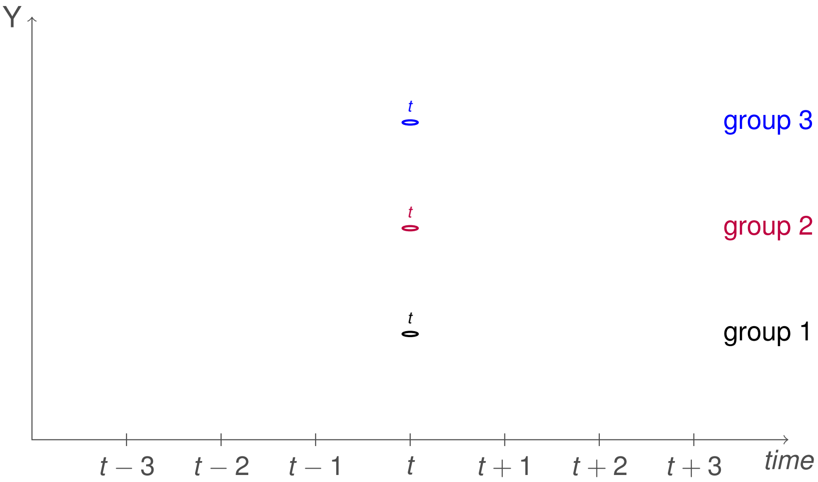
Staggered DiD: Alignment around zero



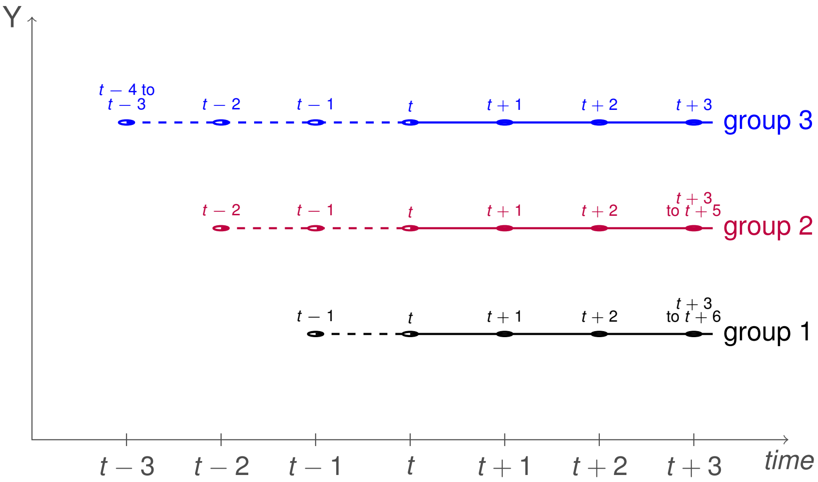
Staggered DiD: Alignment around zero



Staggered DiD: Alignment around zero



Staggered DiD: Alignment around zero



European Commission

Case Study 1: Assessing the economic impact of faster payments in B2B commercial transactions

Antonella Ferrara and Massimiliano Ferraresi

Motivation

Delayed payment in commercial transactions

- is an important concern for businesses
- determine higher cost and liquidity risks for the supplier
- might be particularly harmful for SMEs
- might yield to insolvency (account for one out of four bankruptcies in the EU)

$\Longrightarrow$ **In 2011 the EC adopted a recast of the Late Payment Directive**

Scope of the study

What: Estimating the impact of the LPD provisions in Business–to–Business (B2B) transactions on firms' outcomes, using firm level data in a panel of nine European countries over the period 2008-2018.

How: Exploiting the staggered adoption of the Directive by MS and grouping firms according to their different *degree of exposure* to the directive, using a **Difference-in-Differences** approach.

The Late Payment Directive (LPD)

- The Directive 2011/7/EU (LPD) regulates both commercial transactions between public authorities and businesses (PA2B) and among businesses (B2B)
- **General objectives:** i) contribute to the development of the Single Market; ii) improve the competitiveness of European enterprises and iii) eliminate barriers to cross-border commercial transactions.
- **Specific provisions:**
  1. set a 30-day target for payments in PA2B (with exceptions)
  2. set a 30-day target in B2B, which can be extended to up to a 60-day target, expressly agreed in the contract and as long as it is not grossly unfair to the creditor
  3. interests and compensation claims for late payments
  4. set a statutory interest rate for late payments in all MS of at least 8 percentage points above the ECB's reference rate.
  5. EU-MS lay down recovery procedures for undisputed claims to obtain enforceable titles within 90 calendar days

The expected effects

By stimulating more discipline in payment schedules, these measures were expected to:

- produce substantive improvements in enterprises cash flow
- reduce costs and prevent bankruptcy due to limited self-financing
- remove barriers to cross-border commercial transactions within EU boundaries
- reduce the cost for businesses, especially for SMEs

The impact of PA2B on firms' exit

In the empirical literature, little is still known on the effect of the LPD provisions on firms' performance in the EU. A recent study (Conti et al. 2021) focused on PA2B operations and showed that:

- Firms' exit rates fall relatively more in sectors that sell a larger fraction of their output to the government
- More pronounced effects were found in sectors with a large share of small firms, for countries characterized by longer payment delays
- Taken together, findings indicate that more discipline in governments' payment terms can have considerable effects on economic activity

The Data

Orbis database

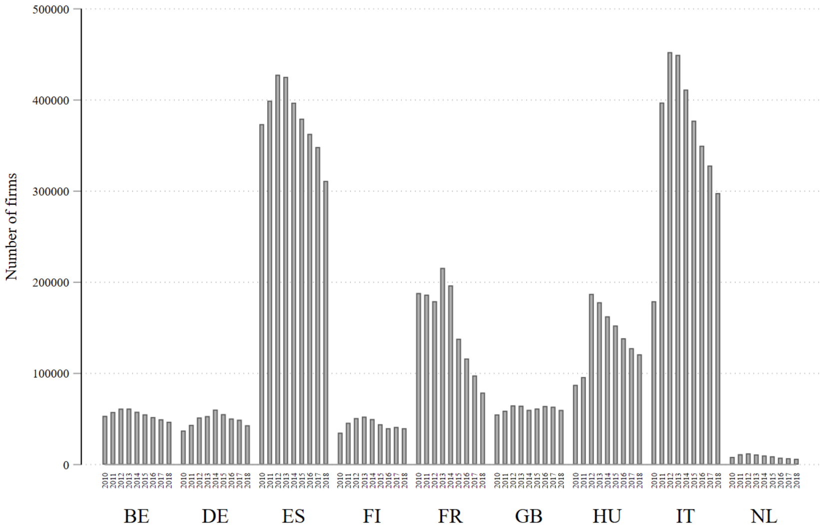
- It is the largest cross-country firm-level database available for economic and financial research
- The coverage varies by firm size, industry, over time and across variables within the data $\rightarrow$ Several limitations!
- The industry coverage reflects the non-farm business sector
- Data at hand covers approximately 10 years, over which it is possible to construct an unbalanced panel of firms
- Despite its well-known limitations, Orbis is one of the most comprehensive sources of information at the firm level

The sample

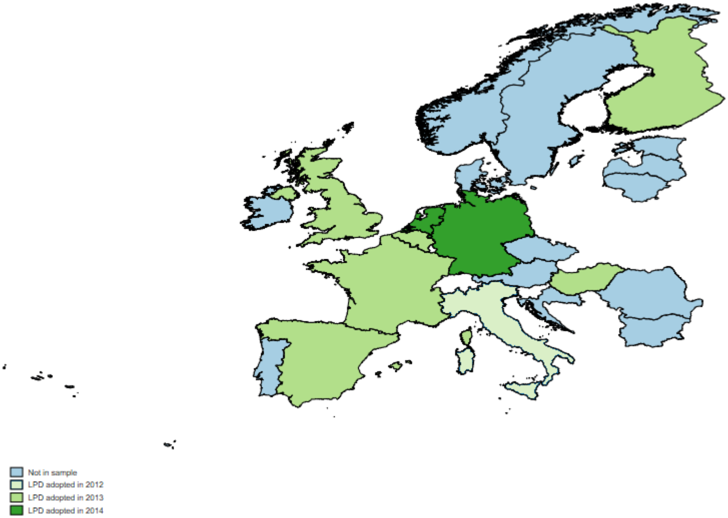
A European panel

- Final sample is based on specific recommendations provided by DG GROW A.2 and on the representativeness of ORBIS (Bajgar et al. 2020)
- Selected countries are:
  - Belgium
  - Finland
  - France
  - Germany
  - Italy
  - Spain
  - Hungary
  - the Netherlands
  - the United Kingdom
- Period 2010-2018; on average 1,772,059 observations per year

Sample composition



The LPD rollout



How to measure the impact

The empirical analysis builds on a **DiD model**

All firms were bound to comply with the LPD, hence disentangling treated and control units is not straightforward

Firms might **differ in their level of exposure to the treatment (*treatment intensity*)**, defined by the average number of days to collect credits (account receivables)

Firms more exposed: those having a higher average credits payment duration before the introduction of the LPD

Firms less exposed: firms that before the introduction of the directive were already collecting their credits in compliance with the LPD B2B provisions

Intuition: Firms that were collecting payments over a term far larger than that set by the LPD are likely to benefit the most from its adoption!

The exposure variable

Collection period days

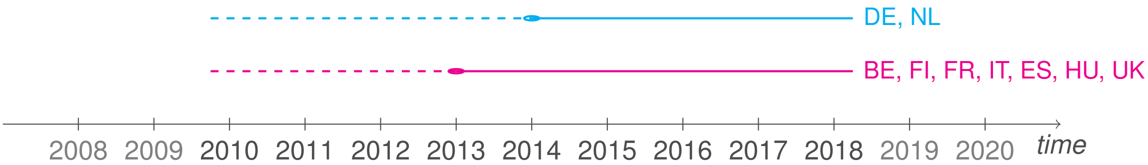
- Computed as the **ratio between Credits and Op Revenue** ($\times$ 360)
- Information about the **average nr of days** a firm takes to collect its credits
- To avoid endogeneity issues, the exposure indicator is defined as the mean of this indicator, over the period **2008-2010**
- The distribution shows very heterogeneous values, with the mean being around 4 months (**96 days**) and with a min of 0 and a max of 1000 days
- The analysis is restricted to firms that collect their credits **within 120 days** to mitigate the distortion due to potential outliers
- Results are sensitive neither to the threshold selected nor to the reference period used to compute the exposure variable

Model setting

- **Time-span:** 2010-2018
- **Treatment:** continuous measure based on firms' exposure to the LPD (i.e. credits pay duration in 2008-2010)
- **Impact:** changes in the outcome of firms that are exposed more to the LPD to those that are exposed less, before and after the implementation of the directive
- We account for firm-level controls and FE at firm, year and sector-by-year level SE double clustered at firm and country-by-year level

Model setting

- **Time-span:** 2010-2018
- **Treatment:** continuous measure based on firms' exposure to the LPD (i.e. credits pay duration in 2008-2010)
- **Impact:** changes in the outcome of firms that are exposed more to the LPD to those that are exposed less, before and after the implementation of the directive
- We account for firm-level controls and FE at firm, year and sector-by-year level



Days to collect credits relative to year (t-1)

Evolution over time of the average number of days to collect credits

Credits collection period, on average, **has decreased** (-22 days) after the implementation of the directive.

A **similar pattern** is confirmed in single-country analyses.

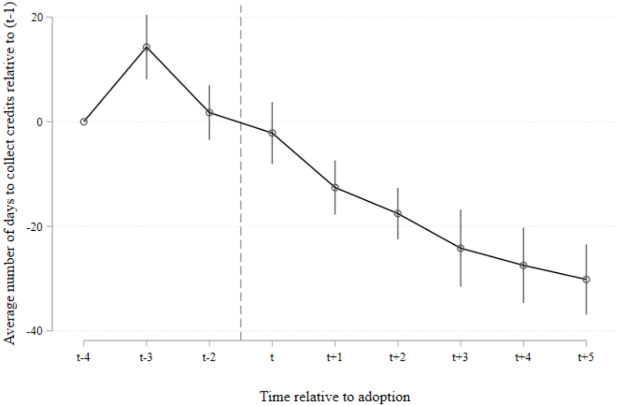


Figure: Effect on collection period days

Cash Flow trending

- Cash Flow considered as a proxy of company's capacity to create value
- The figure shows differences in Cash Flow trending between firms low–and–high exposed to the LPD
- Firms that were already taking more than 60 days to collect credit are considered "high exposed"
- The shaded area identifies the "adoption time window"

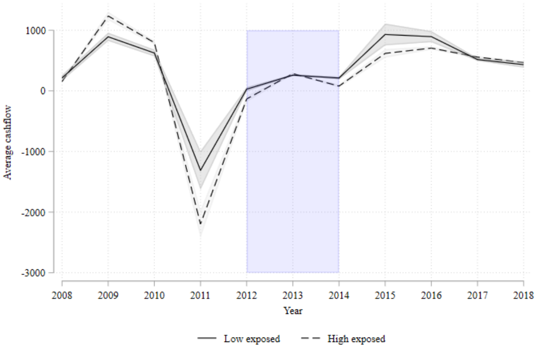


Figure: Cash Flow over time

Main results

Cash Flow

- On average, there is no significant effect of the LPD and the cash flow of firms
- The coefficient is always positive and the magnitude **depends on the degree of the exposure**
- The result is also **robust to several specifications** of the estimated model
- The effect is still **positive but higher** in magnitude, albeit **not distinguishable from zero for SMEs** ($\leq$ 250)
- However, commercial transaction terms might take some time to adapt to the new provisions and have an impact on firms' cash flow

Event Study on Cash Flow

- The existence of a common trend is a key identifying assumption for DiD
- i.e. More and less exposed firms would have experienced the same trend in Cash Flow in the absence of the policy
- Point estimates suggest that there is no evidence against the presence of a common trend between treated and control units

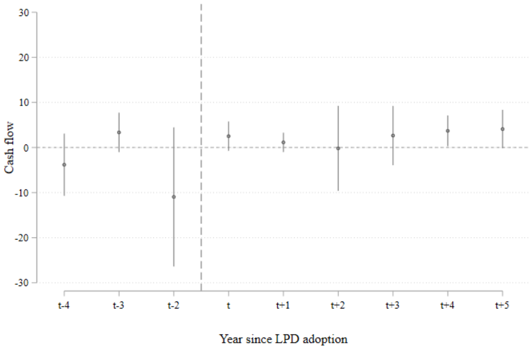


Figure: Whole sample

Event Study on Cash Flow

- The outset of the LPD has increased the cash flow of firms more exposed to the Directive as compared to firms less exposed to it
- The effect ranges from **3.7** (four years after the adoption) to **4**, statistically significant at 5 and 10% respectively

Let's consider the impact in (t+4):

- Average cash flow in highly exposed companies is 60% higher than the pre-LPD period (comparing high and less exposed companies)
- The effect of the LPD on cash flow is stronger the larger was firm's collection period in the past

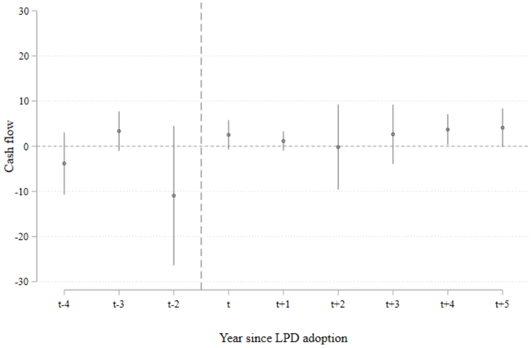


Figure: Whole sample

Heterogeneity

Heterogeneity by country

- Country analysis to isolate specific dynamics
- No evidence of anticipation (except in Spain)
- Previous findings largely confirmed for BE, ES and IT

Heterogeneity by sector

- To highlight sectoral differences
- We group NACE activities in Manufacturing, Manufacturing and Construction and Services
- Previous findings confirmed in manufacturing and manufacturing and construction

Robustness

- Other outcome variables (Investment, Sales, Tangible Assets)
- Results not driven by a country or a sector
- Exposure indicator computed on a different time span
- Different thresholds of credit collection terms

Conclusions

- Late payments are a serious threat for businesses
- The LPD has undoubtedly had a significant role in reducing collection period days **(- 22 days on average)**
- Four years after the adoption of the LPD the average cash flow, in highly exposed companies, is 60% higher than its value the year before the introduction of the directive when comparing firms more exposed to less exposed ones
- More marked effect for firms operating in the manufacturing and construction sectors, characterized by a strong presence of SMEs in the relevant supply chains

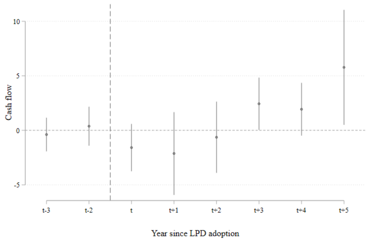
Lessons learned

- By fostering more discipline in firms' payment terms, the LPD had beneficial repercussions on the economy
- The effect does not materialise immediately
- Predictable payment terms within a "standard" time range (30–60 days) increase businesses' cash flow and sales
- Effective mechanism to be preserved and enforced
- Weaknesses: selected sample of countries, Orbis limitations, heterogeneous coverage, imprecise measurement of some variables

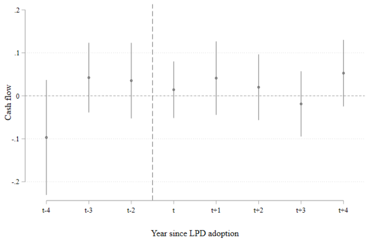
Thank you

European Commission

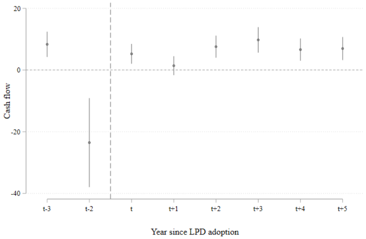
Event Study by country



BE



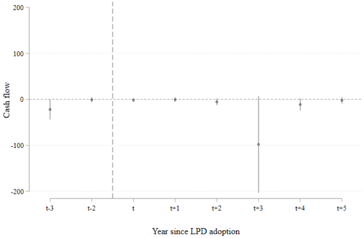
DE



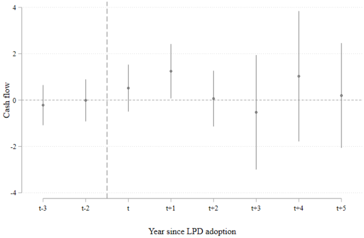
ES

European Commission

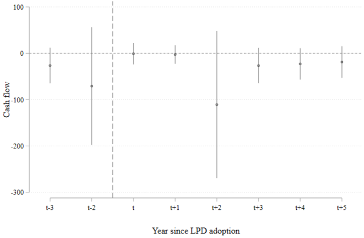
Event Study by country



FI



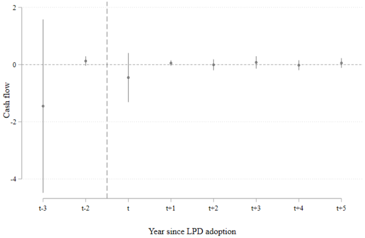
FR



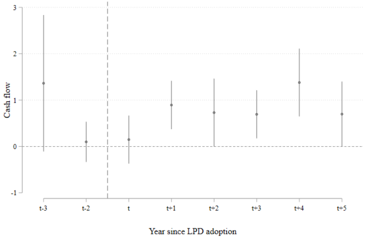
UK

Back

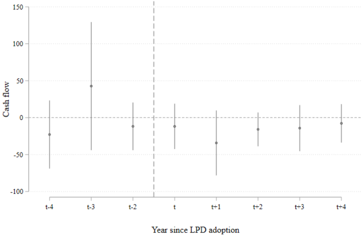
Event Study by country



HU



IT



NL

Back

European Commission

Event Study by sector

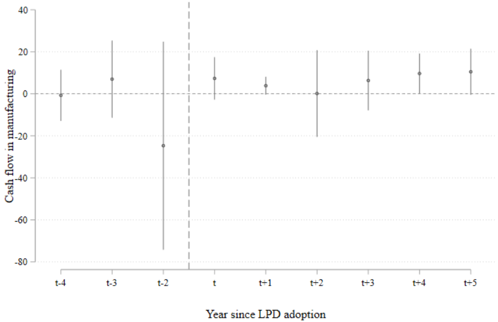


Figure: The dynamics of the effect in Manufacturing

Back

Event Study by sector

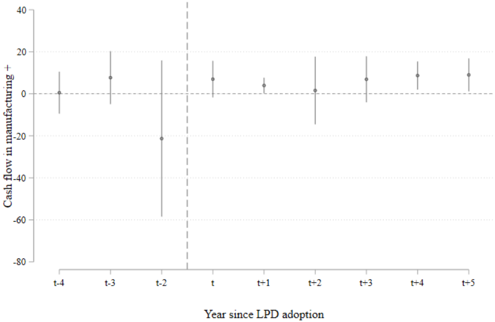


Figure: The dynamics of the effect in Manufacturing and Construction

Back

Event Study by sector

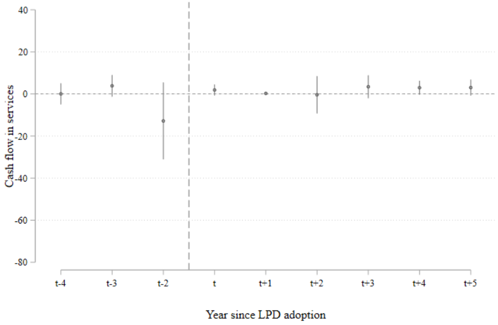
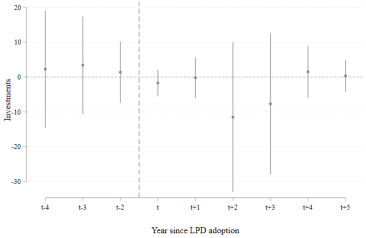
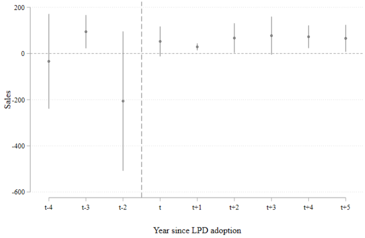


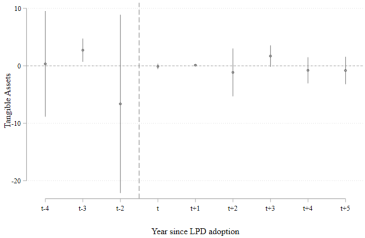
Figure: The dynamics of the effect in Services

Back

Other outcomes







Back

Sensitvity checks

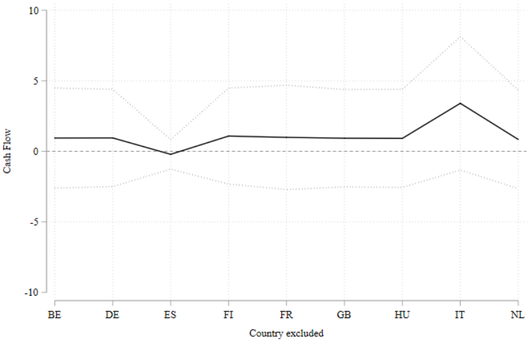


Figure: Exclusion of a country

Back

Sensitvity checks

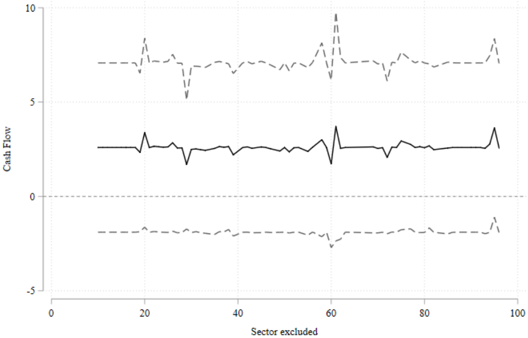


Figure: Exclusion of a sector

Back