Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Investment subsidies and SME performance under staggered treatment: Evidence from matched difference-in-differences

**Eugeni Gil-Ocana**

egiloca@doctor.upv.es

Universitat Politècnica de València

**Ana Garcia-Bernabeu**

Universitat Politècnica de València

**Pablo De Pedraza**

Universitat Politècnica de València

**Research Article**

**Keywords:** Public subsidies, SME policy evaluation, Causal inference, Matching methods, Difference-in-differences, Staggered treatment effects

**Posted Date:** May 12th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-9498631/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Investment subsidies and SME performance under staggered treatment: Evidence from matched difference-in-differences

**Abstract**

Monitoring of support schemes for small and medium-sized enterprises is essential to assess policy effectiveness and the efficient use of public resources. This paper evaluates the causal impact of investment subsidies on SME performance under non-random treatment assignment and staggered policy implementation. The empirical strategy combines propensity score matching, to construct a comparable control group, with a doubly robust estimator, which identifies group-time average treatment effects while accounting for treatment heterogeneity across cohorts and periods. The framework is implemented using administrative subsidy records from the Spanish National Subsidies Database (BDNS) and firm-level accounting data from SABI for the period 2018–2024. Results indicate that subsidies generate a positive and significant effect on fixed asset accumulation, with no significant short-run effects on revenue, employment, or financial ratios. Heterogeneity analysis shows that effects are concentrated among micro-enterprises and increase monotonically with subsidy intensity

**Plain English Summary.** Public investment subsidies boost capital investment in small firms, but do not immediately improve sales or jobs for most recipients.We study the effect of Spanish government grants on small and medium-sized enterprises using a rigorous statistical method that accounts for the fact that firms receive subsidies at different points in time. We find that subsidies lead to a significant increase in fixed assets but have no significant short-run effect on revenue, employment, or financial health for the average recipient. However,the smallest firms, those with fewer than ten employees, do show broader benefits, including employment growth, and larger grants generate stronger effects. These findings suggest that policymakers designing SME support programmes should pay close attention to firm size and grant magnitude when allocating public funds, as blanket interventions may be ineffective for larger firms while proving transformative for the smallest ones.

**Keywords:** Public subsidies, SME policy evaluation, Causal inference, Matching methods, Difference-in-differences, Staggered treatment effects

**JEL Classification:** L25 , C21 , C23 , O38

# 1 Introduction

The weight of small and medium enterprises (SMEs) in employment and economic activity makes them of particular interest when designing public support schemes. Effective policies require careful evaluation, especially when they are aimed at alleviating financial constraints or fostering development (European Commission, 2021; OECD, 2020). The subsidy literature remains contested: one strand shows that subsidies promote investment and firm growth (Bronzini & Iachini, 2014; Criscuolo, Martin, Overman, & Van Reenen, 2019), while another casts doubt on their effectiveness, including the risk of distorting competition or crowding out private investment (see, for example, the General Block Exemption Regulation, GBER — Commission Regulation (EU) No 651/2014 of 17 June 2014). These conflicting views, however, converge on a single point: rigorous evaluation is essential (Srhoj, Lapinski, & Walde, 2021). Modern evaluation relies on rich microdata and empirical strategies capable of constructing credible counterfactuals from observational settings, while incorporating recent methodological advances in causal inference that account for treatment heterogeneity and staggered policy implementation (Callaway & Sant'Anna, 2021).

This paper is grounded in these principles to evaluate the causal effects of public investment subsidies on firm performance. We devote particular attention to two sources of heterogeneity: firm size and subsidy intensity. To address these challenges, we combine propensity score matching with the doubly robust estimator of Callaway and Sant'Anna (2021), which identifies group-time average treatment effects, $ATT(g,t)$, while allowing for heterogeneous effects across cohorts and periods. We focus on the Spanish post-COVID landscape, exploiting subsidies classified as state aid but exempted from notification to the European Commission under the GBER, designed to stimulate productive investment among firms.

The empirical analysis merges two administrative sources at the firm level. Treated firms enter the sample through the Base de Datos Nacional de Subvenciones (BDNS), the Spanish national registry of public transfers, which records every grant awarded under the GBER between 2022 and 2024 with information on the legal basis, the gross grant equivalent, and the award date. Financial data for both treated and control firms come from SABI, a Bureau Van Dijk database covering balance sheets, income statements, and firm demographics for the full population of Spanish limited liability and joint-stock companies. Linking the two sources yields a panel from 2018 to 2024, long enough to provide at least four pre-treatment observations per firm and a covariate set that captures size, age, sector, financial structure, and past growth trajectories.

Our results show that public subsidies primarily affect capital accumulation, with no significant impact on revenue, employment, or financial ratios on average. Heterogeneity analysis reveals that these effects are driven by micro-enterprises, for which the response in fixed assets is substantially larger and accompanied by an increase in employment. Among larger firms, estimated effects are generally indistinguishable from zero. The dose-response analysis, focusing on micro-enterprises, shows that the effect on fixed assets increases monotonically with subsidy intensity, while revenue and employment respond only at the highest levels of support. These findings suggest that

investment subsidies relax binding capital constraints primarily for the smallest firms, and that grant size is a key determinant of effectiveness.

This paper makes three main contributions. First, it provides new causal evidence on the effects of investment subsidies in a staggered treatment setting, leveraging recent advances in the difference-in-differences literature (Callaway & Sant'Anna, 2021; Sun & Abraham, 2021). Second, it moves beyond a binary treatment framework by estimating dose-response effects based on subsidy intensity. Third, it documents substantial heterogeneity across firm size, showing that the effects of subsidies are concentrated among micro-enterprises.

The rest of the paper is structured as follows. Section 2 reviews the relevant literature. Section 3 presents the empirical strategy, including the matching procedure, identifying assumptions, and estimation framework. Section 4 describes the institutional setting and data. Section 5 reports the main results and heterogeneity analysis. Section 6 concludes.

## 2 Causal inference in SME policy evaluation: methods and evidence

The evaluation of public support policies for SMEs has developed along two interconnected axes: the methodological frontier of causal inference, and the growing body of applied evidence on the effects of public support on firm outcomes. On the methodological side, propensity score matching (Rosenbaum & Rubin, 1983) became the workhorse approach for constructing credible counterfactuals from observational data, reducing a high-dimensional covariate space to a single scalar that summarises the probability of treatment assignment. The practical appeal of this reduction comes with well-known costs. King and Nielsen (2019) show that matching on the propensity score can actually worsen covariate balance relative to direct matching methods when the score model is misspecified. Inverse probability weighting (Heckman, Ichimura, & Todd, 1997) and entropy balancing (Hainmueller, 2012) achieve balance by construction rather than by iterative search, though each carries its own sensitivity to model specification.

Matching alone, however, only addresses selection on observable characteristics. To also account for unobserved time-invariant heterogeneity, the literature has combined matching with difference-in-differences estimators, exploiting the panel structure of the data to isolate causal effects (Angrist & Pischke, 2009; Imbens & Lemieux, 2008). The validity of this two-step design rests on the parallel trends assumption, which requires that, absent treatment, treated and control firms would have followed comparable outcome trajectories. Marcus and Sant'Anna (2021) formalise this assumption and show that there is a fundamental robustness-efficiency trade-off in its specification: more restrictive formulations gain efficiency but impose identifying restrictions that are difficult to defend when treatment cohorts differ systematically.

Standard two-way fixed effects estimators, the standard estimator in applied DiD, produce biased estimates when treatment effects are heterogeneous across cohorts or when firms receive support at different points in time (Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021; Sun & Abraham, 2021). In staggered adoption settings, these estimators implicitly use already-treated firms as controls for later-treated ones, contaminating the counterfactual in ways that can generate negative weights and misleading estimates. In response, Callaway and Sant'Anna (2021) propose a group-time estimator that explicitly accounts for treatment timing heterogeneity and restricts comparisons to never-treated firms, providing a more credible identification strategy in exactly these settings. Beyond this framework, machine learning methods such as causal forests (Wager & Athey, 2018) and double machine learning (Chernozhukov et al., 2018) open new possibilities for estimating heterogeneous treatment effects in high-dimensional settings, representing a promising direction for future research in SME policy evaluation.

On the applied side, the evidence is generally positive but far from uniform. Small capital grants to microenterprises in Sri Lanka generate measurable returns on profits and capital stock even when award amounts are modest (de Mel, McKenzie, & Woodruff, 2008). Quasi-experimental work has extended these findings to more complex institutional settings: regional subsidies raise employment in the United Kingdom (Criscuolo et al., 2019), R&D grants in Italy crowd in rather than displace private investment (Bronzini & Iachini, 2014), innovation support in Belgium lifts R&D output (Czarnitzki & Lopes-Bento, 2013), and US innovation grants leave detectable traces in patenting activity and subsequent venture funding (Howell, 2017). These headline results, however, sit alongside less comfortable findings. Italian investment subsidies raise employment and capital accumulation yet appear to drag on total factor productivity, pointing to possible misallocation (Cerqua & Pellegrini, 2017). Czech cohesion grants improve labour productivity without moving the needle on financial performance (Dvouletý & Blažková, 2019). Meta-analytic evidence reinforces the impression of a literature where positive average effects coexist with substantial heterogeneity across programmes, firm types, and methodological choices (Card, Kluve, & Weber, 2018; Dimos & Pugh, 2016; Dvouletý, Srhoj, & Pantea, 2021).

The Spanish context has generated its own strand of evidence, concentrated on R&D support rather than investment subsidies. Blanes and Busom (2004) find that selection into public R&D programmes in Spain is heavily shaped by agency-specific rules and differs systematically between high-tech and low-tech industries. Huergo and Moreno (2017) compare subsidies and soft credits directed at Spanish firms and find that both instruments improve innovation outputs, though through distinct mechanisms: concessional loans impose repayment discipline that direct grants do not, altering how recipients deploy the funds. Surveying five decades of international evidence, Zúñiga-Vicente, Alonso-Borrego, Forcadell, and Galán (2014) argue that the heterogeneity of estimated effects across studies reflects genuine variation in programme contexts rather than methodological noise alone. Taken together, these contributions point to a Spanish support landscape marked by instrument diversity, firm-level selection, and context-dependent returns precisely the features our empirical design is built to accommodate.

COVID-19 complicates the picture for any study using post-2020 data. Canzian et al. (2025) examine state aid allocated to Spanish and Italian firms during the pandemic using difference-in-differences combined with propensity score weighting. Micro-firms in their sample fared notably better when supported, turnover declines were smaller and tangible investment held up while larger firms showed little response, a pattern that echoes the size heterogeneity we document for structural GBER subsidies. Since their analysis runs through 2022 and our pre-treatment window overlaps with that period, some firms in our sample likely received pandemic-related transfers before the investment grants we study. Our restriction to recipients of a single instrument with no concurrent public support is therefore doubly motivated: it guards against treatment contamination and ensures that the pre-treatment trajectories we use for identification are not themselves shaped by an extraordinary policy episode.

This body of evidence points to three lessons that shape our empirical strategy. First, selection bias is pervasive and must be addressed through careful matching on pre-treatment firm characteristics. Second, treatment effects are heterogeneous by firm size, grant intensity, and timing, which calls for a method that accommodates this heterogeneity rather than averaging it away. Third, the staggered nature of subsidy allocation in administrative data makes standard two-way fixed effects unreliable. Our empirical framework is designed to address all three challenges simultaneously, and in doing so contributes with new causal evidence on the effectiveness of investment subsidies for Spanish SMEs

# 3 Methodology

The empirical strategy consists of four stages (see Figure 1). First, we build a matched control sample to ensure comparability of pre-treatment characteristics between treated and untreated firms. Second, we estimate treatment effects using the doubly robust estimator of Callaway and Sant'Anna (2021), which accommodates staggered treatment adoption and heterogeneous effects across cohorts. Third, we partition the sample by firm size and subsidy amount following Srhoj et al. (2021) to examine treatment effect heterogeneity. Fourth, we assess the sensitivity of the results to alternative matching procedures and identification assumptions through a series of robustness checks.

## 3.1 Matching procedure

We start by constructing a control group using one-to-one propensity score matching (PSM) without replacement. The goal is to improve comparability between treated and untreated firms along observable pre-treatment characteristics, reducing selection bias before estimation (Rosenbaum & Rubin, 1983). The procedure rests on the Conditional Independence Assumption (CIA): conditional on observed firm characteristics, treatment assignment is independent of potential outcomes. What makes this assumption credible in our setting is the richness of the data. Although grants are restricted

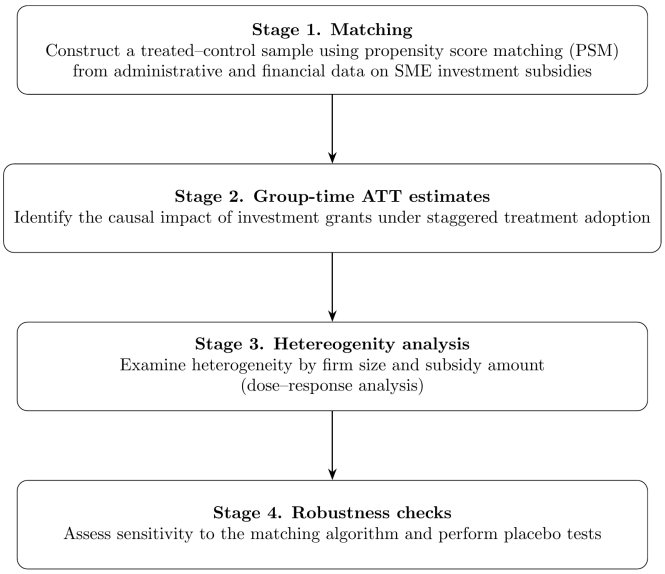



**Fig. 1** Flowchart of the four-stage empirical strategy

to firms in assisted areas or meeting specific eligibility criteria, selection into treatment within this eligible population is primarily driven by observable firm-level characteristics, size, productivity, financial structure, and growth trajectories, all of which are captured by our covariate set.

We prefer PSM over Mahalanobis distance matching for a practical reason: our covariate matrix includes four lags of each outcome variable alongside structural firm characteristics, which introduces substantial collinearity. Inverting a near-singular covariance matrix produces match distances that no longer reflect genuine multivariate proximity, and the problem worsens as the ratio of covariates to observations grows (Hampel, Ronchetti, Rousseeuw, & Stahel, 1986). Propensity score matching handles this differently rather than working in the original covariate space, it reduces the matching problem to a single scalar, the estimated probability of treatment, and so never requires inverting the covariance matrix directly (Rosenbaum & Rubin, 1983). Critics of PSM rightly note that multicollinearity among covariates inflates variance in the score model itself, but this matters less here than it would in a structural regression: we are using the score to balance the sample, not to recover interpretable coefficients on individual predictors (Imbens & Wooldridge, 2009). For that balancing purpose, a score estimated from a collinear covariate set is adequate provided the model is not badly misspecified.

The matching procedure is stratified by industry, imposing an exact match on the two-digit NACE classification. This restriction ensures that treated and control firms operate in comparable competitive environments and face similar demand and production conditions. Within each industry stratum, we match on a broad set of firm-level

pre-treatment characteristics. Firm size and scale are captured through log operating revenue and log number of employees, as prior sales and employment levels are systematically correlated with future performance (Peric & Vitezić, 2016). Financial structure is controlled for through the leverage ratio at $t-1$ and log liquid liabilities. Productive capacity is proxied by log fixed assets, which reflects the pre-treatment capital stock and the firm's capacity to absorb and deploy the investment grant. Overall financial health is captured through the solvency ratio. Trade status is captured through categorical indicators distinguishing exporters, importers, two-way traders, and domestic producers, as it affects both access to public support and performance trajectories (Munch & Schaur, 2018). Finally, firm age is included to account for lifecycle effects on investment behavior and financial constraints.

We include all remaining outcome variables across four pre-treatment periods ($t-1$ to $t-4$). Conditioning on multiple lags of the outcome variables, rather than on a single pre-treatment period, ensures that treated and control firms share comparable dynamic paths prior to treatment (Chabé-Ferret, 2017). This directly targets the Parallel Trends Assumption (PTA), though the validity of our design rests on the assumption itself rather than on this design choice alone.

Following Marcus and Sant'Anna (2021), we adopt a formulation of the PTA suited to our data and institutional context. Formally:

$$\mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid G_i = g\right] = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid C_i = 1\right] \tag{1}$$

where $i$ indexes firms, $G_i = g$ denotes the period in which it first adopts the treatment, and $C_i = 1$ identifies firms that are never treated. This assumption requires that, in the absence of treatment, the evolution of untreated potential outcomes would have followed parallel trends across treated cohorts and never-treated firms over the observation window.

The choice of this formulation over more restrictive alternatives reflects the robustness-efficiency trade-off discussed in Marcus and Sant'Anna (2021). Stronger versions, such as those requiring parallel trends conditional on covariates across all cohort pairs, gain efficiency at the cost of identifying restrictions that are difficult to defend in settings with staggered adoption and macroeconomic shocks. Given the availability of a sufficiently large never-treated group and the fact that treatment cohorts correspond to different grant-award years, we favour the more robust formulation.

We assess both the quality of the match and the plausibility of the PTA through a set of complementary diagnostics. Covariate balance after matching is assessed using the absolute standardized mean difference (ASMD), defined as:

$$\text{ASMD} = \left| \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{S_T^2 + S_C^2}{2}}} \right| \tag{2}$$

where $\bar{X}_T$ and $\bar{X}_C$ denote the sample means for treated and control firms, and $S_T^2$ and $S_C^2$ their respective variances. Following Austin (2009), an ASMD below 0.1 is taken as evidence of adequate covariate balance. The plausibility of the PTA is assessed through a dynamic event study. Pre-treatment coefficients do not constitute formal overidentifying restrictions but provide diagnostic evidence on the assumption. As additional diagnostics, we verify sufficient overlap in the distribution of estimated propensity scores between treated and control firms, examine the distributional similarity of key covariates through quantile-quantile plots, and plot the pre-treatment evolution of each outcome variable for treated and control firms separately.

### 3.2 Estimating treatment effects: group-time ATT

To identify the causal impact of investment grants under staggered treatment adoption, we follow Callaway and Sant'Anna (2021). We avoid Two-Way Fixed Effects (TWFE) estimators which produce biased estimates when treatment effects are heterogeneous across cohorts or dynamic over time (Borusyak, Jaravel, & Spiess, 2024; Goodman-Bacon, 2021; Sun & Abraham, 2021). The estimand of interest is the group-time average treatment effect $ATT(g,t)$, defined as the average effect at time $t$ for units first treated at time $g$. Under assumption (1), $ATT(g,t)$ is nonparametrically identified by comparing the outcome evolution of cohort $g$ to that of never-treated firms relative to the pre-treatment base period $g-1$:

$$\widehat{ATT}^{OR}(g,t) = \frac{\sum_{i=1}^{n} G_{ig}(Y_{it} - Y_{i,g-1})}{\sum_{i=1}^{n} G_{ig}} - \frac{\sum_{i=1}^{n} C_i(Y_{it} - Y_{i,g-1})}{\sum_{i=1}^{n} C_i} \tag{3}$$

where $G_{ig}$ is an indicator for units first treated in period $g$ and $C_i$ identifies never-treated units. To guard against misspecification of either the outcome or the selection model, we implement the doubly robust (DR) extension of Callaway and Sant'Anna (2021), which augments the outcome regression (OR) estimator with an inverse probability weighting component:

$$\widehat{ATT}^{DR}(g,t) = \frac{\sum_{i=1}^{n} G_{ig}\left[\widehat{Y}_{it}^{c} + \dfrac{D_{it}\left(Y_{it} - \widehat{Y}_{it}^{c}\right)}{\widehat{p}_{it}}\right]}{\sum_{i=1}^{n} G_{ig}} - \frac{\sum_{i=1}^{n} C_i\, \widehat{Y}_{it}^{c}}{\sum_{i=1}^{n} C_i} \tag{4}$$

where $\widehat{Y}_{it}^{c}$ is the predicted untreated potential outcome and $\widehat{p}_{it}$ is the estimated propensity score. The DR estimator remains consistent if either the outcome model or the propensity score model is correctly specified, but not necessarily both (Sant'Anna & Zhao, 2020). Aggregate treatment effects are obtained by averaging $ATT(g,t)$ across cohorts and post-treatment periods, weighting by cohort size.

Even when applying a doubly robust estimator, consistency and semiparametric efficiency are asymptotic guarantees that may not hold in finite samples. When the distribution of treated and control units exhibits limited overlap, the inverse probability weighting component produces extreme weights that inflate both bias and

variance (Ma, Sant'Anna, Sasaki, & Ura, 2023). By restricting the estimation sample to matched pairs with good balance the matching step explicitly enforces a common support condition prior to estimation, ensuring that the DR estimator operates in a region of adequate overlap.

### 3.3 Heterogeneity analysis

The baseline ATT aggregates treatment effects across all treated firms and post-treatment periods. Because subsidy effects are likely heterogeneous across firm types and grant sizes, this aggregate estimate may obscure economically relevant variation, across firm types or grant sizes. We examine heterogeneity along two dimensions firm size and subsidy amount following Srhoj et al. (2021).

First, we partition the sample by firm size, distinguishing micro-enterprises (fewer than 10 employees) from larger firms, in line with the European Commission's definition. This distinction reflects the theoretical literature on financial constraints: micro-enterprises are more likely to face binding credit constraints, higher costs of external finance, and limited access to capital markets (Beck & Demirgüç-Kunt, 2006; Hall & Lerner, 2010), implying that a public grant may relax a binding constraint for smaller firms while representing a marginal addition to the financing mix of larger ones. If this mechanism operates, we would expect the treatment effect to be larger in magnitude for micro-enterprises.

Second, we estimate a dose-response function by examining heterogeneity in subsidy amount, grouping firms into terciles of the grant distribution and defining low, medium, and high intensity categories at the 33rd and 66th percentiles (Srhoj et al., 2021). This approach allows us to assess whether treatment effects scale with grant size, and in particular whether a minimum grant threshold is required for effects on capital accumulation to materialise.Since micro-enterprises face more severe financing frictions and have limited access to external capital markets, the marginal value of an additional unit of subsidy is predicted to be higher for them than for larger firms (Srhoj et al., 2021). A monotonically increasing pattern of effects across intensity terciles would be consistent with the presence of indivisibilities in investment or fixed costs of project implementation that smaller grants cannot cover.

All subgroup estimates are obtained within the doubly robust Callaway and Sant'Anna (2021) framework, treating each subgroup as an independent estimation sample with its own never-treated comparison group. Standard errors are clustered at the firm level throughout.

### 3.4 Robustness checks

The baseline results are subjected to three robustness checks. First, we examine sensitivity to the matching algorithm. The baseline specification uses one-to-one nearest-neighbor propensity score matching; as an alternative, we consider two specifications with two and three nearest neighbors per treated firm, respectively. Expanding

the number of matches reduces variance but may introduce less comparable control units. Consistency of results across both configurations indicates that the findings are not driven by the choice of matching ratio.

Second, we re-estimate the models replacing propensity score matching with Mahalanobis distance matching, which matches firms directly in the multivariate space of observed characteristics and accounts for the covariance structure of the covariates. Stable estimates across matching methods indicate that the results are not sensitive to the specific balance criterion employed.

Third, we conduct a placebo test by assigning fictitious treatment dates within the pre-treatment period and re-estimating the baseline model. Under valid identification and the absence of anticipatory effects, the placebo coefficients should be statistically indistinguishable from zero.

All analyses are implemented in Python 3.12 using `scipy`, `scikit-learn`, and `statsmodels`. The Callaway-Sant'Anna (2021) doubly robust estimator is implemented via the `diff_diff` package (version 3.2.0; available on PyPI as `diff-diff`), using the `CallawaySantAnna` class with `estimation_method='dr'` and `anticipation=0`. Replication code is available upon request.

# 4 Empirical setting

Public investment subsidies to SMEs represent one of the most widely deployed instruments of industrial policy in the European Union, yet evidence on their effectiveness remains mixed, in part because credible identification is hampered by the non-random allocation of aid. This paper exploits a setting in which the universe of subsidies is centrally registered, the legal basis of each award is documented, and detailed firm-level financial accounts are available for the population of potential recipients, allowing us to construct a clean quasi-experimental design.

## 4.1 Institutional setting

The regulatory framework for the grants examined in this study is governed by the European Union's competition policy, specifically the General Block Exemption Regulation (GBER), Commission Regulation (EU) No 651/2014, which exempts certain categories of state aid from prior notification to the European Commission. We focus on interventions granted under two articles of the GBER: Article 14 (Regional Investment Aid), which targets firms in designated assisted areas with the aim of stimulating investment and job creation, and Article 17 (Investment Aid for SMEs), which provides a broader horizontal framework acknowledging the systemic disadvantages smaller enterprises face in accessing capital markets. Both articles require that the aided investment would not have been undertaken, or would have been smaller in scope, absent the public transfer. In the Spanish administrative context, these subsidies are

deployed across three tiers of government: national, regional (Autonomous Communities), and local, resulting in a heterogeneous landscape of calls targeting various sectors and investment types.

## 4.2 Data

The analysis draws on two data sources. The first is the Base de Datos Nacional de Subvenciones (BDNS), the official administrative registry of all public transfers in Spain, which provides information on the universe of grants allocated under Articles 14 and 17 of the GBER between 2022 and 2024, including the specific legal basis, the gross grant equivalent of the aid, and the date of the award. The second is the Sistema de Análisis de Balances Ibéricos (SABI), provided by Bureau Van Dijk, which contains official annual accounts, profit-and-loss statements, and demographic variables for the population of registered limited liability and joint-stock companies in Spain.

The panel covers the period 2018–2024, providing at least four pre-treatment years and up to two post-treatment years per firm, which is sufficient to verify pre-treatment parallel trends and to allow investment-induced effects to materialise. The dataset is constructed applying the following sample restrictions.

We focus on treated firms that received a single relevant investment aid instrument during the 2022–2023 window and no additional public support of any kind during the observation period. It is a similar restriction as those imposed in other impact evaluation literature (Czarnitzki & Delanote, 2015; Srhoj et al., 2021) and ensures a clean identification of the treatment effect: any post-treatment divergence between treated and control firms can be attributed exclusively to the instrument under evaluation, without contamination from concurrent policy interventions. Given that Spanish SMEs frequently access multiple streams of public funding, averaging approximately 3 subsidies per firm in the raw BDNS data, this restriction reduces the treated sample but substantially strengthens the internal validity of the design. Firms with missing financial records or inconsistent longitudinal data are also discarded. The remaining sample consists of 274 treated firms drawn from a pool of 6884 potential units.

The treatment variable encompasses the full range of aid instruments notified under Articles 14 and 17 of the GBER, including direct grants, subsidised loans, and guarantees. The amount used throughout the analysis is the gross grant equivalent (GGE), which standardises the value of each instrument in terms of its effective public transfer. For direct grants, the GGE coincides with the nominal award; for subsidised loans, it corresponds to the interest rate subsidy, that is, the difference between the market rate and the concessional rate applied rather than the full loan principal. This standardisation ensures comparability of aid intensity across instrument types. As displayed in Table 1, the resulting distribution of aid amounts is highly dispersed: the mean GGE among treated firms is €140 426, with a standard deviation of €634 382 and a median of €31 179. This dispersion reflects the heterogeneous nature of the instruments and investment projects covered by Articles 14 and 17, which range from small subsidies to big investment projects.

**Table 1** Descriptive statistics by treatment status (pre-treatment period)

| | Treated ($N = 274$) | | | Control ($N = 17\,649$) | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | Mean | SD | Median | Mean | SD | Median | $\Delta$ Mean | $p$-value |
| *Performance outcomes* | | | | | | | | |
| Log operating revenue | 14.56 | 1.75 | 14.27 | 13.69 | 1.17 | 13.56 | 0.87 | 0.00 |
| Log employment | 2.73 | 1.29 | 2.56 | 1.98 | 0.87 | 1.79 | 0.75 | 0.00 |
| Log value added | 13.21 | 3.14 | 13.33 | 12.48 | 2.45 | 12.59 | 0.73 | 0.00 |
| Log fixed assets | 13.40 | 2.08 | 13.23 | 11.82 | 2.11 | 11.90 | 1.59 | 0.00 |
| *Financial structure* | | | | | | | | |
| Log liquid liabilities | 13.18 | 1.85 | 12.97 | 12.14 | 1.39 | 12.09 | 1.04 | 0.00 |
| Leverage ratio | 0.49 | 0.26 | 0.48 | 0.47 | 0.55 | 0.40 | 0.01 | 0.44 |
| Solvency ratio | 0.03 | 0.04 | 0.02 | 0.06 | 0.81 | 0.02 | −0.03 | 0.00 |
| *Firm characteristics* | | | | | | | | |
| Firm age (years) | 23.24 | 13.26 | 23.00 | 24.43 | 10.94 | 24.00 | −1.19 | 0.14 |
| Grant amount (€) | 140 426 | 634 382 | 31 179 | — | — | — | — | — |

*Notes:* Statistics are computed over the pre-treatment period using firm-level observations at $t - 1$. $\Delta$ Mean reports the difference in means between treated and control firms. $p$-values correspond to two-sided $t$-tests for equality of means, with standard errors clustered at the firm level. The leverage and solvency ratios exhibit large standard deviations in the control group due to a small number of firms with near-zero equity; results are robust to winsorizing these variables at the 1st and 99th percentiles. Grant amount is reported for treated firms only.

Table 1 also presents descriptive statistics for the key firm-level variables at $t-1$, separately for treated and control firms. Treated firms exhibit systematically higher means in operating revenue, employment, value added, fixed assets, and liquid liabilities, with all differences statistically significant at conventional levels. These pre-treatment differences motivate the matching procedure described in Section 3.1.

# 5 Results

The empirical findings are organized in four parts. We first assess matching quality and pre-treatment diagnostics, then report the core group-time ATT estimates, and conclude with the heterogeneity analysis by firm size and subsidy magnitude, followed by robustness checks.

## 5.1 Matching quality and pre-treatment diagnostics

After implementing the matching procedure, all 274 treated firms are retained. Figure 2 displays the absolute standardized mean differences before and after matching for all covariates. Prior to matching, imbalances are substantial, with several variables including log fixed assets, log employees, log operating revenue, and log liabilities exhibiting standardized differences in excess of 0.6. After matching, all financial and demographic variables fall below the conventional threshold of 0.1 with the exception

of firm age (ASMD = 0.14) and the dummy for two-way trading (ASMD = 0.22). Two-way trading status is correlated with firm productivity and trade exposure; however, we note that treated and control firms are already matched on exact two-digit NACE sector and on four lags of all seven outcome variables (28 lagged covariates in total). Conditional on this rich pre-treatment history, residual imbalance in trade status is unlikely to generate differential outcome trends, and the doubly robust estimator provides an additional layer of protection against residual confounding via propensity score reweighting. Solvency ratio ASMD remains below the 0.1 threshold after matching but is marginally higher at $t-1$ than before matching (ASMD 0.07 vs. 0.02); this occurs because solvency was already very well balanced in the full sample, and the restriction to matched pairs shifts the joint distribution slightly, though the absolute level remains unproblematic.

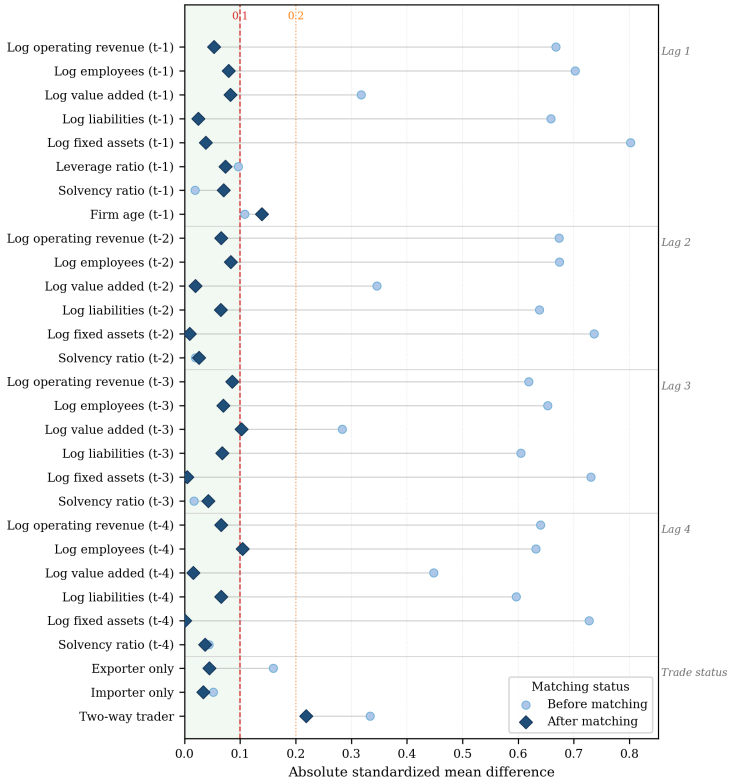



**Fig. 2** Absolute standardized mean differences (ASMD) before and after matching for all covariates. The first dashed vertical line indicates the conventional balance threshold of 0.1. Values to the left of the threshold indicate adequate covariate balance.

Figure 3 reports the event-study coefficients for periods $t-4$ through $t+2$. For most outcomes, pre-treatment coefficients are small and statistically indistinguishable from

zero throughout all four pre-treatment periods. Log operating revenue, log employees, leverage ratio, and solvency ratio display no discernible pre-treatment trend. Log value added shows negative coefficients at $t-4$ ($-0.380$) and $t-3$ ($-0.246$), though neither is statistically significant. Log liquid liabilities and log fixed assets exhibit negative and significant coefficients at $t-4$ ($-0.084$ and $-0.106$, respectively), which revert toward zero and become statistically indistinguishable from zero by $t-2$ and $t-1$. No outcome displays a significant coefficient at $t-1$. We note that for cohorts first treated in 2024, $t-4$ maps to 2020 — the height of the COVID-19 shock — and the data span both the disruption and the subsequent recovery. Treated and control firms in the same NACE sector are exposed to the same aggregate shock, and the exact-sector matching ensures that any differential recovery occurring between $t-4$ and $t-2$ reflects compositional rather than systematic differences. The absence of any significant pre-trend at $t-1$ confirms that parallel trends hold in the period immediately preceding treatment, which is the relevant condition for identification under Callaway and Sant'Anna (2021).

Complementary diagnostics on propensity score overlap and covariate distributions are reported in Appendix A

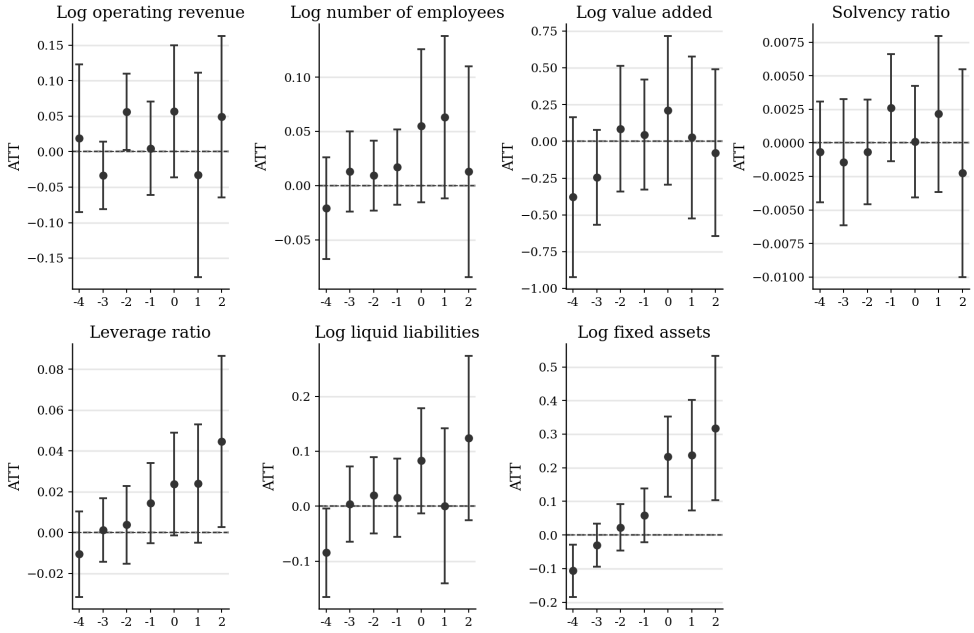



**Fig. 3** Event-study estimates following Callaway and Sant'Anna (2021) for periods $t-4$ through $t+2$, where $t=0$ denotes the year of grant receipt. Each panel reports the group-time ATT for a given outcome variable. X-axis represent periods relative to treatment. Bands represent 95% confidence intervals. Standard errors are clustered at the firm level.

## 5.2 Average group-time (ATT) estimates

Table 2 reports the main results obtained using two alternative estimation strategies: the outcome regression estimator and the doubly robust estimator, both following the framework proposed by Callaway and Sant'Anna (2021). All specifications include firm-level controls, and standard errors are clustered at the firm level to account for serial correlation. Results are consistent across both estimation methods. We find that the effects of the subsidies are highly concentrated on investment-related outcomes. Fixed assets exhibit a significant increase of approximately 19.9% in the regression specification and 19.8% under the doubly robust approach ($p < 0.01$ in both cases). Other dimensions of firm performance do not show statistically significant changes within the observation window. Variables such as the number of employees, operating revenue, and liquid liabilities show positive coefficients, but do not reach conventional levels of significance. The post-treatment window of two years is unlikely to be sufficient for investment-induced gains to translate into measurable improvements (Bernini, Cerqua, & Pellegrini, 2017).

**Table 2** Average Treatment Effect on the Treated (ATT)

| | Outcome regression (OR) | | Doubly Robust (DR) | |
|---|---|---|---|---|
| Variable | Coef. (SE) | $p$-value | Coef. (SE) | $p$-value |
| Log Operating Revenue | 0.017 (0.038) | 0.666 | 0.017 (0.040) | 0.670 |
| Log Employees | 0.036 (0.028) | 0.201 | 0.037 (0.030) | 0.218 |
| Log Value Added | 0.009 (0.187) | 0.962 | 0.010 (0.192) | 0.960 |
| Solvency Ratio (%) | -0.002 (0.003) | 0.481 | -0.002 (0.003) | 0.499 |
| Leverage Ratio (%) | 0.011 (0.010) | 0.238 | 0.011 (0.010) | 0.243 |
| Log Liquid Liabilities | 0.033 (0.046) | 0.475 | 0.033 (0.048) | 0.488 |
| Log Fixed Assets | 0.199 (0.055) | 0.000 | 0.198 (0.056) | 0.000 |
| Controls | Yes | | Yes | |
| Observations | 3500 | | 3500 | |

Notes: Standard errors (SE) in parentheses are clustered at the firm level. Estimations follow Callaway and Sant'Anna (2021). All specifications include firm and time fixed effects and firm-level controls.

## 5.3 Heterogeneity analysis

The aggregate ATT estimates reported in subsection 5.2 mask potential heterogeneity in how firms respond to public subsidies. In what follows, we examine heterogeneity along two dimensions: firm size, distinguishing micro-enterprises from larger firms, and dose-response functions for subsidy intensity, grouping firms into terciles of the grant distribution.

To asses the group-time att estimates attending firm size we split the sample in two groups: micro-firms and larger firms. As shown in Table 3, it reveals systematic differences in how firms respond to public subsidies depending on their initial scale.

For micro-enterprises, subsidy effects are both broad and economically meaningful. Fixed assets increase substantially, with a group-time ATT of 0.337 ($p < 0.01$), suggesting that subsidies translate directly into tangible capital accumulation among the smallest firms. Employment also expands significantly (group-time ATT = 0.069, $p < 0.05$). In addition, we observe a marginally significant increase in the leverage ratio (group-time ATT = 0.029, $p < 0.10$). Rather than crowding out private finance, this pattern may reflect a certification effect whereby receiving a public subsidy improves firms' creditworthiness and facilitates access to external debt.

**Table 3** Heterogeneity analysis by firm size: group-time ATT estimates

| | Micro firms (< 10 employees) | | Larger firms (≥ 10 employees) | |
|---|---|---|---|---|
| Variable | ATT (SE) | $p$-value | ATT (SE) | $p$-value |
| Log Operating Revenue | 0.080 (0.061) | 0.190 | −0.021 (0.055) | 0.704 |
| Log Employees | 0.069 (0.034) | 0.043 | 0.017 (0.044) | 0.696 |
| Log Value Added | 0.050 (0.348) | 0.886 | 0.016 (0.249) | 0.949 |
| Solvency Ratio (%) | −0.004 (0.005) | 0.373 | 0.000 (0.003) | 0.979 |
| Leverage Ratio (%) | 0.029 (0.017) | 0.078 | −0.001 (0.012) | 0.943 |
| Log Liquid Liabilities | 0.112 (0.079) | 0.156 | −0.016 (0.063) | 0.797 |
| Log Fixed Assets | 0.337 (0.112) | 0.003 | 0.092 (0.055) | 0.098 |
| Controls | Yes | | Yes | |
| Observations | 1456 | | 2044 | |

*Notes:* Standard errors (SE) in parentheses are clustered at the firm level. Estimates use the doubly robust group-time estimator.

By contrast, subsidy effects among larger firms are considerably more muted. The coefficient on fixed assets remains positive (group-time ATT = 0.092) but is only marginally significant ($p < 0.10$) and is roughly four times smaller in magnitude than the corresponding estimate for micro-enterprises. All remaining outcomes, employment, revenue, value added, and financial ratios are statistically indistinguishable from zero.

To implement the dose-response analysis, we group micro-enterprises into three categories defined at the 33rd and 66th percentiles of the grant distribution among micro-enterprise recipients ($N = 107$). The low-intensity group comprises firms receiving grants up to 7056€ ($N = 36$, median award 3347€); the medium-intensity group spans 7056€–23 805€ ($N = 35$, median 15 034€); and the high-intensity group includes grants above 23 805€ ($N = 36$, median 67 444€). The grant distribution among micro-enterprises is right-skewed (mean 28 735€, median 16 275€), so the high-intensity bin captures a wide range of award sizes with a long upper tail.

**Table 4** Heterogeneity analysis by dose-response functions for micro enterprises: group time ATT estimates

| | Low dose | | Medium dose | | High dose | |
|---|---|---|---|---|---|---|
| Outcome Variable | ATT (SE) | $p$-value | ATT (SE) | $p$-value | ATT (SE) | $p$-value |
| Log Operating Revenue | 0.028 (0.094) | 0.770 | 0.020 (0.064) | 0.756 | 0.118 (0.060) | 0.050 |
| Log Employees | 0.031 (0.036) | 0.396 | 0.038 (0.039) | 0.319 | 0.090 (0.053) | 0.090 |
| Log Value Added | −0.085 (0.519) | 0.870 | −0.336 (0.404) | 0.406 | 0.136 (0.233) | 0.561 |
| Solvency Ratio (%) | 0.001 (0.005) | 0.924 | −0.002 (0.005) | 0.715 | −0.015 (0.010) | 0.128 |
| Leverage Ratio (%) | 0.056 (0.024) | 0.019 | −0.014 (0.020) | 0.474 | 0.029 (0.028) | 0.290 |
| Log Liquid Liabilities | 0.098 (0.097) | 0.313 | −0.018 (0.104) | 0.860 | 0.258 (0.114) | 0.024 |
| Log Fixed Assets | 0.239 (0.100) | 0.017 | 0.338 (0.136) | 0.013 | 0.619 (0.224) | 0.006 |
| Controls | Yes | | Yes | | Yes | |
| Observations | 1001 | | 931 | | 938 | |

*Notes:* Standard errors (SE) in parentheses are clustered at the firm level. Estimates use the doubly robust group-time estimator.

First, fixed asset accumulation is positive and statistically significant across all three intensity percentiles, increasing monotonically from group-time ATT = 0.239 at the lowest tercile to group-time ATT = 0.338 at the medium tercile and group-time ATT = 0.619 at the highest tercile. This gradient is considerably steeper than in the full-sample estimates, indicating that micro-enterprises are more responsive to subsidy scale than larger firms. Second, the leverage ratio increases among low-intensity recipients (group-time ATT = 0.056, $p < 0.05$) but not at higher intensities, consistent with the view that small grants prompt firms to complement public funds with external debt, while larger grants may be sufficient to finance investment independently. Third, high-intensity recipients exhibit a significant increase in liquid liabilities (group-time ATT = 0.258, $p < 0.05$), which may reflect short-term debt taken on to support larger investment projects. No statistically significant effects are detected for operating revenue, employment, value added, or solvency at any intensity level, with the exception of marginally significant coefficients for revenue and employment among high-intensity recipients ($p = 0.050$ and $p = 0.090$, respectively).

To assess the economic magnitude of these effects, we translate the log-point estimates into approximate euro impacts using the pre-treatment median fixed assets within each intensity bin. The median treated micro-enterprise holds €144 017 in fixed assets in the low-intensity bin, €121 105 in the medium-intensity bin, and €269 566 in the high-intensity bin. Applying the group-time ATT estimates, the implied increase in fixed assets is approximately €38 882 for low-intensity recipients ($\exp(0.239) - 1 \approx 27.0\%$), €48 701 for medium-intensity recipients ($\exp(0.338) - 1 \approx 40.2\%$), and €231 037 for high-intensity recipients ($\exp(0.619) - 1 \approx 85.7\%$). Expressed relative to the median grant received by each group, these figures imply a return on public investment of approximately 13.4€ per euro of grant for low-intensity recipients (median grant €2900), 3.0€ for medium-intensity recipients (median grant €16 275), and 5.0€ for high-intensity recipients (median grant €46 267). The non-monotonic pattern reflects

the interaction between grant size and the pre-treatment capital stock of each group rather than a simple diminishing returns story, and all three groups exhibit investment responses that substantially exceed the value of the public transfer.

## 5.4 Robustness checks

The baseline results are subjected to three robustness checks based on alternative matching specifications. The first two extend the baseline one-to-one nearest-neighbor PSM to two and three neighbors per treated firm (NN = 2 and NN = 3), respectively. The third replaces PSM with Mahalanobis distance matching. As noted in Section 3, the high collinearity of the temporal covariates renders the inverse covariance matrix ill-conditioned under Mahalanobis matching, reducing the number of firms with adequate balance to 129. All treatment effects are estimated within the doubly robust group-time framework.

**Table 5** Robustness Checks: alternative matching specifications

| | NN = 2 | | NN = 3 | | Mahalanobis | |
|---|---|---|---|---|---|---|
| Outcome Variable | ATT (SE) | $p$-value | ATT (SE) | $p$-value | ATT (SE) | $p$-value |
| Log Operating Revenue | −0.018 (0.040) | 0.653 | −0.011 (0.038) | 0.769 | 0.013 (0.048) | 0.787 |
| Log Employees | 0.031 (0.026) | 0.227 | 0.033 (0.025) | 0.191 | 0.046 (0.020) | 0.021 |
| Log Value Added | −0.071 (0.172) | 0.679 | −0.064 (0.162) | 0.692 | −0.088 (0.082) | 0.286 |
| Solvency Ratio (%) | −0.003 (0.003) | 0.340 | 0.003 (0.005) | 0.631 | −0.005 (0.003) | 0.063 |
| Leverage Ratio (%) | 0.010 (0.009) | 0.293 | 0.008 (0.009) | 0.393 | 0.037 (0.012) | 0.003 |
| Log Liquid Liabilities | 0.019 (0.044) | 0.674 | 0.015 (0.043) | 0.719 | 0.079 (0.056) | 0.159 |
| Log Fixed Assets | 0.180 (0.050) | 0.000 | 0.174 (0.049) | 0.000 | 0.219 (0.075) | 0.003 |
| Observations | 5145 | | 6650 | | 1792 | |

*Notes:* The table reports group-time ATT estimates under the doubly robust group time estimator for alternative matching specifications. NN = 2 and NN = 3 refer to nearest-neighbor PSM with two and three matches per treated unit, respectively. Mahalanobis refers to matching based on Mahalanobis distance; the reduced sample reflects the imposition of a balance threshold under high covariate collinearity. Standard errors in parentheses are clustered at the firm level.

Table 5 shows that the results are qualitatively stable across specifications. The positive and statistically significant effect on fixed assets persists under all three alternative matching procedures, with coefficients ranging from 0.174 to 0.219. Operating revenue, value added, liquid liabilities, and solvency ratio remain statistically insignificant throughout. Under Mahalanobis matching, the leverage ratio and employment show significant coefficients (group-time ATT = 0.037, $p < 0.01$ and group-time ATT = 0.046, $p < 0.05$, respectively) that are not significant under PSM specifications, which may reflect compositional differences in the reduced Mahalanobis sample rather than a genuine treatment effect. Overall, the main finding is that a positive and robust effect on fixed asset accumulation, is not driven by the specific matching algorithm employed.

**Table 6** Robustness check: placebo test

| Outcome Variable | ATT | (SE) | $p$-value |
|---|---|---|---|
| Log operating revenue | 0.058 | (0.028) | 0.037 |
| Log number of employees | 0.018 | (0.019) | 0.344 |
| Log value added | 0.108 | (0.180) | 0.547 |
| Solvency ratio | 0.001 | (0.002) | 0.746 |
| Leverage ratio | 0.011 | (0.013) | 0.377 |
| Log liquid liabilities | 0.028 | (0.040) | 0.483 |
| Log fixed assets | 0.052 | (0.042) | 0.215 |

*Notes:* Placebo treatment dates are assigned two years prior to the actual treatment date. The model is re-estimated using only pre-treatment observations. Estimates follow the doubly robust procedure. Standard errors (SE) in parentheses are clustered at the firm level. Some group-time cells could not be estimated due to insufficient observations in the shorter placebo window.

Table 6 reports the placebo estimates obtained by assigning fictitious treatment dates two years prior to the actual grant award and re-estimating the baseline model on pre-treatment observations. Six of the seven outcomes show coefficients that are small in magnitude and statistically indistinguishable from zero, consistent with the absence of anticipatory effects. The exception is log operating revenue, which yields a statistically significant coefficient at the 5% level (ATT = 0.058, SE = 0.028, $p = 0.037$). We note that the placebo window is shorter than the main estimation window, which reduces the number of group-time cells available for estimation and may inflate variance in individual coefficients. Importantly, the pre-treatment event-study coefficients for revenue in the main specification are consistently small and statistically indistinguishable from zero across all four pre-treatment periods (see Figure 3), and the six remaining outcomes show no pre-treatment movement in the placebo exercise. We therefore interpret this result with caution: while it does not by itself invalidate the identification strategy, it does limit the strength of the null revenue result and should be borne in mind when interpreting the absence of a revenue effect in the main analysis.

# 6 Conclusion

Public investment subsidies are among the most widely deployed instruments of SME policy in the European Union, yet the conditions under which they actually work remain poorly understood. This paper contributes with causal evidence on that question using Spanish administrative data for the period 2018–2024, combining propensity score matching with the doubly robust group-time estimator of Callaway and Sant'Anna (2021) to account for staggered treatment timing, non-random selection, and treatment effect heterogeneity across firms and cohorts.

Fixed asset accumulation is the one outcome that moves consistently and significantly across all specifications. Subsidies raise the capital stock of treated firms by roughly 20 log points, a finding that survives alternative matching procedures and a placebo

test. Revenue, value added, employment, and financial ratios show no statistically significant short-run response, at least in the general specification. The natural reading of this pattern is that investment grants do what they are intended to do: they relax capital constraints for firms that, absent public support, would have lacked the means to undertake the investment on their own.

Firm size turns out to be a relevant moderating variable. Among micro-enterprises, the investment response is roughly four times larger than among firms with ten or more employees, and subsidies also generate modest but statistically significant employment growth in this subgroup. Larger firms show effects that are generally indistinguishable from zero. The dose-response analysis for micro-enterprises tells the same story from a different angle: the investment response grows with grant size, and revenue and employment only move at the top of the support distribution. For micro-enterprises, grant size is what determines whether the transfer is large enough to shift behaviour in the short run.

Among the single-instrument recipients in our sample, the evidence points in a consistent direction. Spreading grants thinly across firms of all sizes generates investment responses mainly in micro-enterprises: larger firms show little response, while micro-enterprises exhibit effects that grow with grant size. For the firms we observe, concentrating public resources on financially constrained micro-enterprises with grants large enough to matter is what drives the investment response; fewer but larger awards appear to be better value for public money.

Several limitations bound these conclusions. The post-treatment window of two years rules out any assessment of medium-run productivity or survival effects, and the null results for revenue and employment should be read accordingly. The design cannot separate the effect of the transfer itself from a certification channel, whereby selection as a subsidy recipient signals creditworthiness to external lenders regardless of the funds received (Bronzini & Iachini, 2014; Howell, 2017); the leverage increase among micro-enterprises is consistent with both stories. The restriction to single-instrument recipients strengthens internal validity but retains only 274 of roughly 6884 potential treated firms, and whether the findings generalise to firms embedded in more complex support portfolios remains open. Finally, the exact industry matching that underpins the identification strategy concentrates treated and control firms within the same narrow sectors; if treated firms generate spillovers on nearby competitors through labour markets or supply chains, the estimated effects would understate the true treatment impact (Criscuolo et al., 2019). Addressing these limitations will require longer panels, richer instrument-level records, and research designs capable of tracing the full chain from capital accumulation to firm growth.

# 7 Declarations

**Conflict of interest.** The authors declare no conflict of interest.

**Data availability.** The subsidy data used in this study are drawn from the Base de Datos Nacional de Subvenciones (BDNS), which is publicly available at https://www.infosubvenciones.es. Firm-level financial data are sourced from the Sistema de Análisis de Balances Ibéricos (SABI), provided by Bureau Van Dijk. SABI data are proprietary and subject to licensing restrictions.

# Appendix A Additional Diagnostics

**Table A1**: Covariate balance before and after matching

| | Before matching | | | | After matching | | | | % Improvement | |
|---|---|---|---|---|---|---|---|---|---|---|
| Covariate | $\bar{X}_T$ | $\bar{X}_C$ | ASMD | eCDF | $\bar{X}_T$ | $\bar{X}_C$ | ASMD | eCDF | ASMD | eCDF |
| *Performance outcomes (t-1)* | | | | | | | | | | |
| Log operating revenue | 14.56 | 13.57 | 0.67 | 0.26 | 14.29 | 14.36 | 0.05 | 0.10 | 92.09 | 63.75 |
| Log employment | 2.73 | 1.96 | 0.70 | 0.30 | 2.50 | 2.58 | 0.08 | 0.10 | 88.68 | 66.98 |
| Log value added | 13.21 | 12.34 | 0.32 | 0.29 | 13.04 | 13.22 | 0.08 | 0.09 | 74.03 | 68.93 |
| Log liquid liabilities | 13.18 | 12.10 | 0.66 | 0.27 | 12.87 | 12.91 | 0.02 | 0.08 | 96.25 | 71.85 |
| Log fixed assets | 13.40 | 11.75 | 0.80 | 0.33 | 13.07 | 13.00 | 0.04 | 0.06 | 95.24 | 80.77 |
| *Financial structure (t-1)* | | | | | | | | | | |
| Leverage ratio | 0.49 | 0.52 | 0.10 | 0.07 | 0.48 | 0.46 | 0.07 | 0.07 | 23.71 | −9.60 |
| Solvency ratio | 0.03 | 0.07 | 0.02 | 0.04 | 0.03 | 0.03 | 0.07 | 0.05 | −272.54 | −21.44 |
| *Firm characteristics (t-1)* | | | | | | | | | | |
| Firm age (years) | 23.24 | 21.93 | 0.11 | 0.08 | 22.66 | 24.36 | 0.14 | 0.09 | −28.34 | −16.00 |
| *Performance outcomes (t-2)* | | | | | | | | | | |
| Log operating revenue | 14.41 | 13.43 | 0.67 | 0.28 | 14.14 | 14.23 | 0.07 | 0.10 | 90.19 | 65.31 |
| Log employment | 2.67 | 1.94 | 0.67 | 0.28 | 2.45 | 2.53 | 0.08 | 0.10 | 87.63 | 62.63 |
| Log value added | 13.10 | 12.05 | 0.35 | 0.31 | 12.81 | 12.87 | 0.02 | 0.09 | 94.34 | 70.37 |
| Log liquid liabilities | 13.05 | 12.03 | 0.64 | 0.26 | 12.75 | 12.85 | 0.06 | 0.08 | 89.82 | 69.26 |
| Log fixed assets | 13.26 | 11.72 | 0.74 | 0.30 | 12.93 | 12.92 | 0.01 | 0.07 | 98.73 | 77.70 |
| *Financial structure (t-2)* | | | | | | | | | | |
| Solvency ratio | 0.03 | 0.10 | 0.02 | 0.05 | 0.03 | 0.03 | 0.03 | 0.06 | −33.68 | −11.19 |
| *Performance outcomes (t-3)* | | | | | | | | | | |
| Log operating revenue | 14.34 | 13.43 | 0.62 | 0.25 | 14.08 | 14.20 | 0.09 | 0.12 | 86.12 | 53.14 |
| Log employment | 2.67 | 1.95 | 0.65 | 0.28 | 2.44 | 2.51 | 0.07 | 0.09 | 89.31 | 69.24 |
| Log value added | 12.95 | 12.05 | 0.28 | 0.29 | 12.77 | 13.03 | 0.10 | 0.11 | 63.99 | 63.26 |
| Log liquid liabilities | 12.99 | 12.02 | 0.60 | 0.25 | 12.71 | 12.81 | 0.07 | 0.09 | 88.81 | 64.59 |
| Log fixed assets | 13.20 | 11.69 | 0.73 | 0.30 | 12.87 | 12.88 | 0.01 | 0.05 | 99.31 | 82.53 |

*Continued on next page*

Table A1 — *continued*

| Covariate | Before matching | | | | After matching | | | | % Improvement | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\bar{X}_T$ | $\bar{X}_C$ | ASMD | eCDF | $\bar{X}_T$ | $\bar{X}_C$ | ASMD | eCDF | ASMD | eCDF |
| *Financial structure (t-3)* | | | | | | | | | | |
| Solvency ratio | 0.03 | 0.08 | 0.02 | 0.05 | 0.03 | 0.03 | 0.04 | 0.06 | −153.27 | −37.99 |
| *Performance outcomes (t-4)* | | | | | | | | | | |
| Log operating revenue | 14.41 | 13.48 | 0.64 | 0.25 | 14.14 | 14.23 | 0.07 | 0.11 | 89.72 | 57.27 |
| Log employment | 2.65 | 1.96 | 0.63 | 0.27 | 2.43 | 2.53 | 0.10 | 0.11 | 83.48 | 60.54 |
| Log value added | 13.27 | 12.21 | 0.45 | 0.30 | 12.99 | 12.96 | 0.02 | 0.09 | 96.46 | 70.32 |
| Log liquid liabilities | 13.00 | 12.05 | 0.60 | 0.24 | 12.72 | 12.82 | 0.07 | 0.11 | 88.95 | 54.19 |
| Log fixed assets | 13.14 | 11.65 | 0.73 | 0.29 | 12.83 | 12.82 | 0.00 | 0.06 | 99.85 | 79.74 |
| *Financial structure (t-4)* | | | | | | | | | | |
| Solvency ratio | 0.03 | 0.05 | 0.04 | 0.04 | 0.03 | 0.03 | 0.04 | 0.06 | 16.16 | −42.99 |
| *Trade status* | | | | | | | | | | |
| Exporter only | 0.07 | 0.04 | 0.16 | 0.04 | 0.07 | 0.08 | 0.04 | 0.01 | 72.08 | 67.05 |
| Importer only | 0.06 | 0.07 | 0.05 | 0.01 | 0.06 | 0.06 | 0.03 | 0.01 | 34.75 | 37.05 |
| Two-way trader | 0.20 | 0.09 | 0.33 | 0.12 | 0.16 | 0.25 | 0.22 | 0.09 | 34.40 | 23.97 |

*Notes:* $\bar{X}_T$ and $\bar{X}_C$ denote the mean for treated and control firms, respectively. ASMD is the absolute standardized mean difference. eCDF is the maximum difference in empirical cumulative distribution functions. % Improvement reports the percentage reduction in each imbalance measure after matching; negative values indicate a deterioration in balance for that covariate.

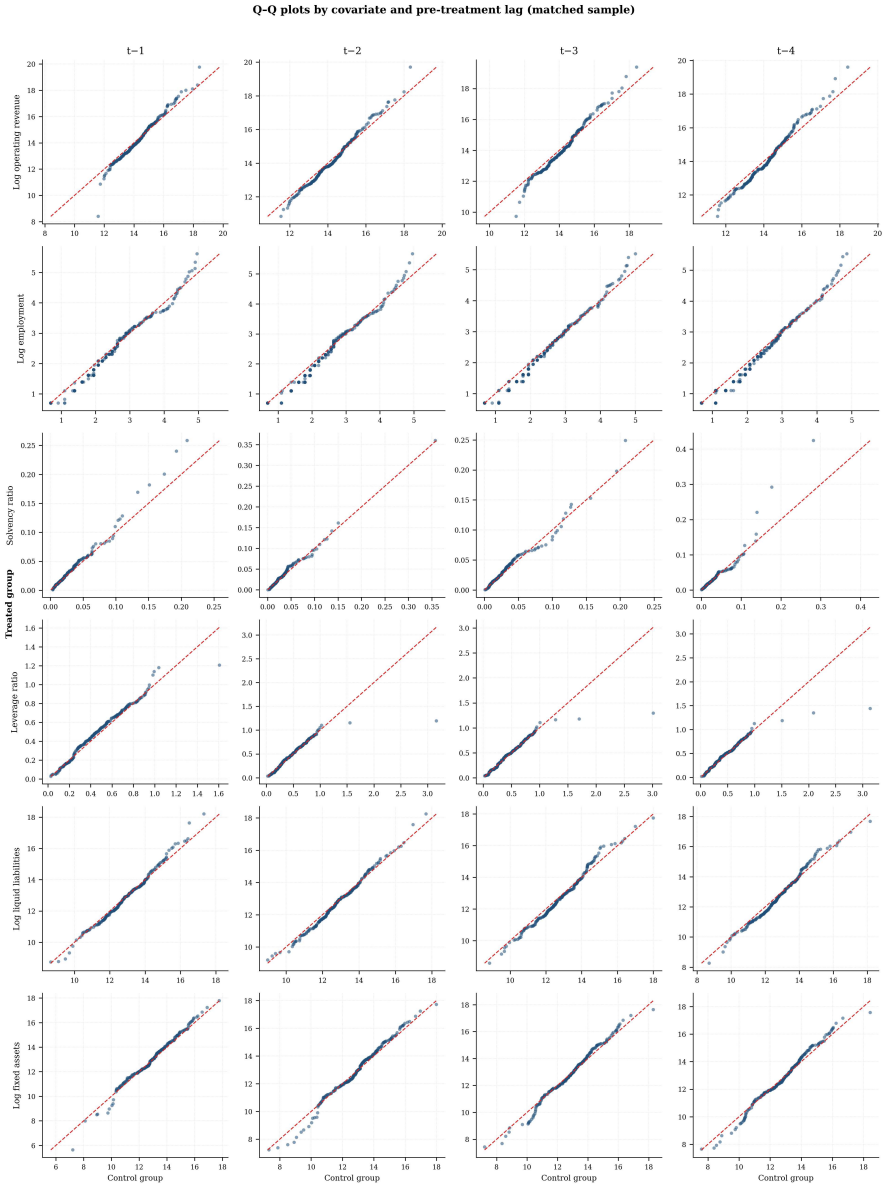



**Fig. A1** Q-Q plots by covariate and pre-treatment lag (matched sample). Each panel compares the empirical quantile distributions of treated and control firms for a given variable and pre-treatment period. Proximity to the 45-degree line indicates distributional similarity between groups after matching.

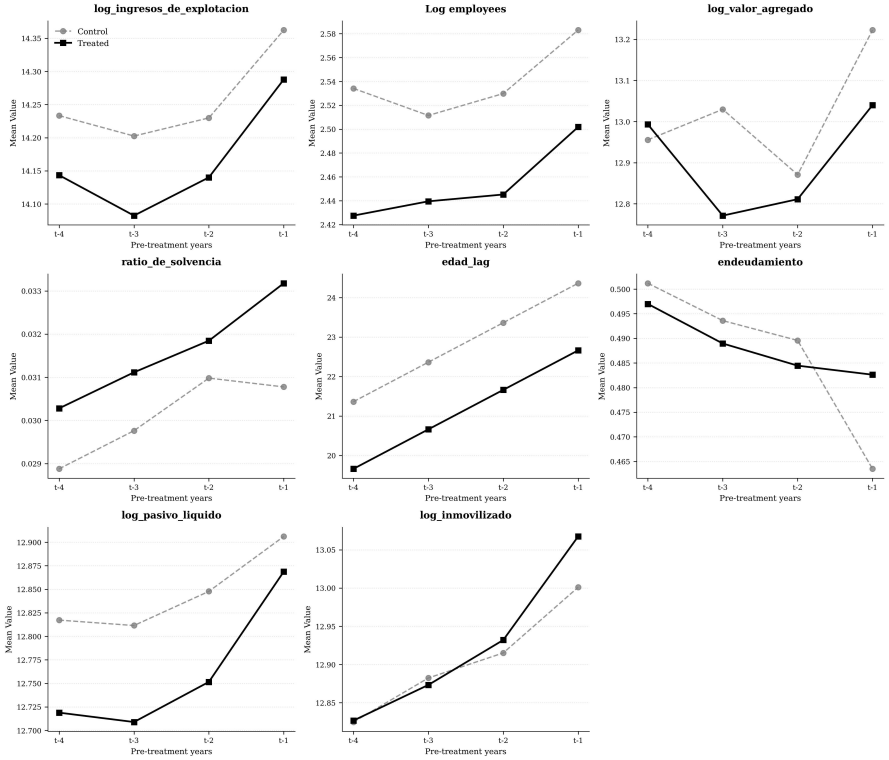



**Fig. A2** Pre-treatment outcome trajectories for treated and control firms across four pre-treatment periods ($t-4$ to $t-1$). Each panel displays the mean evolution of a given outcome variable separately for each group. Comparable trends prior to treatment provide graphical support for the parallel trends assumption.

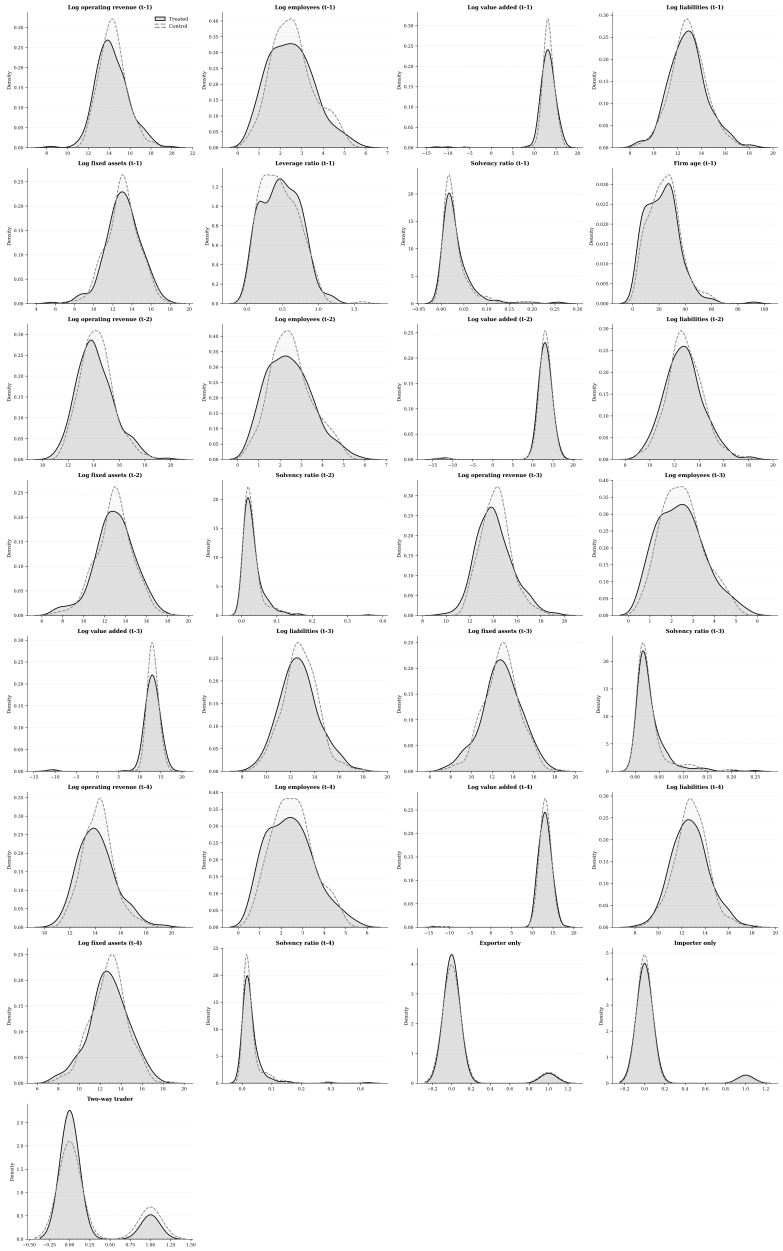



**Fig. A3** Distribution of estimated propensity scores for treated and control firms after matching. Each panel corresponds to a covariate and pre-treatment period. Sufficient overlap between the two distributions supports the common support assumption underlying the matching procedure.

# References


Angrist, J.D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Austin, P.C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, *28*(25), 3083–3107, https://doi.org/10.1002/sim.3697

Beck, T., & Demirgüç-Kunt, A. (2006). Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking & Finance*, *30*(11), 2931–2943, https://doi.org/10.1016/j.jbankfin.2006.05.009

Bernini, C., Cerqua, A., Pellegrini, G. (2017). Public subsidies, TFP and efficiency: A tale of complex relationships. *Research Policy*, *46*(4), 751–767, https://doi.org/10.1016/j.respol.2017.02.001

Blanes, J.V., & Busom, I. (2004). Who participates in R&D subsidy programs?: The case of spanish manufacturing firms. *Research Policy*, *33*(10), 1459–1476, https://doi.org/10.1016/j.respol.2004.07.006

Borusyak, K., Jaravel, X., Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, *91*(6), 3253–3285, https://doi.org/10.1093/restud/rdae007

Bronzini, R., & Iachini, E. (2014). Are incentives for r&d effective? evidence from a regression discontinuity approach. *American Economic Journal: Economic Policy*, *6*(4), 100–134, https://doi.org/10.1257/pol.6.4.100

Callaway, B., & Sant'Anna, P.H.C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230, https://doi.org/10.1016/j.jeconom.2020.12.001

Canzian, G., Crivellaro, E., Duso, T., Ferrara, A.R., Sasso, A., Verzillo, S. (2025). The impact of financial support to firms during crises: The case of covid aid in the eu. *Journal of Economic Behavior & Organization*, *240*, 107339, https://doi.org/10.1016/j.jebo.2025.107339

Card, D., Kluve, J., Weber, A. (2018). What works? a meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, *16*(3), 894–931, https://doi.org/10.1093/jeea/jvx028

Cerqua, A., & Pellegrini, G. (2017). Industrial policy evaluation in the presence of spillovers. *Small Business Economics*, *49*(3), 671–686, https://doi.org/10.1007/s11187-017-9855-9

Chabé-Ferret, S. (2017, June). *Should we combine difference in differences with conditioning on pre-treatment outcomes?* (Tech. Rep. No. 17-824). Toulouse School of Economics. Retrieved from https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/wp/2017/wp_tse_824.pdf

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*(1), C1–C68, https://doi.org/10.1111/ectj.12097

Criscuolo, C., Martin, R., Overman, H.G., Van Reenen, J. (2019). Some causal effects of an industrial policy. *American Economic Review*, *109*(1), 48–85, https://doi.org/10.1257/aer.20160034

Czarnitzki, D., & Delanote, J. (2015). R&d policies for young smes: input and output effects. *Small Business Economics*, *45*(3), 465–485, https://doi.org/10.1007/s11187-015-9661-3

Czarnitzki, D., & Lopes-Bento, C. (2013). Value for money? new microeconometric evidence on public R&D grants in Flanders. *Research Policy*, *42*(1), 76–89, https://doi.org/10.1016/j.respol.2012.04.008

de Mel, S., McKenzie, D., Woodruff, C. (2008). Returns to capital in microenterprises: Evidence from a field experiment. *The Quarterly Journal of Economics*, *123*(4), 1329–1372, https://doi.org/10.1162/qjec.2008.123.4.1329

Dimos, C., & Pugh, G. (2016). The effectiveness of r&d subsidies: A meta-regression analysis of the evaluation literature. *Research Policy*, *45*(4), 797–815, https://doi.org/10.1016/j.respol.2016.01.002

Dvouletý, O., & Blažková, I. (2019). Assessing the microeconomic effects of public subsidies on the performance of firms in the Czech food processing industry: A counterfactual impact evaluation. *Agribusiness*, *35*(3), 394–422, https://doi.org/10.1002/agr.21575

Dvouletý, O., Srhoj, S., Pantea, S. (2021, 6). Public SME grants and firm performance in european union: A systematic review of empirical evidence. *Small Business Economics*, *57*, 243-263, https://doi.org/10.1007/s11187-019-00306-x

European Commission (2021). *Annual report on european smes 2020/2021* (Tech. Rep.). European Commission. Retrieved from https://single-market-economy.ec.europa.eu/smes/sme-strategy_en

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277, https://doi.org/10.1016/j.jeconom.2021.03.014

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, *20*(1), 25–46, https://doi.org/10.1093/pan/mpr025

Hall, B.H., & Lerner, J. (2010). The financing of r&d and innovation. B.H. Hall & N. Rosenberg (Eds.), *Handbook of the economics of innovation* (Vol. 1, pp. 609–639). Amsterdam: North-Holland.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. (1986). *Robust statistics: The approach based on influence functions.* New York: John Wiley & Sons.

Heckman, J.J., Ichimura, H., Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, *64*(4), 605–654, https://doi.org/10.2307/2971733

Howell, S.T. (2017). Financing innovation: Evidence from R&D grants. *American Economic Review*, *107*(4), 1136–1164, https://doi.org/10.1257/aer.20150808

Huergo, E., & Moreno, L. (2017). Subsidies or loans? evaluating the impact of R&D support programmes. *Research Policy*, *46*(7), 1198–1214, https://doi.org/10.1016/j.respol.2017.05.006

Imbens, G.W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635, https://doi.org/10.1016/j.jeconom.2007.05.001

Imbens, G.W., & Wooldridge, J.M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86, https://doi.org/10.1257/jel.47.1.5

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, *27*(4), 435–454, https://doi.org/10.1017/pan.2019.11

Ma, Y., Sant'Anna, P.H.C., Sasaki, Y., Ura, T. (2023). Doubly robust estimators with weak overlap. *arXiv preprint arXiv:2304.08974*, ,

Marcus, M., & Sant'Anna, P.H.C. (2021). The role of parallel trends in event study settings: An application to environmental economics. *Journal of the Association of Environmental and Resource Economists*, *8*(2), 235–275, https://doi.org/10.1086/711509

Munch, J., & Schaur, G. (2018). The effect of export promotion on firm-level performance. *American Economic Journal: Economic Policy*, *10*(1), 357–387, https://doi.org/10.1257/pol.20150410

OECD (2020). *Oecd economic outlook, volume 2020 issue 1* (Tech. Rep.). Organisation for Economic Co-operation and Development. Retrieved from https://www.oecd.org/economic-outlook/

Peric, M., & Vitezić, V. (2016). Impact of global economic crisis on firm growth. *Small Business Economics*, *46*(1), 1–12, https://doi.org/10.1007/s11187-015-9671-z

Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55, https://doi.org/10.1093/biomet/70.1.41

Sant'Anna, P.H.C., & Zhao, J.B. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101–122, https://doi.org/10.1016/j.jeconom.2020.06.003

Srhoj, S., Lapinski, M., Walde, J. (2021). Impact evaluation of business development grants on sme performance. *Small Business Economics*, *57*(3), 1285–1301, https://doi.org/10.1007/s11187-020-00348-6

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199, https://doi.org/10.1016/j.jeconom.2020.09.006

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242, https://doi.org/10.1080/01621459.2017.1319839

Zúñiga-Vicente, J.Á., Alonso-Borrego, C., Forcadell, F.J., Galán, J.I. (2014). Assessing the effect of public subsidies on firm R&D investment: a survey. *Journal of Economic Surveys*, *28*(1), 36–67, https://doi.org/10.1111/j.1467-6419.2012.00738.x