TI 2022-061/V
Tinbergen Institute Discussion Paper

Economic performance and investments under emissions trading: Untangling the effects of a staggered regulation

Revision: February 2024

Leon Bremer[1]
Konstantin Sommer[2]

1 Vrije Universiteit Amsterdam and Tinbergen Institute

2 University of Amsterdam and Tinbergen Institute

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at https://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Economic performance and investments under emissions trading: Untangling the effects of a staggered regulation

Leon Bremer[*] [‡]   Konstantin Sommer[†] [‡]

February 9, 2024

Abstract

We study the effects of the EU Emissions Trading System on the economic performance and investments of Dutch manufacturing firms. Motivated both by sizable differences between firms that became regulated in different phases and by a gradual increase in regulatory stringency, we pay close attention to the staggered design of the ETS as well as to potential treatment effect heterogeneity. We base our estimation on recent advances in the estimation of treatment effects and make use of administrative microdata. Our results align with those of the previous literature. Even when studying the more stringent third phase and when using estimators fit to the staggered ETS setting, there seems to be no discernible effects of the ETS on firms' economic performance. We also do not find any statistically significant effect on the investment behavior of regulated firms.

JEL codes: C23; H23; L51; Q52; Q58
Keywords: Emissions trading, Environmental regulation, Staggered Difference-in-Differences, Treatment heterogeneity, Manufacturing

[*] Vrije Universiteit Amsterdam, Department of Spatial Economics, De Boelelaan 1105, 1081 HV Amsterdam, and Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam. E-mail: l.bremer@vu.nl.

[†] University of Amsterdam, Department of Macro and International Economics, Roetersstraat 11, 1018 WB Amsterdam, and Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam. E-mail: k.h.l.sommer2@uva.nl. (corresponding author)

[‡] The authors are grateful for funding by the A Sustainable Future (ASF) platform. The authors are also grateful for the opportunities to present at various conferences and seminars, and the feedback received by various participants of these sessions.

1 Introduction

This paper studies the effects of the European Union Emissions Trading System (ETS) on the economic performance and investment decisions of regulated manufacturing firms in the Netherlands. The literature on the ex-post evaluation of the ETS has commonly used two-way fixed effects (TWFE), an OLS estimator that controls for time and unit fixed effects. However, recent econometric advances have demonstrated that TWFE can lead to biased results in staggered treatment settings. The ETS presents such a staggered setting, as firms enter the regulation in different phases. We therefore extend this literature by using estimation techniques that adequately control for this staggered design of the ETS. We also show that firms that became regulated in different phases of the ETS differ from each other and that the stringency of the regulation varies significantly over time, making it important to adequately think about the appropriate level of treatment effect heterogeneity.

The EU ETS is the world's largest cap-and-trade system, regulating the largest emitters of greenhouse gases in Europe and covering about 40% of the EU's emissions. The ETS started operating in 2005 and has been amended throughout its four phases, in each of which additional installations were regulated and in which the regulation and its stringency were adapted. Where the need for climate action is widely acknowledged, policy makers have expressed concerns about unilateral climate action. Stringent policies could unintentionally lead to reductions in profits and employment. At the same time, however, climate policies could incentivize firms to adapt production through investments.

This study aims to estimate the causal effect of the ETS on both of these potential policy side effects. So far the literature has found little to no effect from the ETS regulation on such measures. We, however, deviate from the existing studies in two important ways. First, we carefully fit the staggered treatment of the first three EU ETS phases by using appropriate econometric methods. We take into account that treatment varies between phases and that different groups of firms (hereafter *cohorts*) enter treatment in different phases. Based on this, we carefully discuss and distinguish several forms of effect heterogeneity, and show that ignoring these could result in misleading estimates. Second, we include the more recent years of the regulation in which allowance prices rose, free allowance allocations were reduced, and in which amendments like the Market Stability Reserve (MSR) were introduced. Studying these different periods allows us to see if phases with higher regulatory stringency strengthen the effects of the ETS on firm outcomes.

We use detailed firm-level microdata from Statistics Netherlands (in Dutch: CBS), the Dutch national statistics agency, and link those to the European Union Transaction Log (EUTL) for information on regulated ETS firms. We are able to link most of the treated firms in the anonymized CBS data. We show descriptively that firms that were regulated in phase 1 are far more energy-intensive than later treated firms, and that regulation stringency, measured both in terms of carbon prices and through the allocation of allowances, varied substantially by phase.

We start our analysis by presenting a decomposition of the TWFE estimator that shows the extent of its bias in staggered treatment settings, based on Goodman-Bacon (2021). The main reason for this bias in the ETS context is that firms became regulated in different phases, which causes the TWFE estimator to implicitly use earlier regulated firms as controls for later regulated firms (a "forbidden comparison"). This would only be valid if treatment effects would materialize once and stay constant thereafter, which is unlikely in the context of the ETS. In our setting, the Goodman-Bacon (2021) decomposition shows that TWFE estimates of the ETS are based on some problematic comparisons, but that the weights on these comparisons are low.

For the ETS treatment effect estimation, we then employ a recent estimator developed in Callaway and Sant'Anna (2021) (hereafter *CS*) that is designed for settings with staggered treatment timing. The estimator allows us to carefully control for pre-treatment differences between regulated and non-regulated firms, and estimates the treatment effect for each cohort and year combination. It thereby avoids the forbidden comparison. We then aggregate these individual cohort-year estimates to find reasonable estimates of the ATT for different cohort and phase aggregations. We present aggregations of our results from the CS estimator for the whole ETS, for different cohorts of firms, and for each cohort and phase combination. We compare the findings of the CS estimator with those of a matched TWFE estimator.

We capture economic performance by both firm size and profitability. Firm size is measured by employment and turnover, which is defined as the aggregate revenue derived from industrial product sales. For profitability we study firms' profit margins. Investment is captured by the investment intensity, i.e. total investments as a share of turnover, and we study both investments into total fixed assets as well as into machines only.

Our results align with those of previous studies on the ETS. Even when studying the more stringent phase 3 and even when adequately controlling for the staggered design of the ETS, we do not find evidence for a reduction in economic performance of regulated firms.

We also do not establish any significant effects on the investment intensity of regulated firms, neither into total assets nor into machines. Testing the robustness of our results and discussing the underlying assumptions confirms our results.

Our estimation does result in some statistically negative coefficient estimates for the effect of the ETS on turnover. This effect is largely driven by the most energy-intensive firms in cohort 1. The estimate is statistically significant at the

10 percent level in some specifications, but does not survive all robustness tests. The effect is thus uncertain, but if anything driven by the firms where the effect is most likely to be observed.

In most cases the TWFE estimates align with those of our preferred CS estimation. In the few cases where one of the estimators leads to a statistically significant estimate, the other estimator aligns qualitatively. Taken together with the results of the Goodman-Bacon (2021) decomposition, this implies that the problem of the non-allowed comparisons seems to be small in our setting.

The paper continues as follows: Section 2 discusses the related literature. The data and policy background are discussed in Section 3. The methodology and results are presented in Section 4 and Section 5, respectively. Section 6 shows the robustness of these results and discusses the assumptions underlying our identification. Section 7 concludes this research.

2 Related literature

Our study contributes to the existing body of research on the unintended consequences of environmental policies and carbon pricing schemes on regulated firms. In a review article, Venmans et al. (2020) come to the conclusion that most carbon pricing schemes have shown little to no statistically significant effects on economic performance. The authors attribute this to low prices (indicating low regulation stringency) and industry protection. Our study, therefore, places particular emphasis on the examination of periods marked by varying degrees of regulatory stringency. The EU's carbon trading scheme, one of the most established and stringent carbon pricing schemes globally, provides an ideal framework to explore diverse levels of stringency across extended time periods.

There are several studies on the effects of the EU ETS on the competitiveness and economic performance of regulated firms. Several of these studies use administrative firm-level data in other countries and apply standard difference-in-differences methods. Other studies use a larger set of EU ETS firms combined with publicly available data sets (e.g. Calel & Dechezleprêtre, 2016). Underlying all studies is the complexity of finding appropriate control firms that are unregulated but sufficiently similar in order to draw causal conclusions. However, we are aware of no study that fully appreciates the staggered nature of the ETS and relies on the recent methodologies developed for these situations.

Most studies estimate the treatment effect on the treated by using the semi-parametric estimator of Heckman et al. (1997) or by using a TWFE regression. Using the Heckman et al. (1997) methodology, Petrick and Wagner (2014) and Jaraite-Kažukauske and Di Maria (2016) find no negative effects of the ETS on productivity and employment for Germany and Lithuania, respectively, and Colmer et al. (2022) find no evidence of outsourcing in France. Marin et al. (2018), using non-administrative microdata from Bureau van Dijk for a larger set of countries, also do not find negative effects on economic performance, but do find an increase in labor productivity. Although the Heckman et al. (1997) estimator in theory would allow researchers to study heterogeneity over time and between different treatment cohorts, this has not been done yet. Löschel et al. (2019) additionally use a TWFE setting to analyze the ETS's effect on productivity in Germany. The authors interestingly find significant positive effects on productivity using the Heckman-style estimator, but not in the regression estimation. According to the authors, this effect is likely driven by a positive EU ETS effect on efficiency in some of the regulated industries. Dechezleprêtre et al. (2023) estimate the effect of the EU ETS on revenue, fixed assets, employees, and EBIT, using a matched TWFE estimator. The authors also do not find evidence for any significant effect on profits or employment and even find an increase in revenues and assets.

However, all of these studies only use data for the first phase and some years into the second phase. Since the stringency of the ETS increased significantly in the second and third phases, adding later phases may lead to stronger and clearer results. However, two studies that look at later phases also do not find strong negative effects of ETS regulation. Dechezleprêtre et al. (2022) use data on multinational firms up to 2014 and analyze carbon shifting within these firms, again without finding much evidence of leakage. Klemetsen et al. (2020) do look at phase 3 and analyze firms in Norway. Their fixed effect regression methodology accounts for different effects between phases, but not between companies starting in different phases. They find a slight increase in productivity in phase 2, but no significant effect in the other phases.

In a literature survey, Verde (2020) comes to the conclusion that there is no convincing evidence of leakage and losses in competitiveness due to the ETS yet. The authors also highlight that this might be due to the short time span covered in almost all studies and point to the importance of analyzing more long-term indicators, such as investments. This study aims to address both of these gaps. In another review, Joltreau and Sommerfeld (2019) additionally argue that energy costs remained a small share of their costs for most regulated firms. Our division into cohorts will as a byproduct also study energy-intensive firms separately.

When it comes to the ETS's effect on innovation, the literature is smaller but still contains important contributions. Calel and Dechezleprêtre (2016) show in a large multi-country panel that the ETS has increased green patenting, and Borghesi et al. (2015) show in an Italian phase 1 firm-level panel that regulated sectors have increased innovation, but that this varied by treatment stringency of the sector. However, a survey by Teixidó et al. (2019) comes to the conclusion that evidence on the ETS's effect on innovation is still too sparse for a coherent conclusion.

Our contribution to the existing research is threefold. First, we employ recent advances in econometric techniques to estimate the EU ETS's effects, comparing them with those obtained from a traditional DiD estimator to assess potential biases in previous studies. Second, we provide new insights into the potential treatment heterogeneity of the EU ETS, utilizing longer time series to estimate the effects of later phases and long-term effects from earlier phases. Third, we have access to detailed administrative data, including information on investments, along with traditional performance indicators. Furthermore, the Netherlands, with its export-oriented and energy-intensive industrial structure, is an ideal context to observe potential competitiveness effects.

3 Data and policy background

3.1 EU ETS policy background

The EU ETS regulates installations, which we will also refer to as plants. Each of these plants is registered under one owner, the account holder, at a time in the European Union Transaction Log. The number of active installations regulated in the Netherlands and their account holders can be found in Figure 1.[1] After its initial implementation in 2005 the ETS has been largely revised 3 times when new phases came into effect, in 2008, 2013 and 2021. Most of these revisions aimed at making the system more restrictive and effective, and, especially in the Netherlands, also led to an increase in the number of regulated plants. This study uses data until 2020, thereby excluding Phase 4.

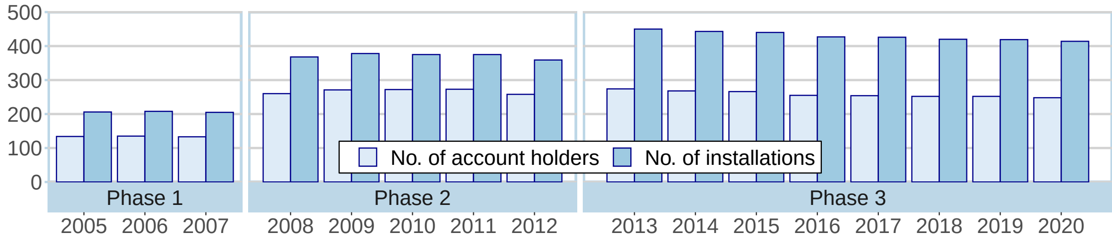


Figure 1: *Account holders and installations regulated under the EU ETS.*

Note: Number of active installations regulated under the EU ETS in the Netherlands and their account holders. Source: authors' calculations based on EUTL data accessed through EUETS.INFO.

[1] Only (former) Operator Holding Accounts that are registered in the Netherlands are selected. The connected installation must have positive verified emissions for that year.

In phase 1 (2005-2007), the number of allowances that were handed out was so high that the price of an allowance approached zero towards the end of the phase (Narassimhan et al., 2018), see Figure 2. In the Netherlands, actual emissions were almost 15% below the number of allocated allowances (Ellerman & Buchner, 2008). An important feature of the early phases of the ETS was that almost all allowances were distributed for free, called *grandfathering*, to avoid a reduction in the competitiveness of regulated firms. As grandfathering does not alter the opportunity costs of emissions, the idea was that free allocations would not change the abatement incentives of regulated firms (Woerdman et al., 2008).

Phase 2 (2008-2012) included nitrous oxide in the list of regulated greenhouse gases and increased the penalty for non-compliance from €40 to €100 per tonne of CO_2-equivalent. The number of regulated installations within the Netherlands increased from 205 to 368 (see Figure 1), mainly because in Phase 1 150 Dutch installations were excluded from the ETS.[2] This increase in regulated plants has important implications for the estimation of the ETS effect, as it turns the treatment design into a staggered one.

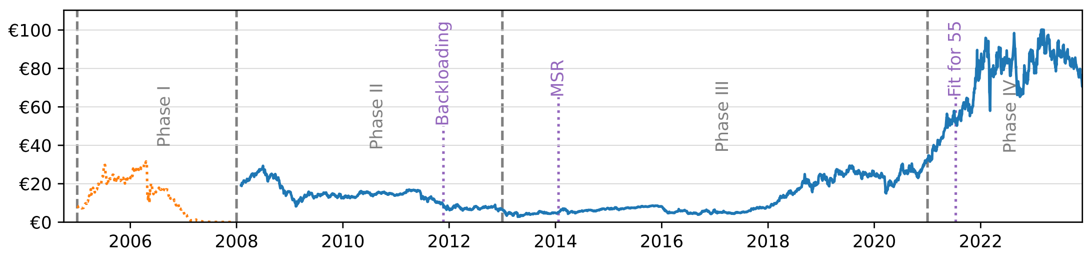


Figure 2: *EU ETS allowance price.*

Note: The EU ETS's allowance price in Euros per tonne of CO_2-equivalent. These are day closing prices for its futures contracts. The futures montage ECF00-NDEX is plotted in solid blue, and this data is accessed through FactSet. The December 2007 futures price for phase 1 allowances is plotted as a dotted orange line. These allowances were not transferable to later phases. The phase 1 data come from the European Environment Agency. Vertical dashed grey lines indicate the starts of a new phase, while purple dotted vertical lines indicate early proposal dates of amendments to the EU ETS.

More greenhouse gases were added to the regulation in Phase 3 (2013-2020)

[2]The following decisions by the European Commission (EC) provide further details of the phase 1 exemptions for Dutch installations. In October 2004 the EC exempted 93 installations and in March 2005 the EC exempted a further 57 installations (European Commission, 2004, 2005). Other countries that have exempted some firms from the regulation in the first phase were the UK, Sweden and Belgium. More information can be found on the EU Commission's website.

and the default allowance allocation method switched from grandfathering to auctioning. To counteract the low emission prices, the European Commission implemented two sets of new rules to the ETS. First, starting from 2014 the auctioning of new allowances was postponed until 2019-2020, which was referred to as Backloading. Second, in 2019 the Market Stability Reserve (MSR) began operating. The MSR takes the backloaded allowances and puts them in a reserve. Depending on demand and supply, allowances will be added to the reserve or released from the reserve. As of 2023 excess allowances in the reserve might be permanently canceled.

In addition, manufacturing sectors in the aluminium and chemicals production were added to the coverage in phase 3. This did not change the number of regulated account holders much, but it increased the number of regulated plants (see Figure 1). Arguably, in phase 3 more plants of the same owners were regulated.

Phase 4 (2021-2030) mainly sped up the rate at which the cap decreases over time, the Linear Reduction Factor, and it strengthened the MSR.[3]

The changing degree of regulatory stringency is also reflected in the allowance price path, as depicted in Figure 2.[4] Prices decreased to zero at the end of phase 1, then started around €20 in phase 2, but stayed around only €10 for several years. Although economists argue about the optimal price of carbon, a recent survey by Pindyck (2019) among economists and other scientists puts the average optimal carbon price among respondents clearly above 100$ per ton and thus much higher than those realized prices. Prices have started to increase since 2018 and reached €100 in February 2023, making the ETS much more restrictive in recent years.

3.2 EUTL and Dutch microdata

The data for this project come from two main sources. First, the European Union Transactions Log (EUTL) data is accessed through EUETS.INFO, a free service that provides cleaned data from the EUTL (Abrell, 2021). Second, Dutch firm-level data are accessed through the microdata services of Statistics Netherlands (in Dutch: CBS).

The data collected from the EUTL contain information on the free allocations of allowances, verified emissions, surrendered allowances, and the use of international credits, both by installation and account holder. Account holders in the EUTL can potentially own several regulated plants and are registered under a national identification number. As installations are assets, they can be purchased from or transferred to other firms. Such ownership changes are not perfectly captured by

[3]Please refer to the European Commission's webpage for more details.

[4]ETS stringency is not the only driver of the allowance price. A body of literature studying the ETS price drivers has, for example, identified fossil fuel prices as a driver of ETS prices (see e.g. Hintermann, 2010), but much of the variation is in fact hard to explain (Koch et al., 2014).

the data. Many installations do not change ownership between EU ETS phases in our data, but for the ones where it does change, we manually look up the date of ownership change using online public sources. Sources can be online news articles or websites that provide information about ownership structures. The list of manually assigned ownership changes and their respective source is available upon request. The data are organized in an unbalanced panel spanning the years 2005-2020 and a total of 439 unique account holders, owning 598 installations in the Netherlands.

The CBS data are not publicly accessible and are anonymized. They contain rich firm-level information on economic activity of almost the entire population of Dutch firms with more than 50 employees. The data contain information like the number of employees, costs of goods sold and turnover, as well as investment data. This study is limited in scope to manufacturing firms and relies on more than 40,000 firms over a time span of 21 years. To deflate monetary variables, we use Eurostat's industry producer price index for the Netherlands.

We link EUTL data to the administrative firm-level data of CBS. The linking takes place with the use of the chamber of commerce identifiers that are available in the EUTL and in the CBS data. Within CBS, several chamber of commerce numbers can comprise a "business unit", a construct defined by CBS and further explained in Appendix A.3. We will from here on refer to these business units as "firms". After linking the EUTL data to CBS's anonymized data, we are no longer able to identify individual firms.

As a business unit can comprise multiple account holders and plants, it can be the case that a business unit is regulated through more than one plant. We do not make a distinction here and consider each business unit (firm) as regulated if it owns at least one regulated plant in that year. Our level of analysis is at this business unit level, referred to as the firm level.

3.3 Sample selection

One issue with our panel arises from firm exit and, to a lesser degree, firm entry, from and to the sample. As we are dealing with anonymized microdata it is not possible to determine if such an exit is due to closure of the firm, an acquisition by another firm or due to changes in the firm structure. To minimize the effect that sample composition could have on our analyis, we curtail our sample to firms that we observe continuously from two years before to three years after treatment start. Unregulated firms also face this requirement when considered as a control unit.

We also enforce common support for all of our covariates (employment, energy costs, turnover, and total wage bill) between treated and control groups in the baseline years. Importantly, we only keep untreated firms in two-digit industries in which we also observe treated firms, as production processes might be considerably different in untreated sectors.

3.4 Measures of economic performance and investment intensity

We are interested in the ETS's effects on (1) economic performance and (2) investments. We measure these concepts with four dependent variables, namely, (1a) the firm's employment, (1b) its turnover, (1c) its profit margins, and (2) its investment intensity. Tracking employment outcomes also allows us to evaluate whether domestic environmental regulation leads to job losses at home, an often heard counterargument to unilateral environmental policy (see Vona, 2019, for a discussion of this argument). Likewise, turnover captures the overall sales of the firm's operation, potentially related to its market share, that one would expect to shrink if the firm would be harmed by the regulation. Profit margins directly assess the profitability of regulated firms. They also show to what extent regulated firms were able to charge a price above marginal costs, thus they also show if regulated firms were able to pass on additional costs of the regulation to their consumers. This ability probably decreases with the level of competition from abroad. Investment intensity estimates in how far firms are incentivized to invest in new technologies as a response to the regulation.

We measure employment in full-time equivalents (FTE). We define turnover as the aggregate revenue derived from industrial product sales in euros, and use the gross profit margin as the profit margin of interest. Gross profits measure the difference between turnover and the costs of goods sold. We scale it by turnover to transform it into a margin. Gross profits are generally not influenced by a firm's financial operations and thereby, for example, exclude a firm's income from holding activities.

As we are interested in the firm's responses in terms of updating or expanding its production capital, we turn to the investments firms make. We measure both investments in fixed assets and the more restrictive measure of investments in machines. We scale these measures by turnover to obtain a ratio that controls for the size of the firm's activities.

If variables are in monetary terms, they are deflated such that they can be compared over time. For this deflation we use Eurostat's industry producer price index for the Netherlands.

3.5 Heterogeneity between cohorts and phases

In this section, we elaborate on the two most important sources of heterogeneity that this study attempts to disentangle. First, we show the substantial differences between the regulated cohorts, and second, we show the development of the stringency of the ETS treatment over time.

Table 1: *Summary statistics by treatment group.*

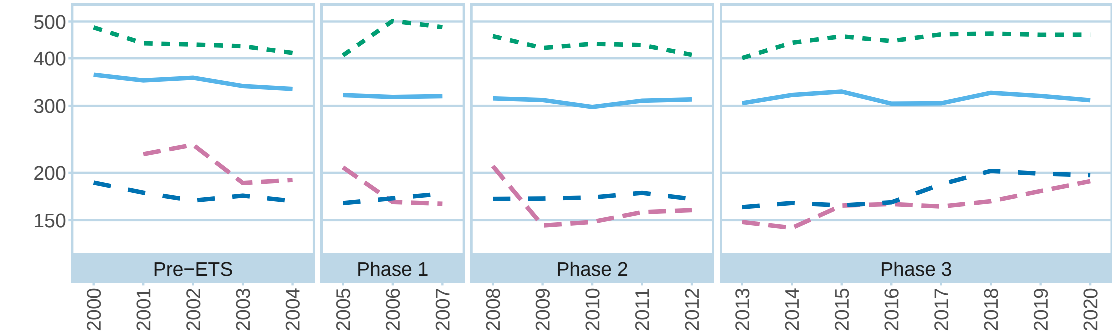
	Mean				P-value		
Variable	Control	C1	C2	C3	C1	C2	C3
Employment (FTE)	236.97	453.43	456.69	277.24	0.00	0.01	0.63
Turnover (Mil Euro)	74.52	196.16	197.20	117.38	0.00	0.14	0.16
Gross Profit Margin	0.45	0.51	0.64	0.39	0.04	0.00	0.09
Operating Margin	0.09	0.11	0.18	0.07	0.29	0.00	0.28
Energy Costs (Mil Euro)	1.60	8.99	4.04	2.55	0.00	0.02	0.57
Wage Bill (Mil Euro)	8.42	20.25	17.86	10.08	0.00	0.01	0.46
Investments fixed assets/Turnover	0.04	0.06	0.06	0.08	0.18	0.10	0.20
Investments Machines/Turnover	0.03	0.05	0.05	0.05	0.16	0.09	0.33
Investments/ Employees (Th Euro/FTE)	16.02	23.09	24.20	33.00	0.14	0.19	0.08
Observations	348	51	37	17			

The used data is based on all pre-estimation adjustments. Values are based on the year 2003, which is before the treatment for all cohorts. P-values are based on T-tests for the difference in means between the treated cohort and the control group.

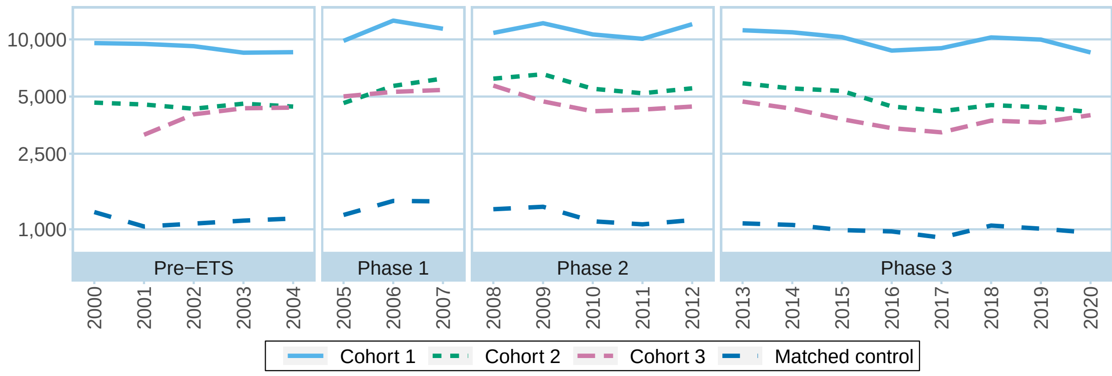
Table 1 shows the mean of different variables for the different treatment cohorts and our remaining set of control firms. Notable is that even though cohorts 1 and 2 are comparable in terms of size, both measured in employment and turnover, cohort 1 is considerably more energy-intensive than cohort 2. Cohort 3 is much smaller in both dimensions. This is not surprising, as the ETS aimed to regulate the most energy-intensive and large firms first. The firms are more comparable in terms of their investment-intensity and profit margins, also with respect to the control sample. Although it is clear that the control sample differs in terms of levels before the treatment, it is important to keep in mind that similarity in levels is not required for a difference-in-differences analysis, where we instead require the parallel trends assumption to hold.

Figure 3 shows the development of the average firm over time for energy expenditure and employment. The plot shows averages for the different ETS cohorts as well as for a set of matched control firms that is chosen to be as similar to the treated firms as possible. The methodology for this is detailed in Appendix C.

One can again see that firms regulated in 2005 are by far the largest energy consumers. Note that these are energy expenses and that energy prices are responsible for some of the time variation. Energy consumption data is unfortunately unavailable within the CBS data. From these plots alone, it is difficult to hypothesize about the estimated treatment effect, as panel (a) does not show clear kinks at the treatment dates. The plots, however, also do not give concern to a violation of the parallel trends assumption, as pre-treatment trends in both variables do not seem to drastically deviate between treated and control firms. We will, however, revisit this issue more thoroughly later.



(a) *Employment (in full-time equivalents)*



(b) *Energy expenses (in thousands 2015 Euros)*

Figure 3: *Averages of selected variables over time.*

Note: Cohorts 1-3 consist of firms first regulated in Phases 1-3, respectively. All groups are the result from the matching algorithm in Appendix C. The groups can therefore slightly differ from the ones in Table 1, which are used for the CS estimation that does not require this matching. The vertical axes are on a log scale.

Another form of heterogeneity lies in the treatment stringency that a firm experiences from the regulation at different moments in time. As almost all allowances were handed out for free in the first two phases, one could argue that regulation was not stringent in these phases. It was also not uncommon that firms were overallocated with free allowances. In theory, the allocation of allowances should not influence the firm's decision, as allocation does not influence the firm's opportunity costs (Woerdman et al., 2008). However, these allocations are likely to have mattered in practice, as firms in the energy sector, for example, have been found to have made large windfall profits from this allocation (Sijm et al., 2006), and therefore one could interpret an overallocation of emissions as a subsidy rather than a binding regulation.

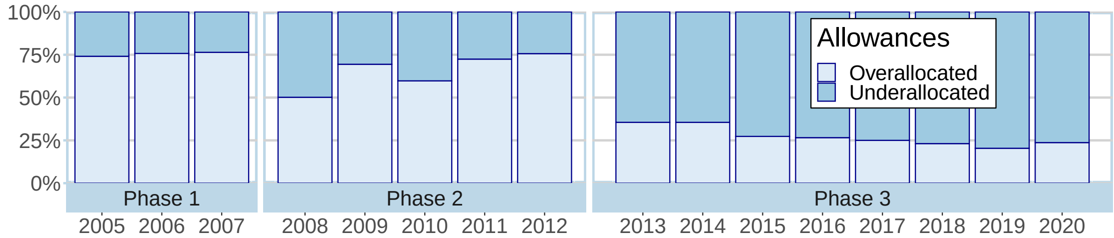


Figure 4: *EU ETS allowance allocations.*

Note: The share of Dutch regulated firms that receive more (or less) allowances for free than their verified emissions. Source: authors calculations based on EUTL data accessed through EUETS.INFO.

In phase 3 allocation mechanisms changed to auctioning as the default option. As many firms were exempted from the switch to auctioning, treatment stringency became more heterogeneous between firms, with some still receiving more allowances than needed, but most now receiving fewer than needed. Figure 4 presents an overview of the share of firms that received more free allowances than they actually emitted by year. One can see the sharp drop in overallocation between phases 2 and 3. Together with changes in prices, shown in Figure 2, this creates significant heterogeneity in policy stringency both over time and between firms.

4 Methodology

4.1 General identification strategy

To identify the effects of the ETS, we use the fact that not all manufacturing firms in the Netherlands are regulated under the ETS. Regulation is on the plant level and there are mainly two criteria for inclusion in the ETS, either through exceeding a certain sector-specific threshold related to energy input or production capacity, or through incorporating specific production processes that imply automatic regulation.[5] This implies that one can attempt to find comparable control firms for each treated firm that are both active in comparable production processes and are comparable in terms of size, employment characteristics and energy input.

In general, two main steps can be identified in any evaluation process that aims at recovering the causal effect of such a regulation, namely (1) matching or weighting, in which one scores firms across treatment status based on their similarity, and (2) comparison, in which one either regresses the outcome variable on treatment status or takes differences in outcome variables across treatment status. The second step utilizes in some way the weights established in the first step.

In this paper, we use a recent estimator developed by Callaway and Sant'Anna (2021) that follows the above approach of weighting and comparison. We use the estimator to obtain causal effects of the EU ETS on all treated firms. We will also use the estimator to obtain cohort and cohort-phase specific treatment effects.

We will argue that the commonly used two-way fixed effects (TWFE) estimator could be biased in this setting, due to the staggered treatment timing of the ETS. We will evaluate the potential severity of the bias and present TWFE results next to the results obtained from the Callaway and Sant'Anna (2021) estimator to keep the comparison to the literature.

4.2 TWFE's problems in staggered DiD settings

In the (environmental) policy evaluation literature it is common to estimate the treatment effect of a regulation by using TWFE, an OLS estimator that controls for time and unit fixed effects. Sometimes a matched TWFE design is adopted, in which the TWFE regression is preceded by sample selection based on matching treated and control units on certain characteristics. The simplest form of such a regression has the following form:

$$y_{jt} = ETS_{jt}\alpha + \gamma_j + \delta_t + \varepsilon_{jt} \tag{4.1}$$

where ETS is a dummy variable that is equal to one if firm j is regulated in year t, and γ_j and δ_t are fixed effects for each firm and year (could alternatively also be a phase fixed effect in our setting). The coefficient of interest is α, which is supposed to capture the aggregated, causal effect of the regulation on the variable of interest y.

[5] For a detailed overview see Annex I of European Parliament, Council of the European Union (2003).

Given that the ETS is divided into several phases, with varying treatment stringency and with firms entering treatment at different phases, the policy design is what is usually referred to as *staggered*. Recent econometric literature has identified several problems with TWFE estimation in staggered settings. This literature focuses on the potential biases in TWFE estimators applied to settings with staggered treatment and potentially heterogeneous treatment effects (see e.g. Daw & Hatfield, 2018; de Chaisemartin & D'Haultfoeuille, 2022). This is exactly the case in our setting, in which firms get treated in different phases. We both expect the different cohorts to react differently to treatment, and we expect the treatment effect to be time-dependent.

The key problem of TWFE in such cases is that the derived estimator for the Average Treatment effect on the Treated (ATT) is a weighted average over the ATTs of the different treatment groups at different times, without explicitly appreciating this and without being able to control the weights of these group-time-specific ATTs. Additionally, the TWFE estimator assigns a positive weight to comparisons that should not reasonably be made. That is, it also includes an ATT that uses already treated units as a control group, which would only be allowed if the treatment effect is completely static.

To assess the severity of this problem in the ETS setting and in our sample, we present a decomposition of the TWFE estimate $\hat{\alpha}$, developed by Goodman-Bacon (2021). For this decomposition, it helps to consider four different groups of firms in our setting, namely the never treated group and the three treated groups, cohorts 1 to 3. As a TWFE regression does not impose rules on which groups can be compared to which other groups, the estimated treatment effect will consist of an implicit weighted average of a difference-in-differences type of comparison between each of the treated groups to all of the other groups.

The estimate overall TWFE effect $\hat{\alpha}$ therefore consists of comparisons between treated and never treated groups, comparisons between treated and not yet treated groups, and comparisons between treated and already treated groups. Especially the latter is problematic, as it makes the "forbidden" comparison, using a control group that would only be suitable if the treatment effect would not be dynamic. In the ETS case this is especially unrealistic, as the severity of the treatment also changes over time and because some adjustment period is likely to exist for firms. For the decomposition, let us denote $\hat{\alpha}_{a,b}$ as the result from comparing group a to group b with the accompanying implicit TWFE weight $s_{a,b}$. Groups consist of the three cohorts (1, 2, 3) and the never treated group (nv). The TWFE estimate can then be decomposed as

$$\hat{\alpha} = \underbrace{\sum_{c\in\{1,2,3\}} s_{c,nv}\hat{\alpha}_{c,nv}}_{\text{Treated to never treated}} + \underbrace{s_{1,2}\hat{\alpha}_{1,2} + s_{1,3}\hat{\alpha}_{1,3} + s_{2,3}\hat{\alpha}_{2,3}}_{\text{Treated to not yet treated}} + \underbrace{s_{2,1}\hat{\alpha}_{2,1} + s_{3,1}\hat{\alpha}_{3,1} + s_{3,2}\hat{\alpha}_{3,2}}_{\text{Treated to already treated}} . \tag{4.2}$$

The weights for each underlying comparison, the ss, in the total TWFE estimate are proportional to (a) the sample size of the cohort serving as treated and the cohort serving as control group, (b) the time span of this subsample as a share on the total sample time, and (c) the identifying variation within the subsample. While the first two parts are rather intuitive, the identifying variation is based on how long the treatment time compared to the non-treated time is within this subsample and how equally sized the two groups are in this subsample. This means that groups that are treated closer to the middle of the sample will get larger weights and that comparisons on equally sized groups will get a relatively larger weight, which is not necessarily intuitive or desired. For more technical details and the precise definition of the weights, we refer to Goodman-Bacon (2021), and especially Theorem 1 therein.

Another shortcoming of the matched-TWFE estimator is that most of the matching information is lost in the regression step. Matching is purely used for sample selection, while the link between matched treated and non-treated units is not taken into account in the estimation. This means that a control firm that is matched to a treated firm in cohort 1 will serve as a control also for treated firms in cohorts 2 and 3, and so on.

We will present results of the estimation of (4.1) alongside our main estimates, to evaluate how problematic the usage of such estimators might be in assessing the ETS's causal treatment effect. In these regressions we allow ε, the error term, to be heteroskedastic and serially correlated and estimate (4.1) using ordinary least squares. In Appendix B.2 we describe the matching and the TWFE regression method in more detail.

4.3 The CS estimator

To address the above mentioned issues, we make use of the estimator developed in Callaway and Sant'Anna (2021) (hereafter *CS*), which was especially developed for staggered treatment settings. It overcomes the problem of the forbidden comparisons by explicitly defining a control group for each treated cohort. One of its main advantages lies in the fact that it estimates ATTs for each treatment cohort and at each year into the treatment. We define a cohort as the group of firms that enters treatment in the same EU ETS phase. It then allows for different aggregations of those estimates, enabling us to restrict the type of heterogeneity in the treatment effect. There have been several estimators designed for such setups in the recent literature. We choose the CS estimator over Borusyak et al. (2021) as it requires a slightly stronger parallel trends assumption to hold and we do not choose De Chaisemartin and d'Haultfoeuille (2020) as our treatment remains constant after its start. For an overview of recent methodologies see Roth et al. (2023).

The estimator is in essence an application of the doubly-robust DiD estimator of Sant'Anna and Zhao (2020) to staggered settings. It pays close attention to the conditioning on covariates, combining both inverse probability weighting (see Abadie, 2005) as well as outcome regression adjustment (see Heckman et al., 1997). The latter is also frequently used in adjusted versions in comparable ETS papers like Martin et al. (2014) or Löschel et al. (2019).

While the inverse probability weighting tries to re-balance the control group based on their probability of being treated, thus in fact on their similarity to the treatment group, the outcome regression adjustment tries to take out covariate-dependent trends in the outcome variable. The CS estimator is therefore consistent as long as the covariate conditioning is correctly specified by either one (or both) of the two covariate conditioning strategies (hence "doubly-robust").

An additional advantage of this weighting is that each treated firm is linked to a specific set of control firms and is only compared to these control firms. This is in contrast with the matched TWFE estimator, where treated groups are compared to the whole set of controls, and in the worst case also to already treated firms.

The estimator for each cohort, c, and year, t, is then a common average treatment effect DiD estimator. It compares the outcome of each firm in year t to the firm's own outcome in the base year, b, and to that of the weighted average difference in outcomes between t and b of the respective control group for this firm. Both inverse probability weighting and the outcome regression adjustment are used.

The following equation specifies the estimated ATT for cohort c and year t:

$$\hat{\alpha}_{ct} = \frac{1}{N}\sum_{j\in\mathcal{J}}\Big[\underbrace{\big(\hat{w}_{jc}^{treated} - \hat{w}_{jc}^{control}\big)}_{\text{Inv. prob. weight.}}\big(y_{jt} - y_{jb} - \underbrace{\hat{m}_{jct}(X_j, \hat{\lambda}_{ct})}_{\text{Outcome reg.}}\big)\Big], \tag{4.3}$$

with N the number of firms and $\mathcal{J}$ the set of all firms, y_{jt} the dependent variable, X as pre-treatment controls, and j, c, t referring to firm, cohort and year. *treated* and *control* refer to the treatment status. The corresponding standard errors are bootstrapped and clustered at the firm level.

$\hat{w}_{jg}^{treated}$ and $\hat{w}_{jg}^{control}$ are the weights that adjust for the probability of being treated. They are 0 if a firm is not in the respective group and give higher weights to control firms that are more similar to the treated firm, given a set of covariates. $\hat{m}_{jct}(X, \hat{\lambda}_{ct})$ represents the bias adjustment from an outcome regression, thus deducting the predicted development of y based on X, under the assumption that the firm had not been treated. More information on both adjustments and their exact definition can be found in Appendix B.

4.3.1 Covariates, anticipation and identifying assumptions

The goal of the matching and weighting that precedes the estimation is to control for all factors that explain the probability of being treated. As mentioned before, the treatment decision is based on crossing certain sector specific energy and capacity thresholds and on relying on certain production processes. We thus base our conditioning on all variables that are related to these factors and try to align it with other studies. As mentioned in Section 3.3, we constrain our sample of control firms to contain only firms in two-digit sectors in which we also observe treated firms, such that we do not keep firms with widely different production processes in the sample. We then additionally incorporate sector fixed effects on a more aggregate level as predictors for the inverse probability weights and the outcome regression. To control for the firm's output capacity, we use a firm's employment, as well as its squared value. For the outcome variables employment and the investment-to-employment ratio, we use turnover and its squared value to avoid multicollinearity issues. To control for the energy intensity, we use energy expenses and its squared value. We additionally include a firm's total wage bill to control for the firm's structure.

It is possible that firms anticipated the regulation and therefore reacted to the policy before the actual start of it. When not controlling for this, this could bias the results, as changes would then be compared to a baseline year, in which the treatment already had an effect. Before phase 1 the important directive for the establishment of an ETS passed in 2003, before phase 2 the national allocation plans had to be published in 2006, and before phase 3 the commission passed directive 2011/540/EU in 2011, extending the scope of regulated greenhouse gases and industries. We therefore assume one year of anticipation, pinning down the base year at two years before the treatment starts.

For each treated cohort, two sets of candidate control firms can be considered, namely (1) the entire population of firms that has not been treated up to t, or (2) only the set of firms that will never be treated. This choice has implications for the assumption on the parallel trends of treated and untreated firms, and one of the big advantages of this estimator is that this choice is completely transparent. Including all not yet treated firms increases the chance of good matches between the control and treated group, especially for cohort 1. However, it does increase the risk of anticipation effects in the control group. Anticipation would violate the parallel trends assumption. We present results for both control groups and discuss the implications in more detail in the findings and discussion.

For either choice, there is no guarantee of parallel trends in absence of the ETS. However, we try to assess whether our treated firms exhibit parallel trends with its control firms in the pre-treatment period. To test for parallel pre-treatment trends, we employ a placebo test. We do this by testing whether pre-treatment ATTs are different from zero. To find these pre-treatment ATTs, we estimate $\hat{\alpha_{ct}}$ for each cohort in all years before its actual treatment, always assuming that the base year is one year before t. As advised in Callaway and Sant'Anna (2021), we then use a Wald test to test the joint statistical significance of these estimates. The test then indicates whether any placebo treatment effects occur between the treated and control groups. In case of significant effects, it would hint at the parallel trends assumption being violated.

Another assumption underlying all DiD approaches is that of stable unit treatment values. This in essence forbids spillovers between regulated and unregulated units. Otherwise the control group would also be affected by the regulation and their outcomes could therefore not be interpreted as the trend of the treated units in absence of the treatment. The fact that we estimate the effects on the firms instead of on the plant level alleviates some, but not all, of these concerns, and we discuss potential implications in detail in Section 6.

4.3.2 Aggregations

We present results for different levels of aggregation that all assume different degrees of heterogeneity in the effect of the ETS. We present an aggregated treatment effect of the ETS on all regulated firms, effects by ETS cohort, and effects by each cohort-phase combination.

These aggregations are based on weighted means of the estimates in (4.3). The aggregation for the total ATT, as well as for the cohort-specific ones, are identical to the ones described in Callaway and Sant'Anna (2021) and we only present the aggregation details by cohort-phase here, which is still based on Callaway and Sant'Anna (2021), but cannot be found in their paper. We trust that the reader can infer the more aggregate aggregations from this exposition.

For our aggregation, the cohort-year specific weights, $v^{\tilde{c},\tilde{p}}(c,t)$ are defined as:

$$v^{\tilde{c},\tilde{p}}(c,t) = P[t|c = \tilde{c} \text{ and } t \in \tilde{p}]\mathbb{1}_{\{c=\tilde{c}\}}\mathbb{1}_{\{t\in\tilde{p}\}}, \tag{4.4}$$

and are thus proportional to the likelihood of being in a given cohort-year and are zero for all $\hat{\alpha}_{ct}$ that are not in the given phase and year. All weights are nonnegative and add up to one within each cohort-phase. These weights are then used in the aggregation:

$$\hat{\theta}^{\tilde{c},\tilde{p}} = \sum_{c\in\{1,2,3\}} \sum_{t=2005}^{2020} \hat{v}^{\tilde{c},\tilde{p}}(c,t)\ \hat{\alpha}_{ct}\ , \tag{4.5}$$

in which θ is the cohort-phase aggregated ATT.

For inference, a bootstrap algorithm calculates the clustered standard error for each estimate $\hat{\theta}^{\tilde{c},\tilde{p}}$. To do this, the algorithm repeatedly draws a subsample of firms from the original sample and estimates the $\hat{\alpha}_{c,t}$s and the respective $\hat{\theta}^{\tilde{c},\tilde{p}}$s. The reported standard error is the standard error of the empirical distribution of $\hat{\theta}^{\tilde{c},\tilde{p}}$ estimates.

5 Findings

5.1 Assessing the bias of the TWFE estimator

To assess the potential bias of TWFE in our DiD setting, we use the decomposition of Goodman-Bacon (2021), as explained in Section 4.2 and equation (4.2).

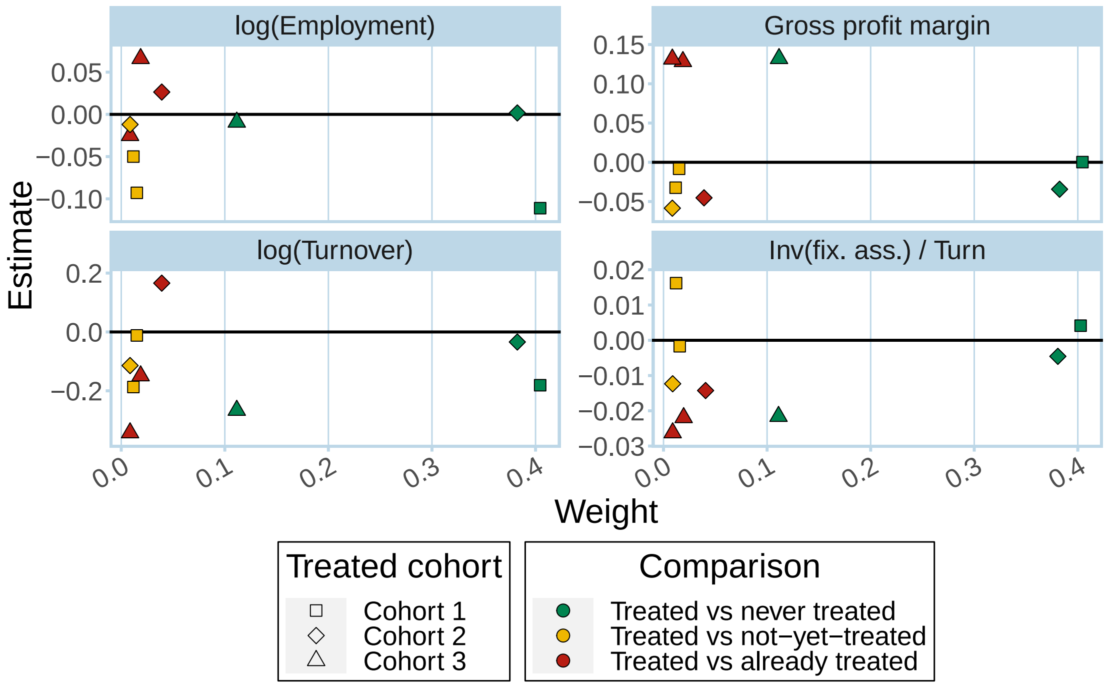


Figure 5: *Goodman-Bacon decomposition for main dependent variables.*

Figure 5 presents all the comparisons and their weights in the TWFE regression, (4.1), using unit and year fixed effects. We can see that in all four estimations, the TWFE estimator assigns non-zero weights to all possible estimates, including the forbidden ones. The desired comparison between treated and never treated firms always receives the highest weight, with the latest treated cohort receiving the lowest weight. This is both because of the shorter treatment time and the smaller sample size of this cohort. Interestingly, the potentially desired comparison between treated and not yet treated firms receives the lowest weights, as they are based on the shortest periods, namely only phases 1 and 2. The forbidden comparisons have weights up to 4 percent for an individual estimate. For turnover and cohort 2, for example, we can see how the forbidden comparison is clearly positive, while the desired one is negative and close to 0. Summing over these two estimates could therefore lead to misleading conclusions.

What also becomes clear is that pooling the different cohorts into one estimate could lead to misleadingly small treatment effect estimates. This is because the different cohorts show clearly different coefficient estimates, ranging from 0 for some and non-zero for others to positive for some and negative for others. In our estimation, we thus pay close attention to differences between cohorts.

5.2 Parallel trends

Before presenting the results of our analysis, we want to provide some evidence on the validity of the parallel trends assumption. Although there is no formal test for parallel trends, one can perform some checks.

Figure 6 plots the ATT estimates for each cohort in each pre-treatment year for our four dependent variables. This allows us to visually inspect if there are substantial pre-treatment trends in the estimated coefficients. This would, for example, be the case if we would observe that treatment effects in a cohort would be continuously significantly different from zero in the pre-treatment years already. The coefficient estimates, however, do not give cause for concern in this regard.

We then use a Wald test on the joint significance of these pre-treatment estimates, as explained in more detail in Section 4.3.1 and as advised in Callaway and Sant'Anna (2021). None of these tests rejects the zero hypothesis of joint statistical insignificance at any conventional significance level. We will always present the test statistic and the p-value of these tests with the respective results. For the TWFE estimation we also check for parallel trends in a placebo test, as explained in detail in Appendix B.2.3.

We add a discussion on the SUTVA assumption in Section 6.

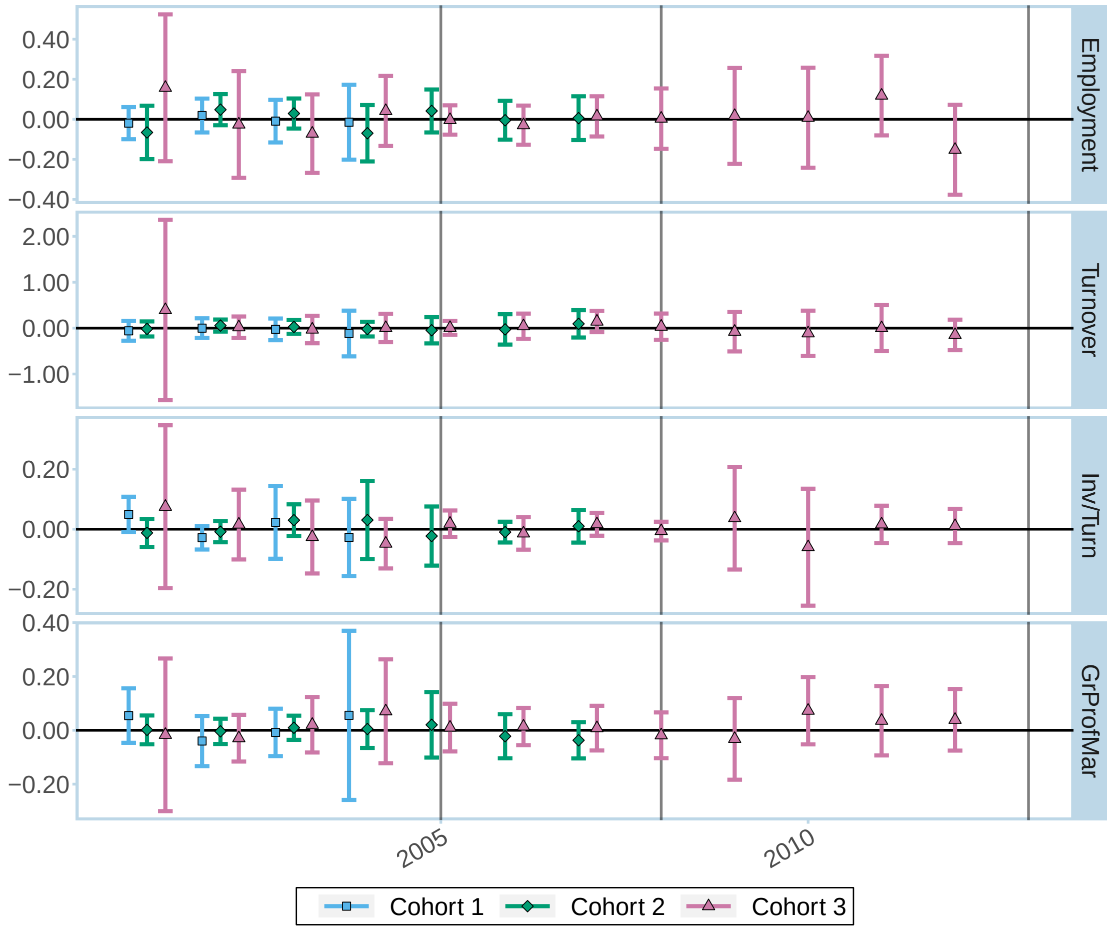


Figure 6: *Pre-treatment estimates to check for parallel trends.*

Note: Non-aggregated coefficient estimates from (4.3) for pre-treatment periods, by using as a baseline year always one year before. All coefficient estimates can be found in Figures D.1 and D.2. Bars represent 95% confidence bars, based on firm clustered standard errors.

5.3 Estimation results for different levels of aggregation

We present results for three different aggregations, each of which assumes a different underlying treatment heterogeneity. We start by presenting an aggregate effect of the ETS on all treated firms, which thus assumes no treatment effect heterogeneity over time and between treated cohorts. We then present results by cohort and finally by cohort-phase. All results are presented for two types of control groups, namely not yet treated and never treated firms. We present TWFE results alongside to determine its performance compared to the unbiased CS estimator.

All results are based on cohort-year-specific ATT estimates, as in (4.3), and aggregated within the desired group as in (4.5). All cohort-year ATT estimates can be found in Figures D.1 and D.2. We present all three aggregations in one table, as they represent the same underlying estimates (in the case of the CS estimator). The results for employment and turnover can be found in Table 2 and those for investment intensity and gross profit margin in Table 3. For TWFE, the statistics at the bottom of the tables refer to the cohort-phase aggregations.

Aggregated estimates, reported in panel A of Table 2 and 3, are mainly small and statistically insignificant. The only exception is the CS coefficient on turnover, which is negative and statistically significant at the 10% level for the CS estimator. This indicates that regulated firms reduced their revenue from product sales compared to the sales from unregulated firms as a result of the regulation. The estimated effect is sizable, as ETS regulation reduces turnover by about 20% on average. We further note that the comparison with the never-treated control group makes the ETS effect larger in magnitude.

The simple aggregate results could hide important heterogeneities, so we proceed with splitting the ETS effect by treatment cohort. The results are presented in panel B. This split does reveal some interesting differences between cohorts, with often opposite signs between cohorts. However, none of these estimates are statistically significant. Looking at the turnover results, we do see that the negative impact of the ETS that we found in the previous aggregation is driven by cohorts 1 and 3, and especially by the larger cohort 1. Cohort 1 sees turnover reduced by 32%, but this is not statistically significant. Cohort 2 firms barely see their turnover affected. Lastly, the TWFE results suggest a negative impact on cohort 1's employment of about 8%, but this is not confirmed by the CS estimates.

We continue by exploring whether the effect of the ETS treatment might have been heterogeneous over time, especially between different treatment phases. The cohort-phase results are presented in panel C. Starting with turnover, we notice that the negative impact that we found above is evenly spread over the three phases. This means that the later phases, with stricter regulation, did not have additional impact on cohort 1 firms, besides a small increase in magnitude between

Table 2: *Regression results for employment and turnover.*

	log(Employment)			log(Turnover)		
	TWFE	CS nyt	CS never	TWFE	CS nyt	CS never
Panel A: Overall aggregation						
Overall ETS	-0.026	-0.002	-0.013	-0.058	-0.203^{*}	-0.235^{*}
	(0.039)	(0.074)	(0.081)	(0.053)	(0.119)	(0.134)
Panel B: Cohort aggregation						
Cohort 1	-0.087^{*}	0.007	-0.009	-0.118	-0.347	-0.393
	(0.052)	(0.124)	(0.151)	(0.079)	(0.197)	(0.229)
Cohort 2	-0.015	-0.028	-0.033	0.006	0.005	-0.012
	(0.060)	(0.044)	(0.046)	(0.088)	(0.123)	(0.123)
Cohort 3	0.098	0.038	0.038	-0.040	-0.150	-0.150
	(0.092)	(0.177)	(0.188)	(0.122)	(0.233)	(0.230)
Panel C: Cohort-Phase aggregation						
Cohort 1 - Phase 1	-0.078^{**}	-0.031	-0.028	-0.023	-0.242	-0.348
	(0.033)	(0.089)	(0.139)	(0.037)	(0.179)	(0.283)
Cohort 1 - Phase 2	-0.088^{*}	0.016	-0.038	-0.126	-0.335	-0.419
	(0.050)	(0.132)	(0.148)	(0.077)	(0.205)	(0.230)
Cohort 1 - Phase 3	-0.092	0.016	0.016	-0.171	-0.394	-0.394
	(0.074)	(0.145)	(0.142)	(0.130)	(0.243)	(0.230)
Cohort 2 - Phase 2	-0.048	-0.073	-0.087	-0.068	0.122	0.077
	(0.058)	(0.046)	(0.048)	(0.081)	(0.123)	(0.103)
Cohort 2 - Phase 3	0.008	0.001	0.001	0.039	-0.068	-0.068
	(0.070)	(0.049)	(0.049)	(0.104)	(0.146)	(0.149)
Cohort 3 - Phase 3	0.101	0.038	0.038	-0.044	-0.150	-0.150
	(0.094)	(0.188)	(0.188)	(0.124)	(0.225)	(0.232)
Observations	6273	11121	11121	6273	11121	11121
Adjusted R2	0.902			0.873		
Placebo test pass	Y			Y		
Wald Stat		25.422	22.991		17.296	22.319
Wald p-value		0.329	0.461		0.794	0.501

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TWFE and CS refer to two-way fixed effects and Callaway & Sant'Anna, respectively, to indicate different regression methods. Note that the table shows three different regressions for TWFE ((4.1), (B.3), (B.4)) and three aggregations based on the same original estimation of year and cohort ATTs for CS, (4.3). The statistics at the bottom of the table are from the cohort-phase aggregation. CS standard errors are bootstrapped and clustered at the firm level both for TWFE and CS estimations. TWFE always includes firm and phase FEs. nyt and nt refer to not yet treated and never treated, respectively, referring to different samples of control firms. The placebo test for TWFE is explained in Appendix B.2.3.

Table 3: *Regression results for investments and profits.*

	Inv(fix. ass.) / Turn			Gross profit margin		
	TWFE	CS nyt	CS never	TWFE	CS nyt	CS never
Panel A: Overall aggregation						
Overall ETS	−0.001	−0.030	−0.033	−0.002	0.022	0.025
	(0.006)	(0.053)	(0.064)	(0.015)	(0.087)	(0.101)
Panel B: Cohort aggregation						
Cohort 1	0.006	−0.056	−0.062	0.001	0.046	0.043
	(0.007)	(0.102)	(0.115)	(0.022)	(0.167)	(0.182)
Cohort 2	−0.002	−0.001	0.001	−0.033	−0.027	−0.013
	(0.008)	(0.013)	(0.015)	(0.025)	(0.038)	(0.037)
Cohort 3	−0.013	0.006	0.006	0.050	0.060	0.060
	(0.014)	(0.027)	(0.029)	(0.038)	(0.085)	(0.092)
Panel C: Cohort-Phase aggregation						
Cohort 1 - Phase 1	0.004	−0.044	−0.050	0.001	0.032	0.002
	(0.011)	(0.064)	(0.101)	(0.010)	(0.122)	(0.182)
Cohort 1 - Phase 2	0.005	−0.076	−0.091	0.007	0.003	0.010
	(0.008)	(0.095)	(0.117)	(0.021)	(0.159)	(0.188)
Cohort 1 - Phase 3	0.007	−0.047	−0.047	−0.004	0.079	0.079
	(0.009)	(0.117)	(0.114)	(0.037)	(0.185)	(0.183)
Cohort 2 - Phase 2	0.009	0.004	0.010	−0.026	−0.068*	−0.031
	(0.013)	(0.026)	(0.027)	(0.022)	(0.033)	(0.034)
Cohort 2 - Phase 3	−0.009	−0.004	−0.004	−0.038	−0.001	−0.001
	(0.012)	(0.014)	(0.013)	(0.031)	(0.043)	(0.045)
Cohort 3 - Phase 3	−0.014	0.006	0.006	0.048	0.060	0.060
	(0.014)	(0.027)	(0.028)	(0.038)	(0.091)	(0.093)
Observations	6207	10945	10945	6273	11121	11121
Adjusted R2	0.110			0.679		
Placebo test pass	Y			Y		
Wald Stat		31.634	23.294		19.096	26.025
Wald p-value		0.108	0.444		0.696	0.300

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TWFE and CS refer to two-way fixed effects and Callaway & Sant'Anna, respectively, to indicate different regression methods. Note that the table shows three different regressions for TWFE ((4.1), (B.3), (B.4)) and three aggregations based on the same original estimation of year and cohort ATTs for CS, (4.3). The statistics at the bottom of the table are from the cohort-phase aggregation. CS standard errors are bootstrapped and clustered at the firm level both for TWFE and CS estimations. TWFE always includes firm and phase FEs. nyt and nt refer to not yet treated and never treated, respectively, referring to different samples of control firms. The placebo test for TWFE is explained in Appendix B.2.3.

phases 1 and 2. Note that these cohort-phase estimates are again not statistically significant. The TWFE estimate for employment suggests a prolonged negative employment effect, although only the effects in phases 1 and 2 are statistically significant. This, however, is again not confirmed by the CS estimator, where the estimates are not statistically different from 0. The cohort 2 and phase 2 coefficient on employment is much larger in magnitude compared to the other estimates, but remains statistically insignificant.

The gross profit margin results are in line with the cohort aggregation results. Only cohort 2 experienced negative effects, and disaggregation shows that phase 2 is driving this effect. Note that only in the comparison with the not yet treated group this effect is statistically significant at the 10% level. Firms' investment intensity seems to be completely unaffected by the ETS.

Overall, the differences between the TWFE and the CS results are not striking. Both estimators yield mostly insignificant coefficients estimates, and if one of them leads to a significant estimate, the estimators mostly qualitatively align. However, we can observe some quantitative differences between the TWFE and the ATT estimates, especially for the employment estimates. Part of this can be explained by the decomposition in Figure 5. For example, the much larger positive coefficient estimate that we find for the employment effect in cohort 3 with TWFE than with CS. The TWFE estimate is based on the aggregation of the correct comparison between cohort 3 and the control group, but also on the comparison of cohort 3 with cohorts 1 and 2 in phase 3. These latter comparisons should not be made, but when looking at the underlying estimates from the decomposition in Figure 5, they might crucially drive the result. While the correct comparison for cohort 3 is negative and tiny, the comparison with cohort 1 is six times larger, in absolute terms, and positive. Even if the weight is much lower, this will nevertheless bias the estimates. For the other differences, these are related to the fact that in the matched TWFE, controls for cohort 3 will also be controls for cohort 1, and the comparison might thus be less accurate.

We also see that controlling for heterogeneity both in terms of treated groups and treatment dynamics is crucial as this uncovers small, but statistically significant effects for cohort 2 right after its treatment start on gross profits. We consider cohort 2 to be the most special group in our sample; as these firms had known that they might become regulated in the future, they are therefore most prone to anticipation effects. In the robustness section we will study cohort 2's results with regard to this anticipation in more detail.

In conclusion, we find little effect of the ETS on the economic outcomes of regulated firms. The only indicators for a reduction in performance are through the two coefficients on turnover and the gross profit margin. However, both are only significant at the 10 percent level and only in specific aggregations. We will use the robustness section to test the robustness of these estimates.

6 Discussion

This section presents additional discussion of the underlying assumptions, and tests the robustness of our results to potential violations. By doing so, we also test how robust our results are to changes in the underlying control group. Further, we present results for slightly different dependent variables, and we discuss our findings in relation to results in the literature.

6.1 SUTVA

The Stable Unit Treatment Value Assumption (SUTVA) is besides the parallel trend assumption the most crucial assumption to hold for identification in a DiD setting. It in essence implies no spillovers between units across treatment status.

The main source of potential spillovers likely comes from within a firm, but between plants. This is because a firm can operate plants of which some are regulated and some are not, and the ETS could thus affect both regulated and unregulated plants, if a firm that operates plants in both groups responds to the regulation by adapting production also in unregulated plants. As our analysis is on the firm level, a large source of these potential spillovers is already accounted for. This, however, also implies that potentially not all activities of the firm are regulated, which could move the estimates towards zero. Note that this is not a drawback per se, as this is simply how the ETS functions. The estimates accurately represent the effect of ETS regulation on firm performance. If we were interested in the question of what would happen if all emissions were regulated, the estimates would be biased toward zero.

On the other hand, there might be positive spillovers between firms through the markets in which they operate. As competition is relative, a reduction in competitiveness of one firm could imply the opposite effect for its competitor. This can both be on the output market, as unregulated firms obtain a relative cost advantage from not being regulated, as well as on the input markets, as potential downsizing of regulated firms allows unregulated firms to obtain employees or input supplies at lower costs. This would inflate the estimates.

As both of these biases to some extent relate to treatment stringency, one can compare the estimates for the different phases in the data. From Figure 2 it becomes clear that later phases are more costly to regulated firms. Figure 1 also shows that more installations of the same owner are regulated in phase 3. Estimates for the later phases should therefore suffer less from the bias towards zero, as more emissions of the firm are regulated, and more from the bias away from zero, as the relative disadvantage from regulation is exacerbated. If these biases exist, in both cases, they should result in larger estimates, in absolute terms, for the same cohorts in later phases. Tables 2 and 3 do not provide evidence of either bias, as estimates for later phases within the same cohort are not consistently further away from zero.

6.2 Strengthening parallel trends

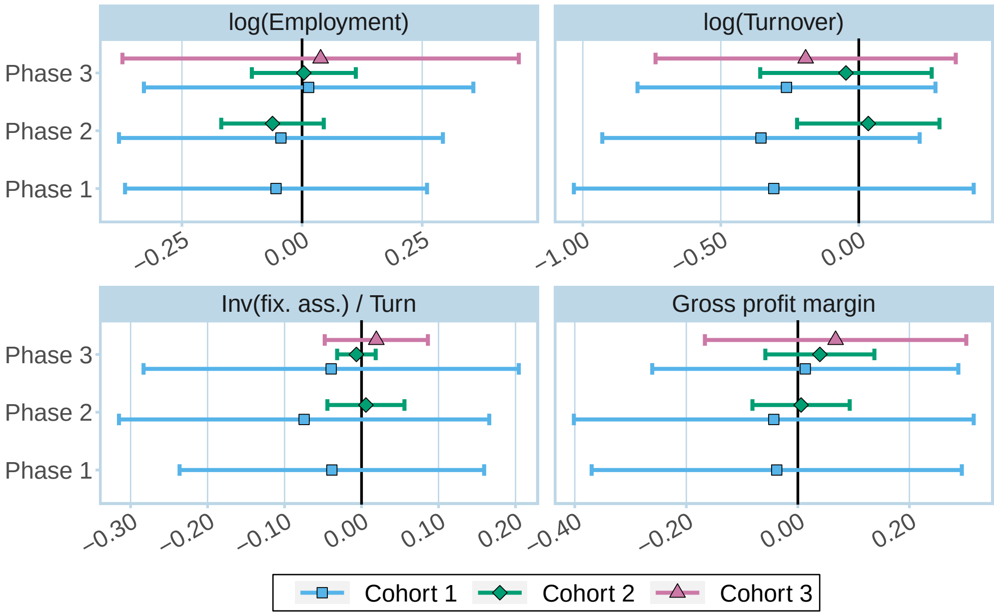


Figure 7: *CS results when matching on pre-treatment trends.*

Note: CS estimates, (4.3), including pre-treatment trends of the dependent variable in the control variables. Aggregated to cohort-phase level and using never treated firms as control group. Results for not yet treated control group can be found in Figure D.5 and disaggregated results in Figure D.4. Bars represent 95% confidence bars, based on firm clustered standard errors.

To strengthen the reader's confidence in the validity of the parallel trend assumption, we add an analysis that aims to enforce at least a pre-treatment similarity in trends between treated and untreated firms. We do this by adding the trend of the dependent variable in the baseline period as a control variable. We calculate this trend in the 4 years before the baseline year for each cohort by regressing, individually for each firm, the dependent variable on a linear time trend. The coefficient on this trend variable is then used in addition to the other control variables in both the outcome regression and the inverse probability weighting.

Figure 7 presents the results for the cohort-phase aggregations using the never treated firms as controls. Disaggregated results and results with a not yet treated control group can be found in Appendix D.

The results are mostly in line with those in Tables 2 and 3. The negative coefficient estimates for turnover remain visible for cohort 1, but they are insignificant at every level of aggregation in this specification.[6] We can again not establish significant effects of the ETS on employment, investments or the gross profit margin. The employment effect that we observed for cohort 2 in phase 2 remains negative, but is still not statistically significant. The negative effect on the gross profit margin for the same cohort phase combination also becomes economically and statistically insignificant.

6.3 Treatment anticipation

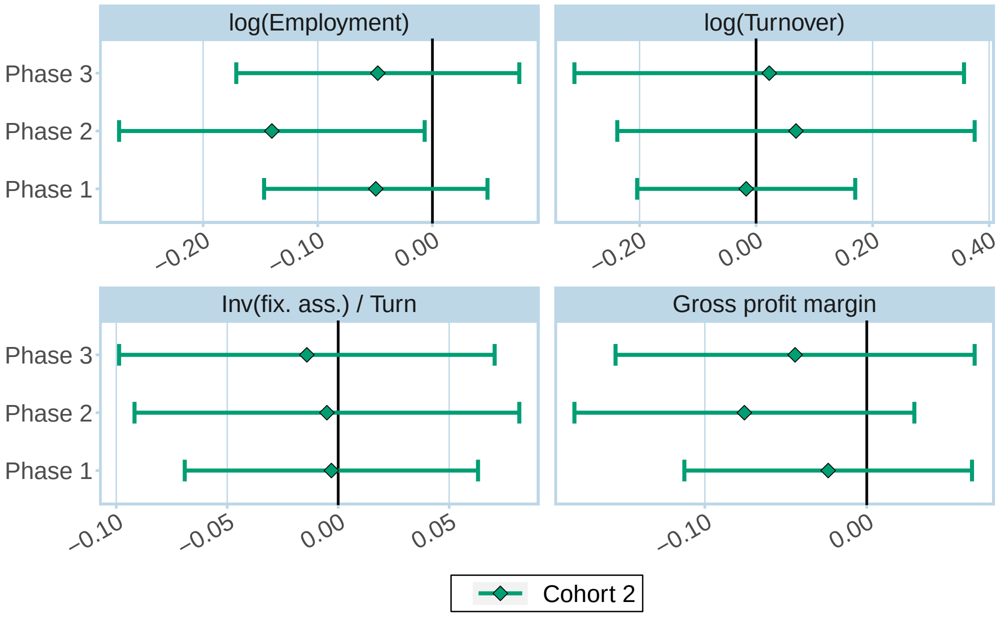


Figure 8: *CS estimates, (4.3), only for cohort 2, when pretending firms in this cohort became treated in phase 1 already. Aggregated to phase level and using never treated firms as control group. Disaggregated results can be found in Figure D.6. Bars represent 95% confidence bars, based on firm clustered standard errors.*

As explained in Section 3, the main reason so many firms enter regulation in the second phase is that the Dutch government excluded many firms from regulation in Phase 1. As these exemptions are public information, it seems likely that these

[6] We do not present all aggregations here, but are happy to provide details on those on request.

firms expected to be regulated in phase 2. If so, the firms in cohort 2 would have already anticipated treatment in 2003, which would violate our assumptions on anticipation, and would also make these firms an improper control for cohort 1 in phase 1.

On the other hand, this policy peculiarity enables us to roughly disentangle an anticipation effect from an actual treatment effect by treating cohort 2 as already being regulated in phase 1. The estimate from such an experiment also provides cohort 2's anticipation effect of being regulated in phase 1, and provides an adjusted estimate of the effect in phases 2 and 3, with an adjusted control group and base year (2003).

The results of this estimation can be found in Figure 8. Most importantly, we do not observe a statistically significant treatment effect in phase 1 for any of the dependent variables, potentially indicating that anticipation effects were small or non-existent. For employment, we still find the negative effect in phase 2, which is now statistically significant at the 5 percent level.

This potential anticipation effect also has important ramifications for the choice of the control group in our setting. If cohort 2 did indeed anticipate the treatment, this would make them ineligible as a control group for cohort 1. The close similarity between results for never treated and not yet treated controls in Tables 2 and 3 also indicates that the choice does not appear to be crucial in our setting.

One could extrapolate and also assume that potentially firms in cohort 3 could have anticipated a future treatment, potentially through insider knowledge in the EU's working, which would weaken the credibility of the not yet treated results for cohort 2. Table 3 showed a negative treatment effect on cohort 2's phase 2's gross profit margins when using not yet treated firms as the control group. When using the never treated firms as control group, the effect halves and becomes statistically insignificant. In case of anticipation effects in cohort 3, the latter estimate would be more appropriate.

6.4 Fully balanced panel

To test the sensitivity of our results to firms entering and exiting our sample, we redo the estimation on a fully balanced panel. This greatly reduces the number of observations and the number of firms in our analysis. This also prevents us from reporting cohort 3 findings due to privacy requirements from CBS.

The results are presented in Figure 9. Potentially due to the lower sample sizes, standard errors are larger and we confirm that we cannot find any statistically significant effects at this level of aggregation.

The negative employment effect for cohort 2 disappears completely, while the negative effect on turnover for cohort 1 remains stable, and at the most aggregate

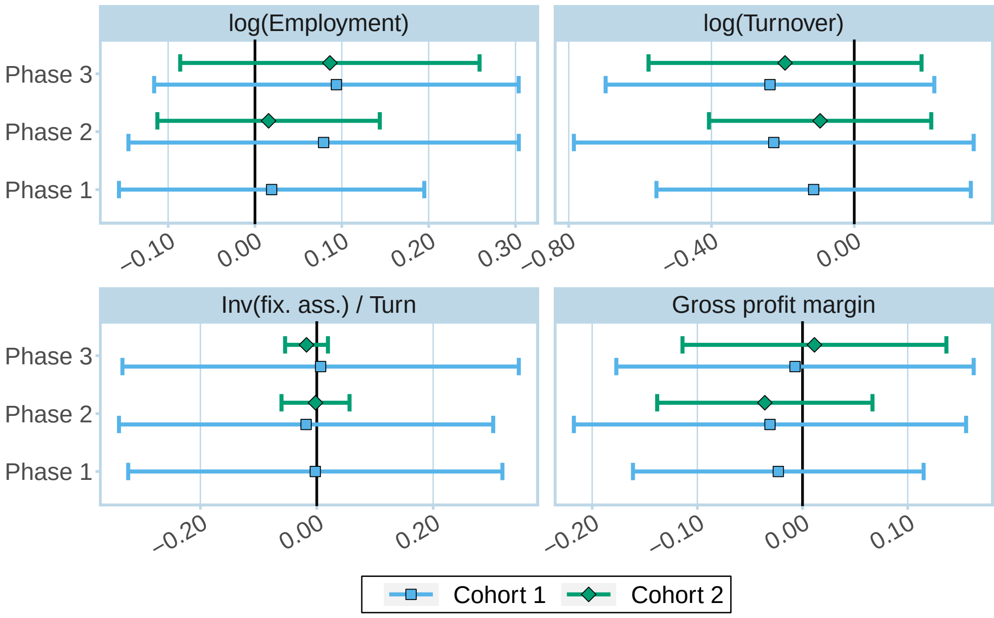


Figure 9: *CS results from a balanced sample.*

Note: CS estimates, (4.3), using a fully balanced panel. Aggregated to cohort-phase level and using never treated firms as control group. Results for not yet treated control group can be found in Figure D.8 and disaggregated results in Figure D.7. Bars represent 95% confidence bars, based on firm clustered standard errors.

aggregation is again significant at the 10% level.[7]

6.5 Alternative measures as dependent variables

To test if our results rely on our choice of the dependent variable, we redo the estimation with alternative dependent variables. We use the operating instead of the gross profit margin and scale investments (into all fixed assets) by the number of employees in full-time equivalents. We also present results for investments into machines only instead of investments into all fixed assets, scaled by turnover.

The results for the economic performance variables are presented in Table D.1. The effect on the operating margin is statistically insignificant. The effect on cohort 2 in phase 2 remains negative but becomes statistically insignificant in our preferred CS estimation.

The scaling of investments into total assets and the choice of a different investment category do not seem to matter, as the investment ratio results remain insignificant. The coefficient estimates are reported in Table D.2.

6.6 Comparison to literature

Most articles that study the ETS's effect find few side effects from the regulation. In that sense our study's overall findings are not too different from the literature's findings. We do not find conclusive indications of a reduction in profit margins and the findings of negative employment effects seem too uncertain to be a cause for concern. However, we do find some evidence that cohort 1 reduced its turnover as a result of the ETS. We have shown that cohort 1 mostly consists of the most energy-intensive firms, which is likely linked to this result, as for example Joltreau and Sommerfeld (2019) hint at low energy-intensity being one of the reasons for why other studies have not found considerably side effects.

The reason for the negative effects could be based in the high export-orientation of the Dutch economy. Exposure to trade could have incentivized regulated firms to downscale their operations within the Netherlands as a response to the regulation, even before the regulation became more stringent. Such anticipation of future strictness of the regulation would also be in line with the potential anticipation effect that we discuss for cohort 2 firms. By disentangling the effects for the very energy-intensive firms in cohort 1 from the later, less energy-intensive cohorts, we show that heterogeneity between the cohorts plays an important role. Not all firms respond the same to the EU ETS. It is, however, important to stress that this effect is statistically uncertain and does not survive all robustness checks.

[7] We only present the cohort-phase aggregation here, but are happy to provide details on other aggregations on request.

Even if we do not find significant effects on investments, the direction of the coefficients for cohort 1 are in line with the effects on turnover and hint at a downsizing of the domestic operations. We therefore cannot support the finding of other studies that find an increase in investments due to the ETS regulation.

Like Löschel et al. (2019) we also find statistically significant effects in the semi-parametric DiD estimation rather than in the TWFE regression, even though our results differ. This again highlights the importance of choosing the right estimation methods when studying the ETS's effects.

7 Conclusion

This paper studies the effects of the EU ETS on the economic performance and investment behavior of regulated manufacturing firms in the Netherlands. Motivated by large differences in energy intensity between firms that became regulated at different times and by an increase in regulatory stringency over time, we pay special attention to treatment effect heterogeneity. We use a decomposition developed by Goodman-Bacon (2021) to show the potential problems of using a classical TWFE estimator when estimating the treatment effect of the ETS. To better allow for heterogeneity and avoid the bias of TWFE in staggered treatment settings, we employ a recent semiparametric DiD estimator introduced by Callaway and Sant'Anna (2021).

We estimate the effects of EU ETS regulation on employment, turnover, the gross profit margin, and investment intensity. To capture the heterogeneities we estimate the effect of phases 1-3 in the EU ETS for each cohort, whereby a firm belongs to one of the three cohorts if it was first regulated in that respective phase. We make use of public data from the European Union Transaction Log and restricted-access microdata from Statistics Netherlands.

In the decomposition exercise, we show that a TWFE estimator implicitly puts a positive weight on comparisons between firms that are both already treated, which is undesirable and could bias results. The decomposition also highlights the importance of heterogeneity between cohorts, as the aggregation consists of the addition of both positive and negative cohort-specific treatment effects.

We find no statistically significant effects on employment and profits. If anything, the results point towards a small reduction in profits for cohort 2, but this effect is not robust across specifications. This implies that even when controlling for the staggered design of the ETS and when accounting for periods with higher carbon prices and fewer free allowance allocations, there seems to be no significant effect of the regulation on employment and profits.

We find an overall negative, but statistically uncertain, treatment effect of the ETS on the turnover of regulated firms, which implies a reduction in the revenue from industrial product sales of around 20%. This effect is driven by a negative effect on firms regulated in cohort 1, for which the effect is visible throughout all three phases, although these disaggregated estimates are not statistically significant. Cohort 1 firms are by far the most energy-intensive in our sample, which makes it likely that they are the most affected by the regulation. We cannot establish any significant effects of the regulation on firms' investment behavior. This is independent of the type of investment that we study.

We thoroughly test the underlying assumptions of our results. Common trends are tested for with pre-treatment placebo tests. Matching on trends and enforcing a balanced sample, do not change the main conclusions. We also discuss the potential violations of the SUTVA assumption and conclude that spillovers play a small role, but that anticipation to treatment might exist. In case of anticipation, the CS estimates based on never treated firms as a control group are preferred.

Our results fit into the literature in two ways. First, we show that there exists large heterogeneity of regulatory stringency over time and that cohorts of firms regulated in different phases have different characteristics. This highlights the importance of the right DiD design and estimator, as heavily discussed in the recent econometric literature. The most important contribution is how the counterfactual is composed and constructed. Second, our findings add to the debate on negative and positive side-effects of environmental regulation. Using data up to the end of phase 3 (2020) allows studying heterogeneity over time. We conclude that most worries over losses in firm performance are not warranted. If anything, there could be some reduction in revenues for the most energy-intensive firms, but these findings are rather uncertain. Profits and investments seem unaffected by the regulation.

Research on the EU ETS's effects remains of interest, as longer time series allow for the evaluation of medium- and long-term effects. This can be informative to policy makers considering the implementation or strengthening of environmental policy. Future research will also allow for the analysis of changes in regulatory stringency, which we have already exploited to some extent here. Analysis of phase 4 reforms and the high EUA prices as of 2021 might provide new insights.

References

Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*, *72*(1), 1–19.

Abrell, J. (2021). Database for the European Union Transaction Log [Database, available at https://euets.info].

Borghesi, S., Cainelli, G., & Mazzanti, M. (2015). Linking emission trading to environmental innovation: Evidence from the Italian manufacturing industry. *Research Policy*, *44*(3), 669–683.

Borusyak, K., Jaravel, X., & Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Calel, R., & Dechezleprêtre, A. (2016). Environmental policy and directed technological change: Evidence from the European carbon market. *Review of Economics and Statistics*, *98*(1), 173–191.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-Differences with multiple time periods [Themed Issue: Treatment Effect 1]. *Journal of Econometrics*, *225*(2), 200–230.

Colmer, J., Martin, R., Muûls, M., & Wagner, U. J. (2022). *Does pricing carbon mitigate climate change? Firm-level evidence from the European Union Emissions Trading Scheme* (Discussion Paper No. DP16982). CEPR.

Daw, J. R., & Hatfield, L. A. (2018). Matching and regression to the mean in difference-in-differences analysis. *Health Services Research*, *53*(6), 4138–4156.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–2996.

de Chaisemartin, C., & D'Haultfoeuille, X. (2022). *Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey* (Working Paper No. 29691). National Bureau of Economic Research.

Dechezleprêtre, A., Gennaioli, C., Martin, R., Muûls, M., & Stoerk, T. (2022). Searching for carbon leaks in multinational companies. *Journal of Environmental Economics and Management*, *112*, 102601.

Dechezleprêtre, A., Nachtigall, D., & Venmans, F. (2023). The joint impact of the European Union Emissions Trading System on carbon emissions and economic performance. *Journal of Environmental Economics and Management*, *118*, 102758.

Ellerman, A. D., & Buchner, B. K. (2008). Over-allocation or abatement? A preliminary analysis of the EU ETS based on the 2005–06 emissions data. *Environmental and Resource Economics*, *41*(2), 267–287.

European Commission. (2004). Commission decision of 29 October 2004 concerning the temporary exclusion of certain installations by the Netherlands from the Community emissions trading scheme pursuant to Article 27 of Directive 2003/87/EC of the European Parliament and of the Council [Decision C(2004)4240-3].

European Commission. (2005). Commission decision of 22 march 2005 concerning the temporary exclusion of certain installations by the netherlands from the community emissions trading scheme pursuant to article 27 of directive 2003/87/ec of the european parliament and of the council [Decision C(2005) 866 final].

European Parliament, Council of the European Union. (2003). Establishing a scheme for greenhouse gas emission allowance trading within the community [Directive 2003/87/EC].

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing [Themed Issue: Treatment Effect 1]. *Journal of Econometrics*, *225*(2), 254–277.

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

Hintermann, B. (2010). Allowance price drivers in the first phase of the EU ETS. *Journal of Environmental Economics and Management*, *59*(1), 43–56.

Jaraite-Kažukauske, J., & Di Maria, C. (2016). Did the EU ETS make a difference? An empirical assessment using Lithuanian firm-level data. *The Energy Journal*, *37*(1).

Joltreau, E., & Sommerfeld, K. (2019). Why does emissions trading under the EU Emissions Trading System (ETS) not affect firms' competitiveness? Empirical findings from the literature. *Climate Policy*, *19*(4), 453–471.

Klemetsen, M., Rosendahl, K. E., & Jakobsen, A. L. (2020). The impacts of the EU ETS on Norwegian plants' environmental and economic performance. *Climate Change Economics*, *11*(01).

Koch, N., Fuss, S., Grosjean, G., & Edenhofer, O. (2014). Causes of the EU ETS price drop: Recession, CDM, renewable policies or a bit of everything?—New evidence. *Energy Policy*, *73*, 676–685.

Löschel, A., Lutz, B. J., & Managi, S. (2019). The impacts of the EU ETS on efficiency and economic performance – An empirical analyses for German manufacturing firms. *Resource and Energy Economics*, *56*, 71–95.

Marin, G., Marino, M., & Pellegrin, C. (2018). The impact of the European Emission Trading Scheme on multiple measures of economic performance. *Environmental and Resource Economics*, *71*(2), 551–582.

Martin, R., Muûls, M., De Preux, L. B., & Wagner, U. J. (2014). Industry compensation under relocation risk: A firm-level analysis of the EU Emissions Trading System. *American Economic Review*, *104*(8), 2482–2508.

Narassimhan, E., Gallagher, K. S., Koester, S., & Alejo, J. R. (2018). Carbon pricing in practice: A review of existing emissions trading systems. *Climate Policy*, *18*(8), 967–991.

Petrick, S., & Wagner, U. J. (2014). *The impact of carbon trading on industry: Evidence from German manufacturing firms* (Working paper No. 2389800). Available at SSRN.

Pindyck, R. S. (2019). The social cost of carbon revisited. *Journal of Environmental Economics and Management*, *94*, 140–160.

Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*.

Sant'Anna, P. H., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101–122.

Sijm, J., Neuhoff, K., & Chen, Y. (2006). CO2 cost pass-through and windfall profits in the power sector. *Climate Policy*, *6*(1), 49–72.

Teixidó, J., Verde, S. F., & Nicolli, F. (2019). The impact of the EU Emissions Trading System on low-carbon technological change: The empirical evidence. *Ecological Economics*, *164*(Article Number: 106347).

Venmans, F., Ellis, J., & Nachtigall, D. (2020). Carbon pricing and competitiveness: Are they at odds? *Climate Policy*, *20*(9), 1070–1091.

Verde, S. F. (2020). The impact of the EU Emissions Trading System on competitiveness and carbon leakage: The econometric evidence. *Journal of Economic Surveys*, *34*(2), 320–343.

Vona, F. (2019). Job losses and political acceptability of climate policies: Why the 'job-killing' argument is so persistent and how to overturn it. *Climate Policy*, *19*(4), 524–532.

Woerdman, E., Arcuri, A., & Clò, S. (2008). Emissions trading and the polluter-pays principle: Do polluters pay under grandfathering? *Review of Law & Economics*, *4*(2), 565–590.

A Data details

A.1 Firms within Statistics Netherlands (CBS)

We refer in this paper to firms as a collection of chamber of commerce numbers. These units in the CBS data are partially constructed by CBS itself. Especially the Business Unit (BEID), which we use as firm in our study, is a construct that is generated by CBS. Here we will discuss how these units are constructed.

A.1.1 Business Unit (BEID)

The business unit (BEID) captures outward-facing (i.e. non-internal) Dutch production or service-provision that can be seen as one unit. This means that legal firm structures are grouped by purpose into BEIDs, e.g. a unit producing wooden furniture. This provides several advantages and disadvantages. The main advantage is that the BE is a unit structure that captures economic activity well. Legal firm structures often only exist for fiscal reasons and do not represent economic activity or choices well. The disadvantage is that BEIDs are constructed and that their composition can change over time, even though these changes might be representative of economic activity within the BE.

A.2 EU ETS

For the data on the EU ETS, coming from EUETS.INFO, a few transformations are needed.

The main problem occurs when installations change owner. This event is poorly captured by the data and therefore requires manual corrections. The corrections of ownership change were done in the following steps.

1. From the European Commission's Union Registry the lists of (stationary) installations for each phase are downloaded.[8]

2. The owners of each installation are compared across phases. If the owners are unchanged between phases, they are assumed to have been the same within that phase.

3. For the installations of which owners have changed between phases, we search the internet for further information to determine whether there was a transfer of ownership and between whom. From sources like news articles or websites that provide ownership data, we deduce when ownership has changed and to who. Two common situations occur, namely (1) ownership of installations

[8]These lists can be found for Phase 1, 2 and 3 on the EC's website.

is transferred within a firm group, which effectively means the installation has the same ultimate owner and (2) another firm purchases the installation, sometimes because the previous owner went bankrupt.

4. For installations that saw their owner change but for which we find no information when this took place, we assumed the change to take place on the day the new phase started.

The dates of ownership change then have to be reconciled with the annual data. For this, the year was chosen in which the ownership change has taken place and this year is considered to be the year in which the new owner takes economic responsibility of the installation.

A.3 Details on merging the EUTL with CBS data

Data that is imported into the CBS environment and that is identified on the chamber of commerce (in Dutch: KvK) number, like the ETS data, is encrypted on the same level. So installations under the EU ETS are imported into the CBS environment and encrypted. Encrypted chamber of commerce numbers can then be used to link EU ETS regulation to the business units.

Based on this encryption, one can find the corresponding CBS person (Dutch: persoon) in each year. This CBS person presents a layer in between the detailed KvK number and the final identifier level, business units (BEs). The CBS persoon itself is just a one to one linking from the KvK number to a CBS internal identifier. In some rare years a KvK number is assigned to two CBS persons within a year. This is because CBS draws from multiple sources which can cause duplicate links. In these cases, we have decided to assign the KvK number to the later created CBS person within that year.

The original ETS plant is thus assigned to a BEID in each year, ownership changes between years are thus uncritically represented here. However, in some years a CBS person is assigned to two BEIDs, which can happen if ownership changes within a year. In these cases, we assign the later BEID to the plant.

The CBS data sets are all identified on the BEID level and so we can in the next step merge the ETS plants to the CBS data sets. In each of these steps some of the companies cannot be assigned to another identifier or data set, such that in the end not all ETS firms can be merged. There is, however, no systemic bias in this. After consultation with CBS, the majority of the firms that we were no able to link stem from site that has merged several ETS installations under one account holder, which are then impossible to link to the BEID in our data.

B Technicalities of estimation strategy

B.1 Further explanation and definitions of the cohort-year specific ATT

We here give the definitions of the inverse probability and outcome regression adjustments as well as their underlying interpretation. We only provide the basic definitions, as they are equivalent to the explanations in Callaway and Sant'Anna, 2021. The propensity score weights used in (4.3) are:

$$\hat{w}_{jc}^{treated} = \frac{G_{jc}}{\frac{1}{N}\sum_i G_{jc}} \text{ and} \tag{B.1}$$

$$\hat{w}_{jc}^{control} = C_{jc}\frac{\frac{p_{jc}(X_j,\hat{\pi}_c)}{1-p_{jc}(X_j,\hat{\pi}_c)}}{\frac{1}{N}\sum_i \frac{p_{jc}(X_j,\hat{\pi}_c)}{1-p_{jc}(X_j,\hat{\pi}_c)}}\ , \tag{B.2}$$

where G_{jc} is a dummy indicating if a firm is in the respective treatment group or not, and C_{jc} is a dummy that is one if the firm can serve as a control for that treatment cohort, here incorporating never treated as well as not yet treated firms. The never-treated case can be found in Callaway and Sant'Anna, 2021. p_{jc} is the estimated propensity score for each firm (giving the probability of being in that treatment cohort), based on the controls and the estimated coefficients $\hat{\pi}_c$ from a logistic regression model. The procedure thus weights controls that are more likely to be treated higher than firms that are unlikely to be treated.

$\hat{m}_{jc}(X_j, \hat{\lambda}_{ct})$ in (4.3) is the estimator of $\mathbb{E}[Y_t - Y_{base}|X, C = 1]$. It is thus the difference in predicted values between year t and the base year for the treated firms, if they were untreated. One thus runs $y_{jt} - y_{jb} = \lambda X_j + \varepsilon_j$ only on the sample of the untreated units, to estimate the change in outcomes that can be predicted by the covariates and then uses this $\hat{\lambda}$ to predict $\hat{m}_{jt}(X_j, \hat{\lambda}_{tc}) = \widehat{y_{jt} - y_{jb}}$, in this case both for the treated and untreated units.

B.2 TWFE estimation details

Even though the TWFE results do not present our main results, we take great care in first matching the treated firms to a reasonable control group, controling for parallel trends, and then estimating the effects for similar aggregations, as in the CS estimation. We break the matching and regression up in the following two subsections. The first one explains the matching that provides the weights, and the second one presents the details of the regression. In the third subsection we then discuss placebo tests that we use to test for the parallel trends assumption. As the matching is a rather complex procedure, we provide all the matching details separately in Appendix C, together with descriptive statistics on the matching.

B.2.1 Matching

The goal of matching is to select similar observations across treatment status from the data. In general a matching algorithm provides a similarity score between each pair of observations in the sample data. If provided with n observations, the matching outcome matrix M has dimensions $n \times n$. For our TWFE application the pair information is dropped, and only those observations with a high enough similarity score to any other observation across treatment status are kept, collapsing the matching information from M to a binary vector of length n, indicating for each firm if it will be kept in the estimation or not. Observations in the non-treated group that do not have a high enough similarity score with a treated observation are thus dropped from our sample. This way matching boils down to sample selection.

The matching outcomes are used to select the sample for our TWFE regression. All observations are kept for firms that are matched, both in the treatment group and the control group (i.e. have a value of 1 in the n-length vector). This effectively is a special form of weighting, as the weights are either 1 (for the matched) or 0 (for the non-matched).

We only match within the two-digit industry code, and base the similarity on a firm's employment, energy costs, turnover and total wage bill, as well as on the squared values of these variables. Matching happens two years before treatment starts, to account for anticipation. Our matching algorithm is further elaborated in Algorithm 1 in Appendix C.

B.2.2 TWFE estimation

Using the resulting matched sample, we can estimate the impact of the EU ETS's phases on each cohort's outcomes. Our main aggregated TWFE regression is presented in (4.1), and the aggregations by cohort and by cohort-phase are presented here.

The regression for the cohort aggregation is:

$$y_{jt} = \sum_{c \in C} ETS_j^c \alpha^c + \gamma_j + \delta_t + \varepsilon_{jt} \ , \tag{B.3}$$

and the one for the cohort-phase aggregations is:

$$y_{jt} = \sum_{c \in C} \sum_{p \in P} ETS_j^c \times P_t^p \alpha^{cp} + \gamma_j + \delta_t + \varepsilon_{jt} \ , \tag{B.4}$$

where the subscripts j, t refer to the firm and year. ETS^c is a dummy variable that is equal to one if firm j is in cohort c from the moment of treatment onwards. P is a dummy equal to one if year t is in ETS phase p. As there are three phases in our data range, we have $C, P \in \{1, 2, 3\}$. The interactions of the two variables present the treatment indicators of our DiD regression. The coefficients of interest are the corresponding αs, with one coefficient for each cohort or for each of the six post-treatment cohort-phase combinations (i.e. cohort 1-phase 1 through cohort 3-phase 3).

For all three regressions, we include firm and phase fixed effects, but abstain from including time-varying controls, as these are likely to be affected by the treatment itself. The error term is allowed to be heteroskedastic and serially correlated. We estimate each model using ordinary least squares (OLS).

B.2.3 Parallel trends in TWFE

To test whether the parallel trends assumption holds for the different cohorts, we devise a placebo test. The placebo test introduces a non-existing treatment in the pre-treatment period and tests whether it has a significant effect on the outcome variables. If so, the treated group's trends deviates from that of the control group.

These tests can be used for the cohort-aggregated regressions from (B.3). It cannot be used for the cohort-phase regressions of (B.4), because the placebo treatment cannot overlap with any actual treatments. For example, introducing a placebo treatment for cohort 2 phase 3 in the year 2011 is flawed, because during that time cohort 2 is actually treated under the ETS. It would then be impossible to disentangle actual treatment effects from placebo treatment effects. Placebo tests have to strictly stick to pre-treatment periods.

The placebo results are presented in Table B.1 and B.2 for all seven dependent variables in this study. We notice that only the operating margin fails the placebo test for cohort 3, as there is a statistically significant effect from the placebo treatment on the operating margin. Further, we note that both the operating margin and employment see a marginally statistically significant effect for cohort 1, indicating there might be some concerns over parallel pre-trends for these two variables in the matched TWFE setting.

C Matching details

Our matching algorithm for the TWFE estimation is presented in Algorithm 1. The algorithm is designed to match treated firms to similar enough control firms in order to make a sensible comparison between their economic outcomes. It also attempts to filter for good data quality, e.g. by only considering firms that are observed for several consecutive years around treatment.

Algorithm 1: *Matching*

Table B.1: *Tests for placebo treatments at $T-2$.*

	Employment	Turnover	Inv/Turn	GrProfMar
Cohort 1	-0.048^{*}	-0.012	0.007	-0.007
	(0.025)	(0.030)	(0.009)	(0.009)
Cohort 2	-0.020	-0.016	-0.002	-0.011
	(0.058)	(0.064)	(0.010)	(0.014)
Cohort 3	0.099	0.064	0.001	0.061
	(0.065)	(0.115)	(0.011)	(0.045)
Observations	4994	4994	4942	4994
Adjusted R2	0.885	0.873	0.129	0.723
Phase FEs	4	4	4	4
Firm FEs	353	353	353	353

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TWFE regressions with dummies for each cohort using the matched sample. The coefficients estimate the treatment effect from a placebo treatment that takes place 2 years before the actual ETS treatment. Standard errors are clustered at the firm level.

Table B.2: *Tests for placebo treatments at $T-2$.*

	OpMar	Inv/Empl	InvMach/Turn
Cohort 1	-0.015^{*}	3.252	0.007
	(0.008)	(3.734)	(0.008)
Cohort 2	-0.009	-2.799	-0.014
	(0.013)	(3.572)	(0.009)
Cohort 3	0.040^{**}	2.137	-0.013
	(0.017)	(8.409)	(0.013)
Observations	4866	4934	4915
Adjusted R2	0.289	0.218	-0.015
Phase FEs	4	4	4
Firm FEs	353	353	353

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TWFE regressions with dummies for each cohort using the matched sample. The coefficients estimate the treatment effect from a placebo treatment that takes place 2 years before the actual ETS treatment. Standard errors are clustered at the firm level.

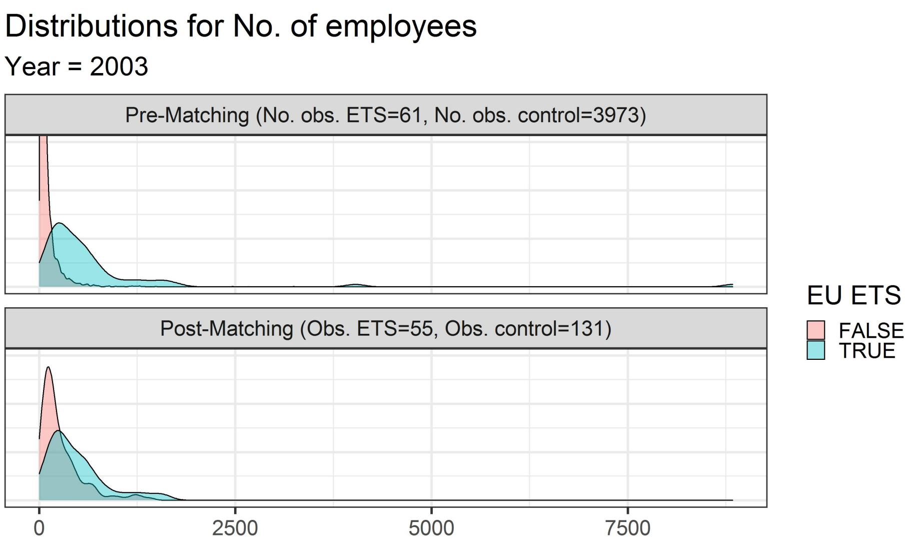
1. Enforce common support between treated and control units
   (a) For each baseline year, we drop all observations that are outside the common support of the treated and control group.
2. Select treatment period
   (a) Take treatment period $T \in T^p$, where T^p is the set of treatment periods, i.e. the years 2005, 2008 and 2013 for phase 1, phase 2 and phase 3 (p) in the EU ETS respectively.
3. Select observations to be potentially matched
   (a) From the ever-treated EU ETS firms, select only those observations that are first regulated in phase p Keep all observations from the never-treated group.
   (b) Only keep units that are observed for all of the years in $(T - pre, T + post)$, where we set $pre = 2$ and $post = 3$. This guarantees that resulting matches can be observed around the treatment period.
   (c) Select only the observations at $T - pre$, dropping the panel structure. This year will be the pre-treatment matching period.
4. Similarity scoring and match decision
   (a) Measure the Mahalanobis distance between all observations in the selected sample across treatment status for the variables X^m.[9] X^m are the matching variables for which we take the number of employees, turnover, wage expenses, energy expenses, and value added and their squared values. We also restrict matches to be only within a 2-digit sector code. Matches across sectors are not allowed.
   (b) For each treated unit collect the H closest neighbors based on the Mahalanobis distance. We opt for $H = 5$ and we do allow for replacement. We also allow for ties, meaning ties are not randomly broken but rather all are included in the result. For the implementation of this step and the previous step we leverage on the `Matching` package's `Match` function in R.

[9] The Mahalanobis distance between treated (T) unit A's covariate vector x_A and control (C) unit B's covariate vector x_B is given by $d(A, B) = \sqrt{(x_A^T - \mu^T)S^{-1}(x_B^C - \mu^C)}$, where S is the variance-covariance matrix between x^T and x^C and where the μs are the means of their respective series. Note that this distance measure is like a variance-corrected normalized Euclidean distance.

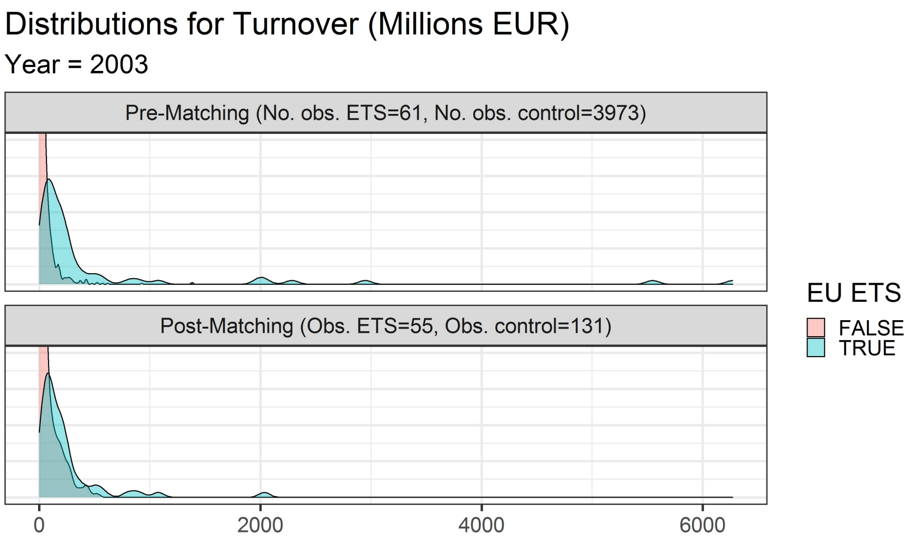
5. Store matching outcome
    (a) Remaining matches are stored under matching year $T - pre$.
6. Next treatment period
    (a) If not all treatment periods in T^p are covered yet, select the next value in T^p and repeat the algorithm from step 2.

Figure C.1, Figure C.2 and Figure C.3 show the distributions of selected variable for regulated versus non-regulated firms before and after the matching procedure for the pre-phase 1 year 2003, pre-phase 2 year 2006 and the pre-phase 3 year 2011 respectively.

C.1 Matching outcomes

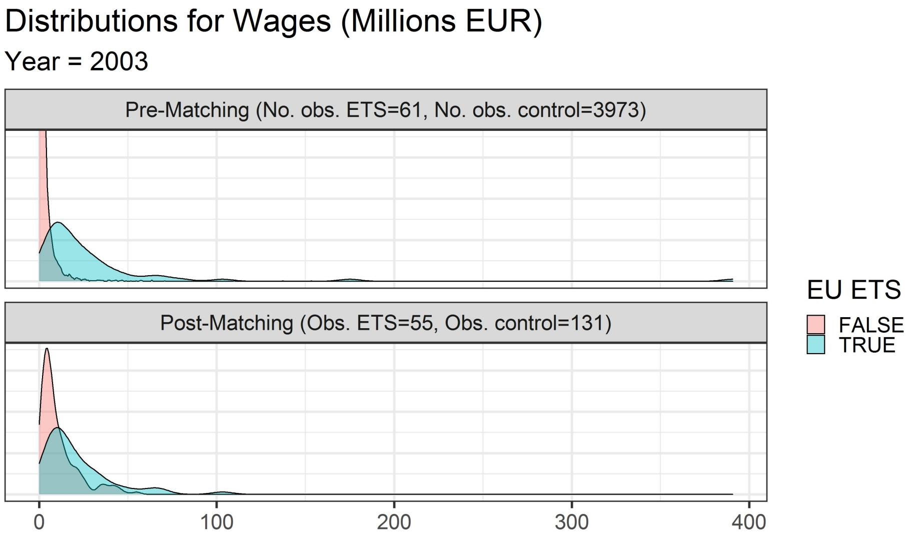


(a) *Number of employees*

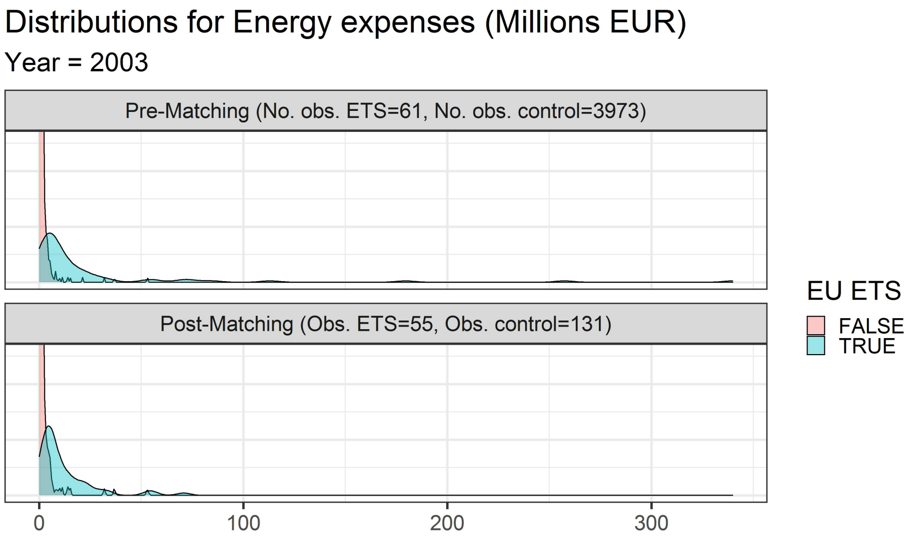


(b) *Turnover*

Figure C.1: *Distributions of variables before and after matching for treated and control firms in 2003.*

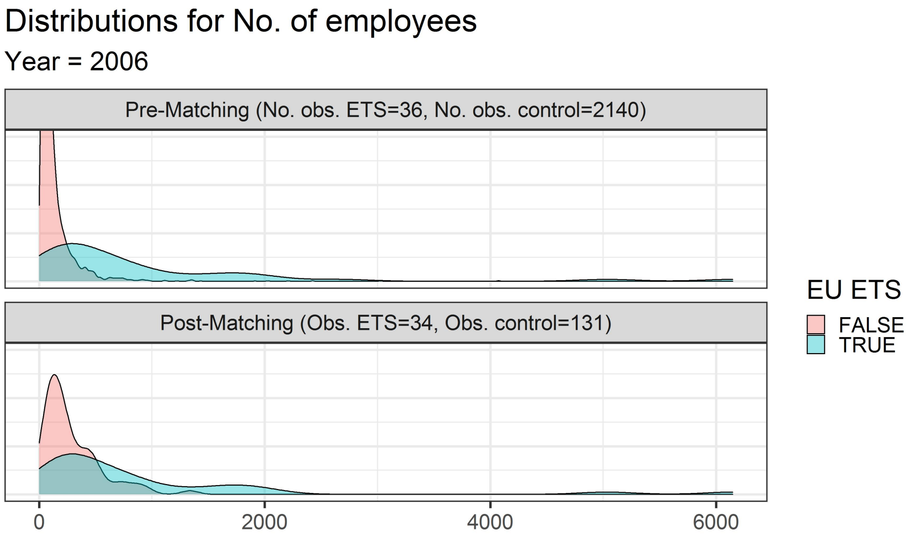


(c) *Total wages*

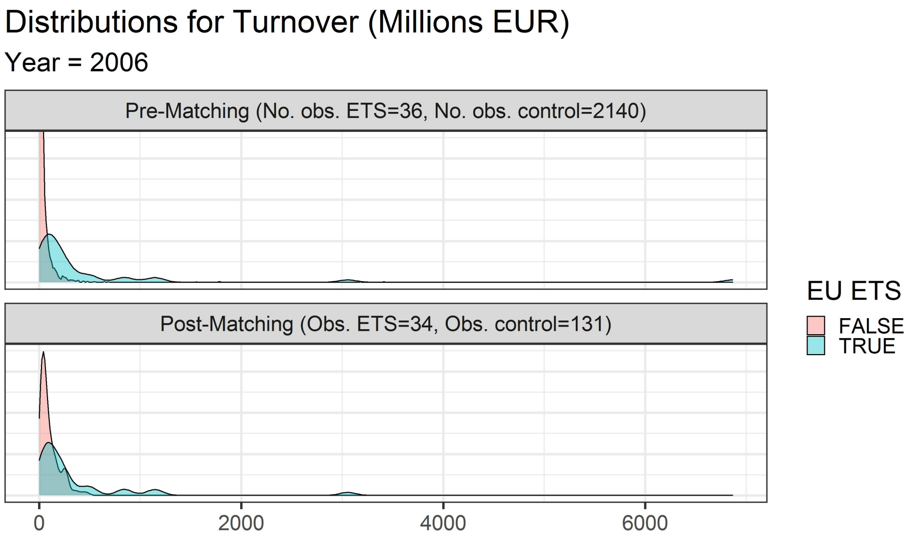


(d) *Energy expenditures*

Figure C.1: *Distributions of variables before and after matching for treated and control firms in 2003. (Cont'd.)*

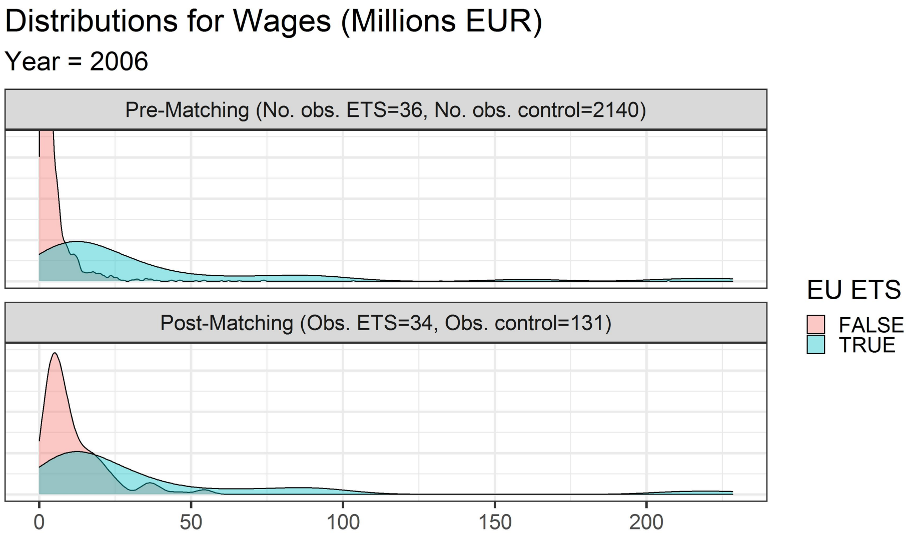


(a) *Number of employees*

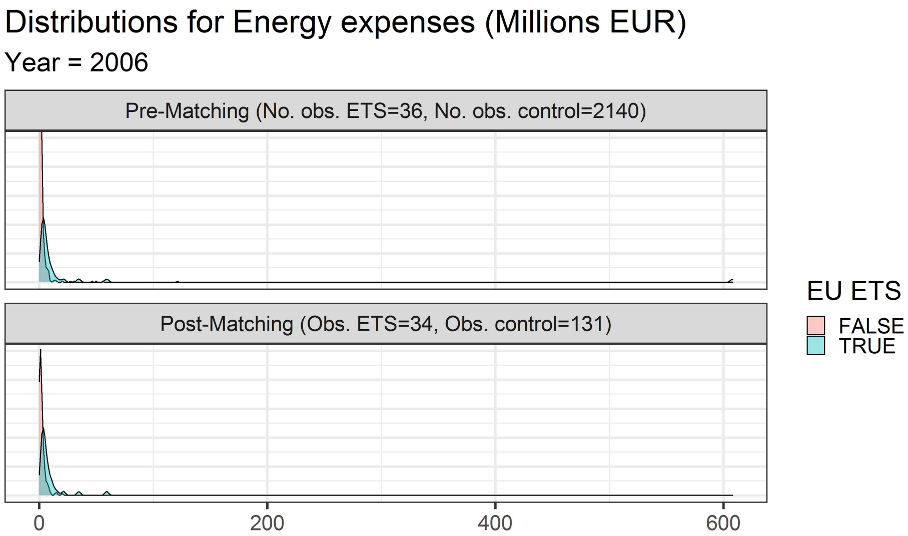


(b) *Turnover*

Figure C.2: *Distributions of variables before and after matching for treated and control firms in 2006.*

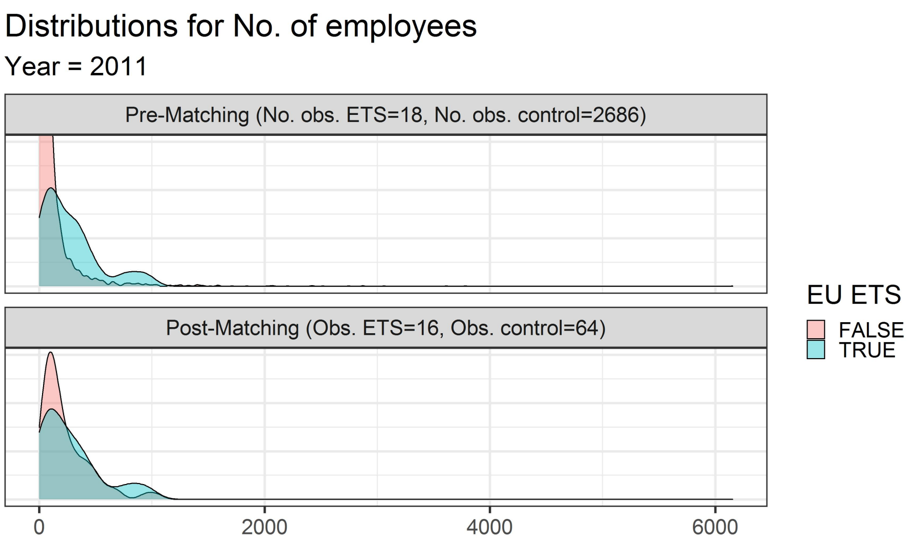


(c) *Total wages*

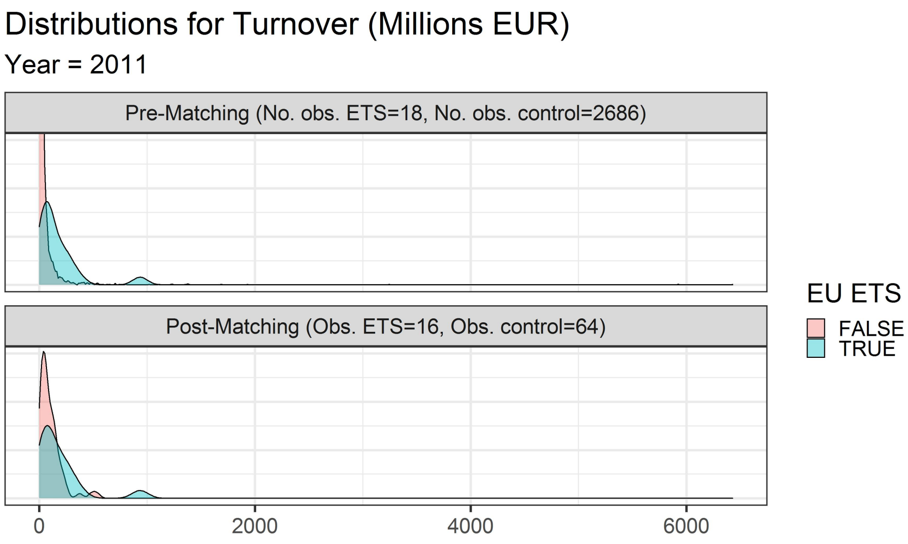


(d) *Energy expenditures*

Figure C.2: *Distributions of variables before and after matching for treated and control firms in 2006. (Cont'd.)*

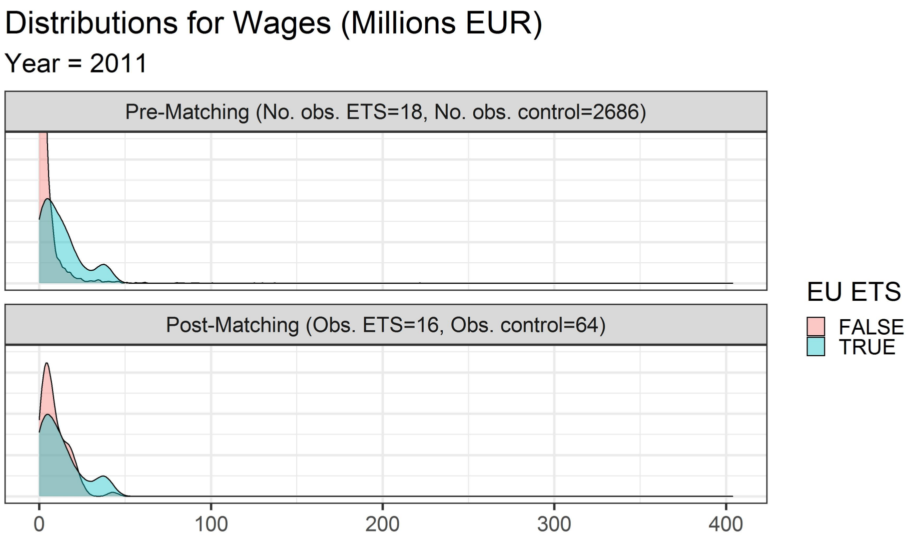


(a) *Number of employees*

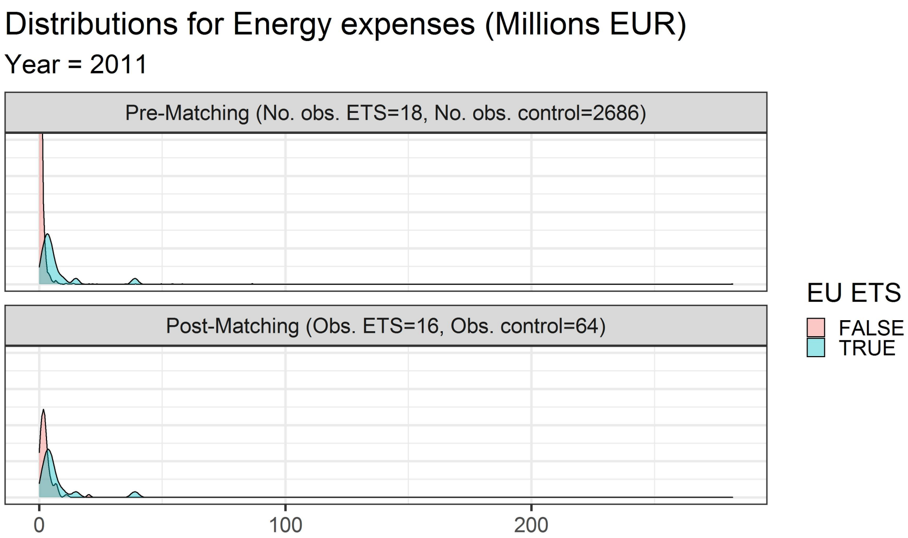


(b) *Turnover*

Figure C.3: *Distributions of variables before and after matching for treated and control firms in 2011.*



(c) *Total wages*



(d) *Energy expenditures*

Figure C.3: *Distributions of variables before and after matching for treated and control firms in 2011. (Cont'd.)*

D Additional tables and figures

D.1 Main results

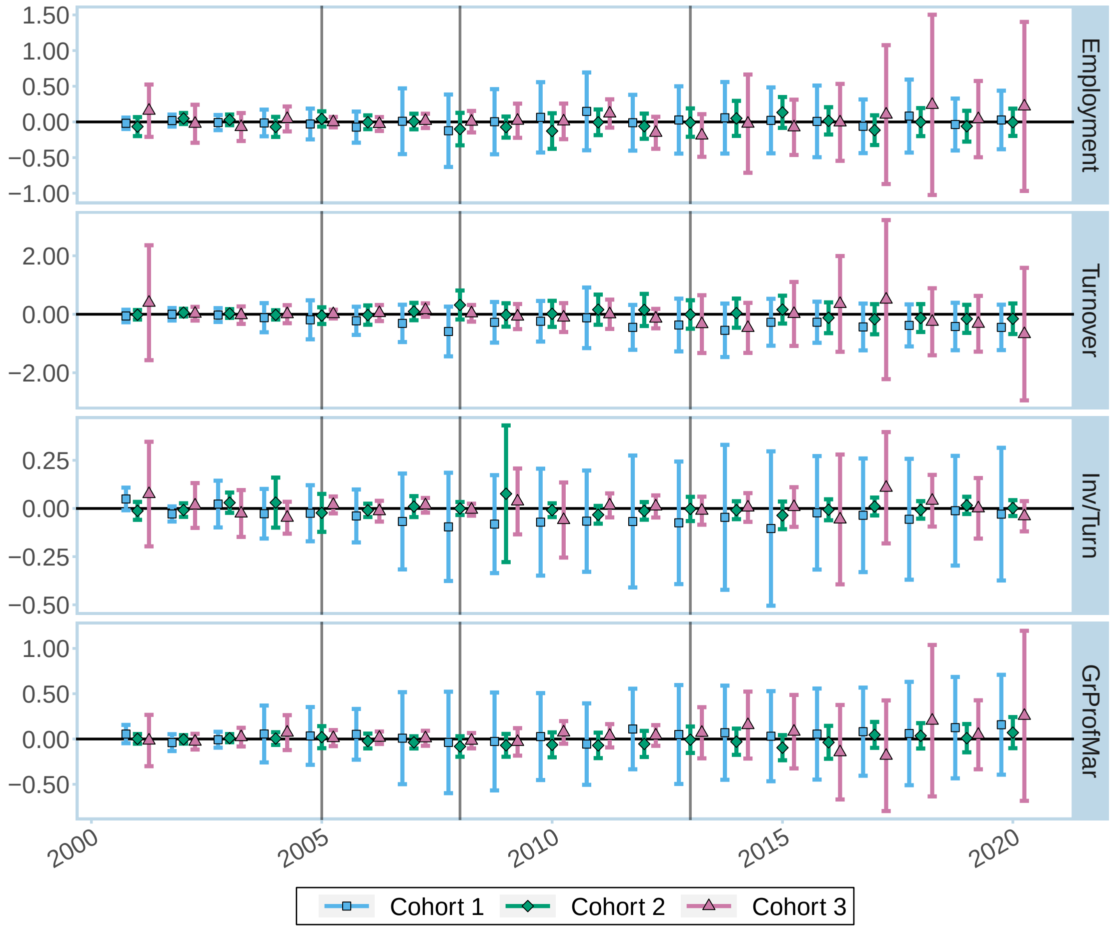


Figure D.1: *Baseline CS estimates, (4.3), using never treated firms as control group. Bars represent 95% confidence bars, based on firm clustered standard errors.*

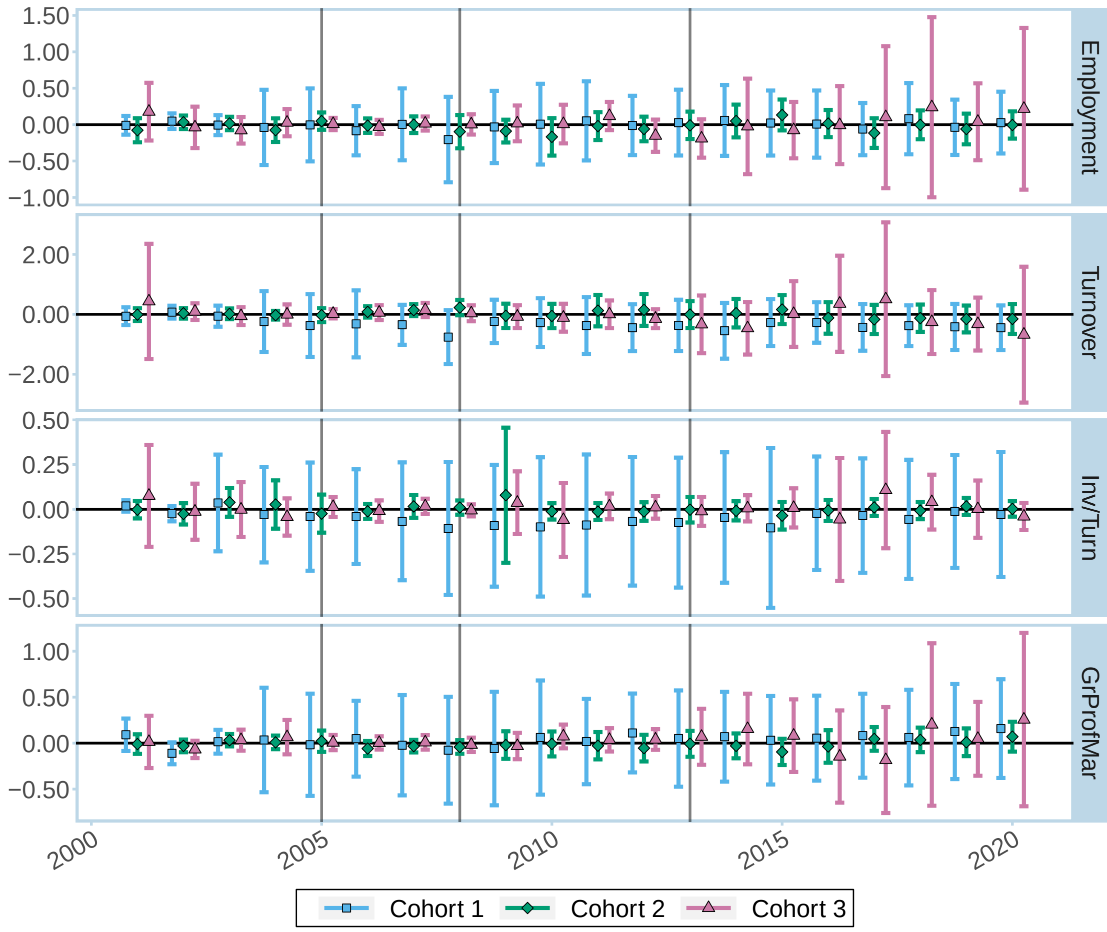


Figure D.2: *Baseline CS estimates, (4.3), using not yet treated firms as control group. Bars represent 95% confidence bars, based on firm clustered standard errors.*

D.2 Discussion

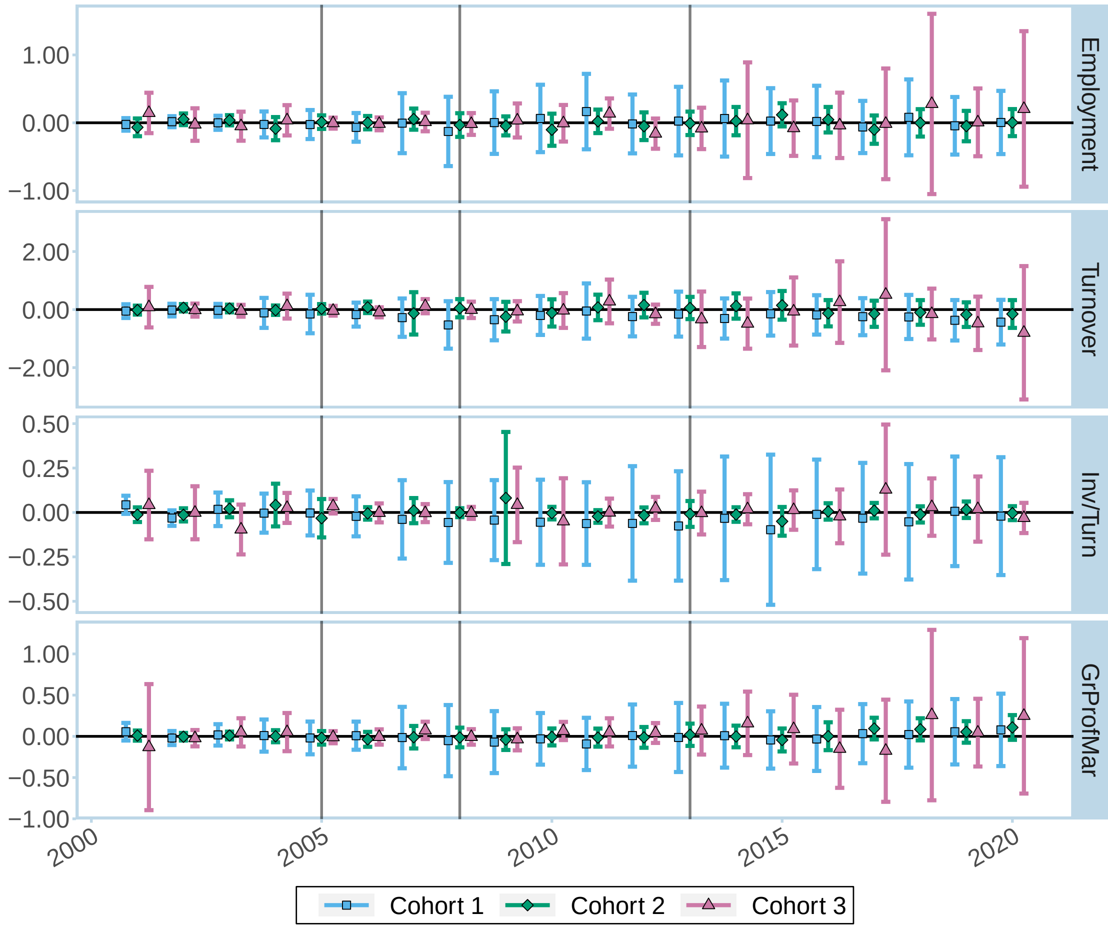


Figure D.3: *CS estimates, (4.3), including pre-treatment trends of the dependent variable in the control variables. Aggregated to cohort-phase level and using not yet treated firms as control group. Bars represent 95% confidence bars, based on firm clustered standard errors.*

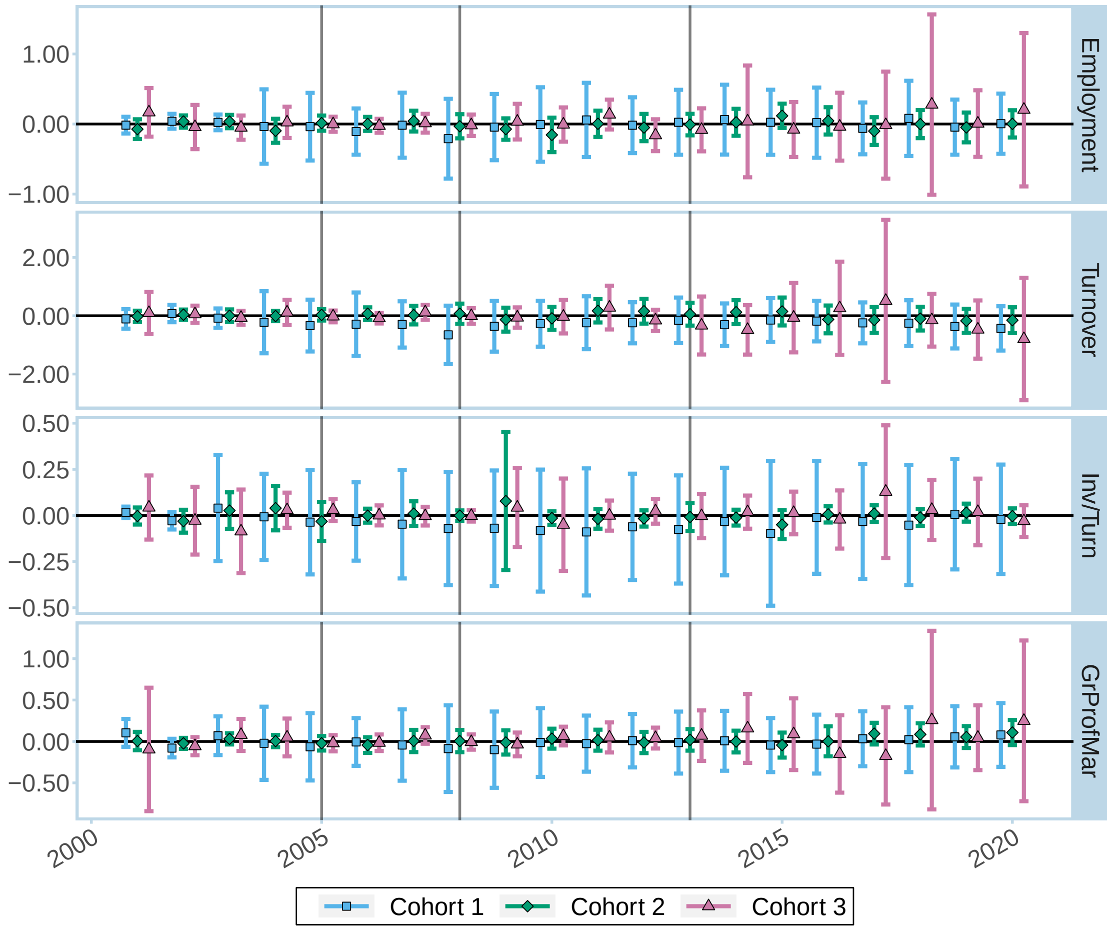


Figure D.4: *CS estimates, (4.3), including pre-treatment trends of the dependent variable in the control variables. Using never treated firms as control group. Bars represent 95% confidence bars, based on firm clustered standard errors.*

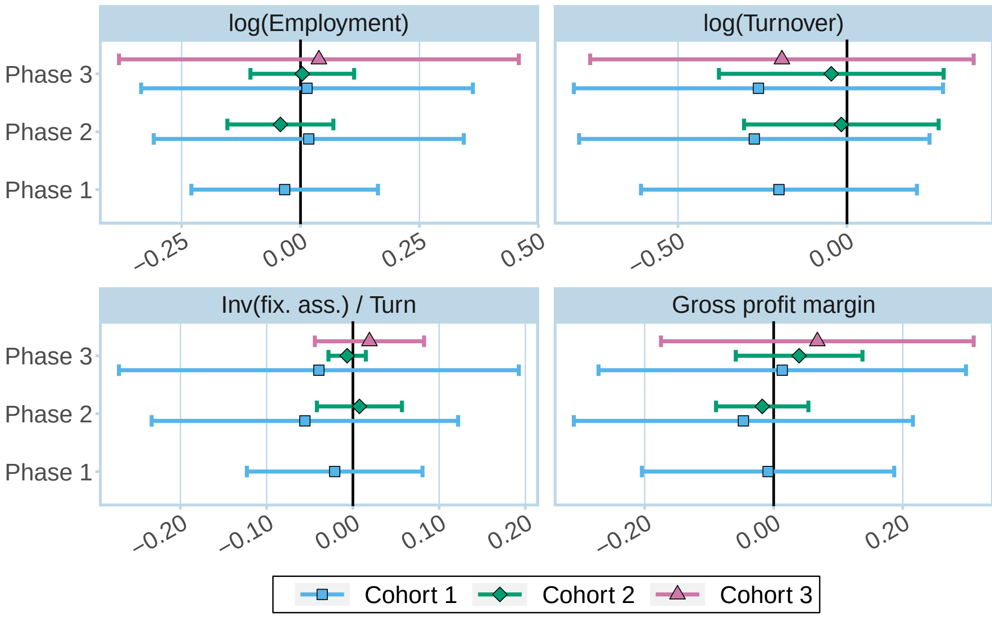


Figure D.5: *CS estimates, (4.3), including pre-treatment trends of the dependent variable in the control variables. Using not yet treated firms as control group. Bars represent 95% confidence bars, based on firm clustered standard errors.*

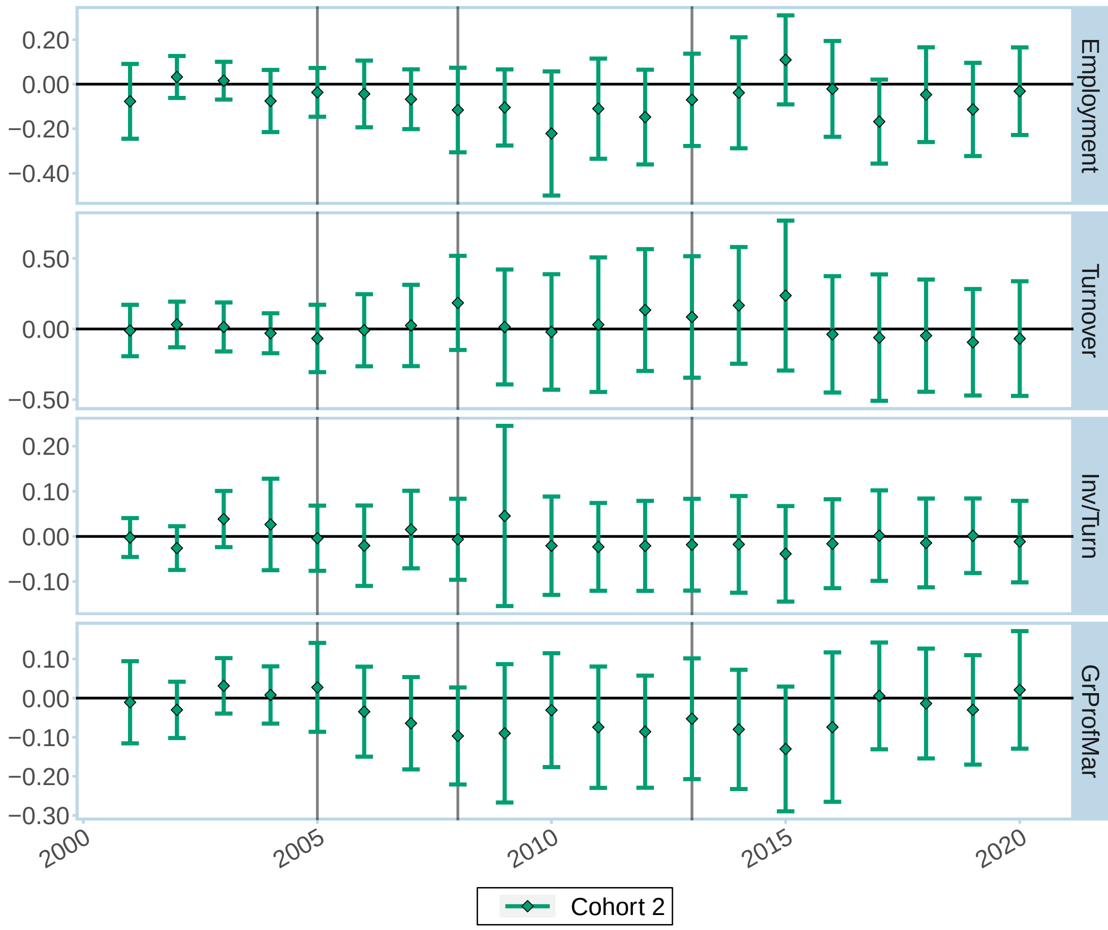


Figure D.6: *CS estimates, (4.3), only for cohort 2, when pretending firms in this cohort became treated in phase 1 already. Bars represent 95% confidence bars, based on firm clustered standard errors.*

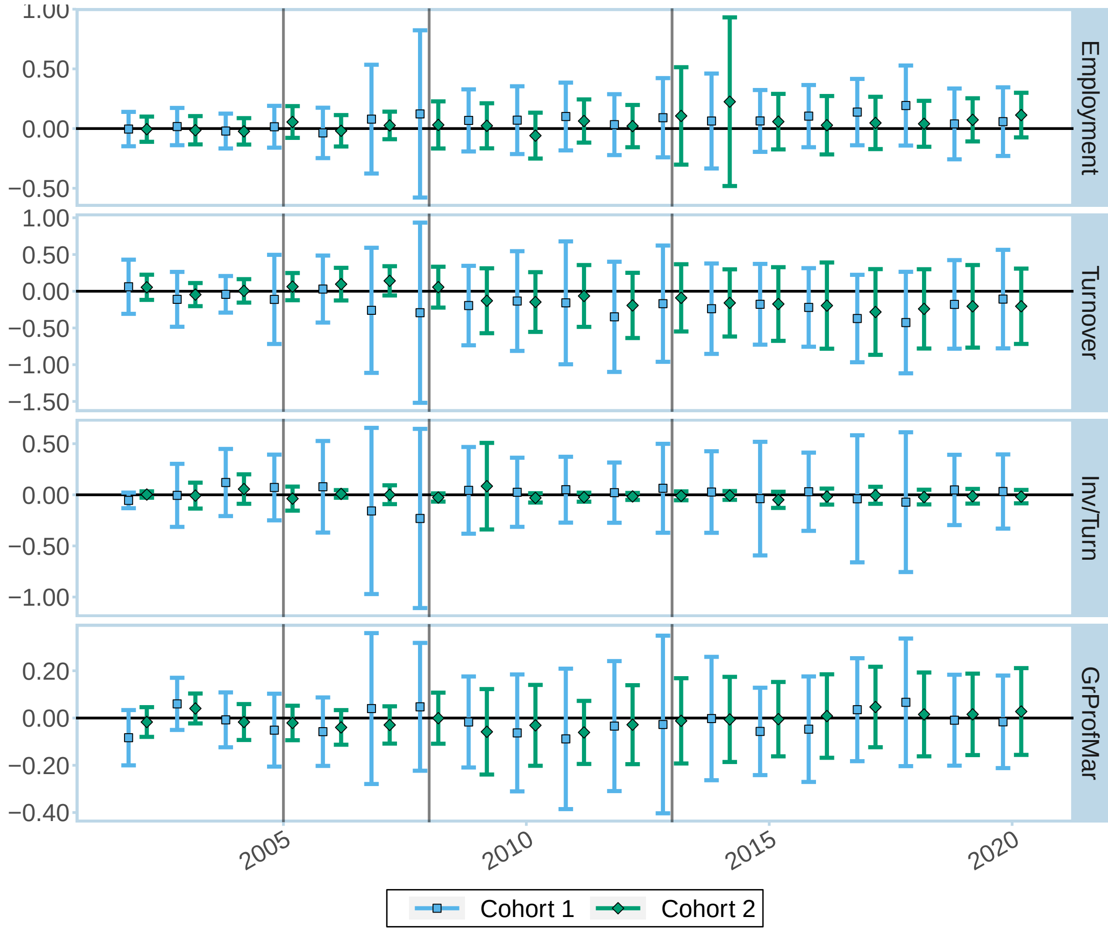


Figure D.7: *CS estimates,* (4.3)*, using a fully balanced panel. Using not yet treated firms as control group. Bars represent 95% confidence bars, based on firm clustered standard errors.*

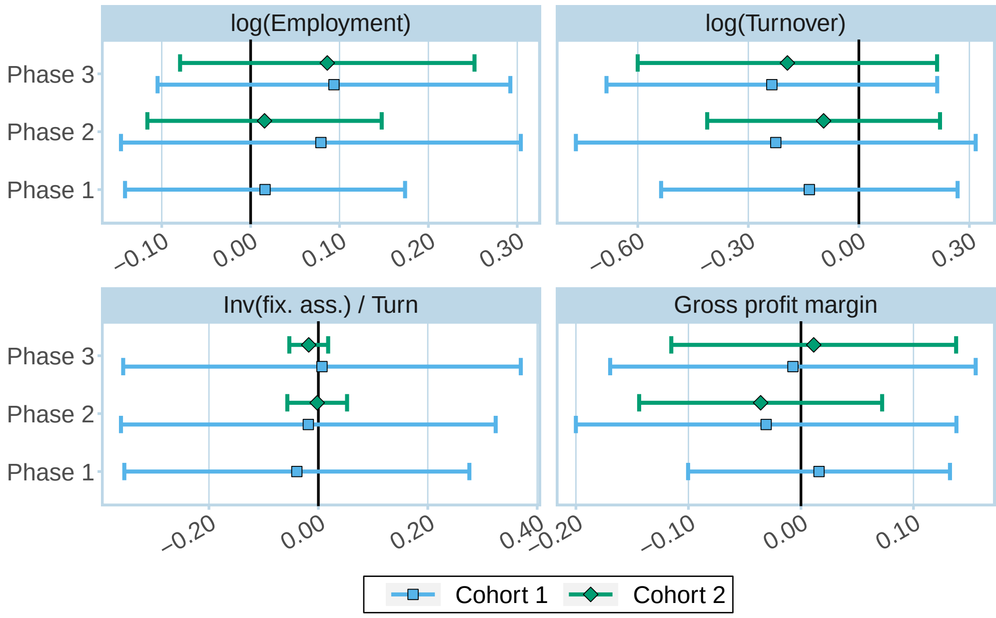


Figure D.8: *CS estimates, (4.3), using a fully balanced panel. Aggregated to cohort-phase level and using not yet treated firms as control group. Bars represent 95% confidence bars, based on firm clustered standard errors.*

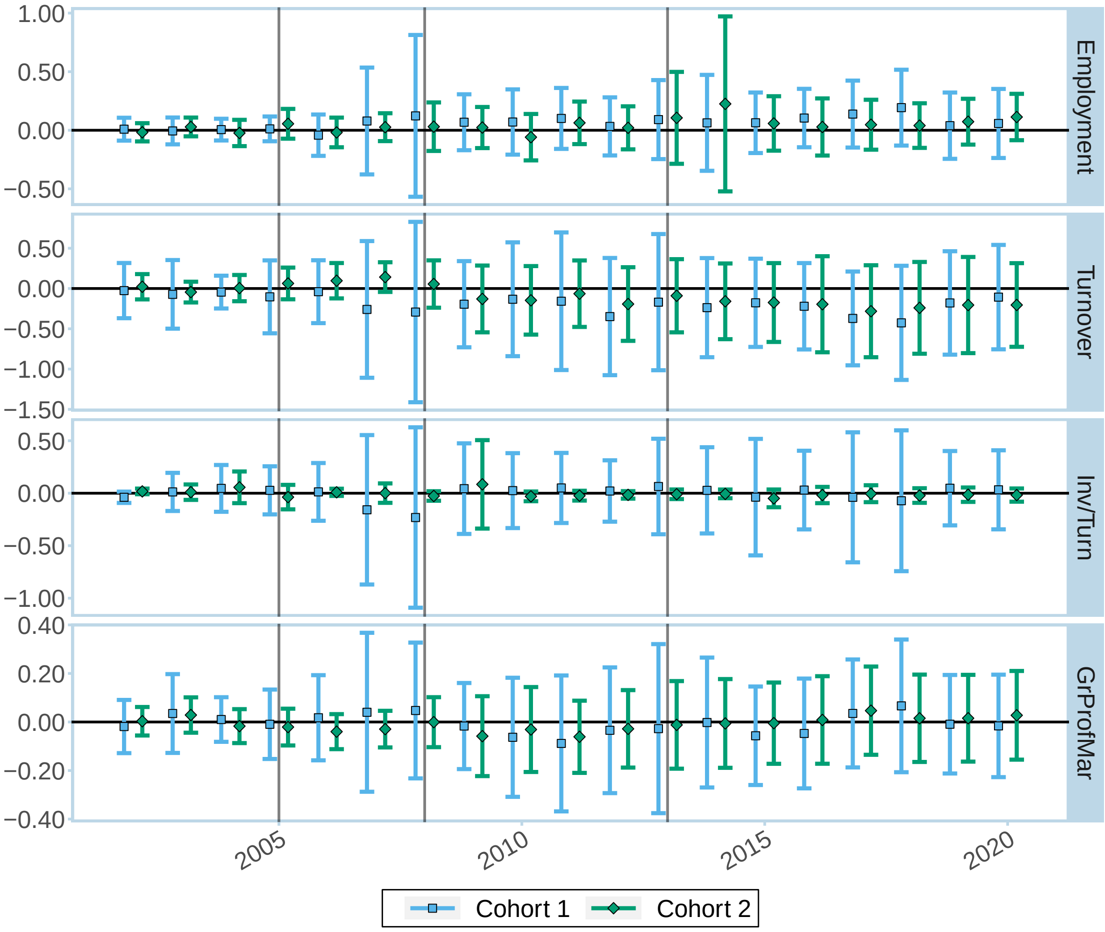


Figure D.9: *CS estimates, (4.3), using a fully balanced panel. Using not yet treated firms as control group. Bars represent 95% confidence bars, based on firm clustered standard errors.*

Table D.1: *Regression results for the operating margin.*

	Operating margin		
	TWFE	CS nyt	CS never
Panel A: Overall aggregation			
Overall ETS	0.010	0.026	0.019
	(0.014)	(0.132)	(0.147)
Panel B: Cohort aggregation			
Cohort 1	−0.008	0.049	0.033
	(0.015)	(0.220)	(0.254)
Cohort 2	0.030	−0.020	−0.016
	(0.034)	(0.034)	(0.034)
Cohort 3	0.015	0.061	0.061
	(0.015)	(0.046)	(0.047)
Panel C: Cohort-Phase aggregation			
Cohort 1 - Phase 1	0.008	0.053	0.027
	(0.009)	(0.162)	(0.247)
Cohort 1 - Phase 2	−0.022	−0.003	−0.037
	(0.022)	(0.201)	(0.264)
Cohort 1 - Phase 3	−0.007	0.079	0.079
	(0.020)	(0.259)	(0.234)
Cohort 2 - Phase 2	0.042	−0.031	−0.019
	(0.043)	(0.038)	(0.039)
Cohort 2 - Phase 3	0.015	−0.014	−0.014
	(0.030)	(0.037)	(0.034)
Cohort 3 - Phase 3	0.013	0.061	0.061
	(0.016)	(0.046)	(0.046)
Observations	6070	10819	10819
Adjusted R2	0.260		
Placebo test pass	N		
Wald Stat		16.834	23.845
Wald p-value		0.817	0.412

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TWFE and CS refer to two-way fixed effects and Callaway & Sant'Anna, respectively, to indicate different regression methods. Note that the table shows three different regressions for TWFE ((4.1), (B.3), (B.4)) and three aggregations based on the same original estimation of year and cohort ATTs for CS, (4.3). The statistics at the bottom of the table are from the cohort-phase aggregation. CS standard errors are bootstrapped and clustered at the firm level both for TWFE and CS estimations. TWFE always includes firm and phase FEs. nyt and nt refer to not-yet treated and never treated, respectively, referring to different samples of control firms.

Table D.2: *Regression results for further investment measures.*

	Inv (machines) / Turnover			Inv(fix. ass.) / Emp		
	TWFE	CS nyt	CS never	TWFE	CS nyt	CS never
Panel A: Overall aggregation						
Overall ETS	−0.114	−0.107	−0.109	−2.244	−9.494	−9.527
	(0.114)	(0.071)	(0.072)	(2.855)	(15.999)	(17.698)
Panel B: Cohort aggregation						
Cohort 1	−0.087	−0.158	−0.163	1.450	−12.572	−12.795
	(0.092)	(0.117)	(0.127)	(4.025)	(28.948)	(31.834)
Cohort 2	−0.113	−0.038	−0.036	−5.643	−3.758	−3.502
	(0.114)	(0.052)	(0.047)	(4.561)	(4.969)	(5.339)
Cohort 3	−0.176	−0.068	−0.068	−4.363	−12.911	−12.911
	(0.163)	(0.072)	(0.074)	(5.938)	(18.175)	(19.143)
Panel C: Cohort-Phase aggregation						
Cohort 1 - Phase 1	−0.002	−0.035	−0.041	4.265	−8.351	−5.577
	(0.013)	(0.052)	(0.086)	(6.393)	(17.769)	(22.120)
Cohort 1 - Phase 2	−0.012	−0.061	−0.073	−2.369	−33.075	−35.454
	(0.017)	(0.084)	(0.105)	(4.518)	(32.862)	(36.214)
Cohort 1 - Phase 3	−0.213	−0.266	−0.266	2.979	−1.340	−1.340
	(0.221)	(0.163)	(0.162)	(3.920)	(28.736)	(29.975)
Cohort 2 - Phase 2	−0.005	0.069	0.076	−6.024	−4.350	−3.684
	(0.015)	(0.076)	(0.077)	(4.422)	(6.367)	(7.303)
Cohort 2 - Phase 3	−0.223	−0.106	−0.106	−5.964	−3.387	−3.387
	(0.218)	(0.060)	(0.068)	(4.824)	(4.838)	(4.645)
Cohort 3 - Phase 3	−0.225	−0.068	−0.068	−3.959	−12.911	−12.911
	(0.211)	(0.077)	(0.073)	(5.951)	(18.522)	(17.455)
Observations	6173	10819	10819	6193	10928	10928
Adjusted R2	0.001			0.187		
Placebo test pass	Y			Y		
Wald Stat		31.836	23.942		31.399	29.169
Wald p-value		0.104	0.407		0.113	0.175

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TWFE and CS refer to two-way fixed effects and Callaway & Sant'Anna, respectively, to indicate different regression methods. Note that the table shows three different regressions for TWFE ((4.1), (B.3), (B.4)) and three aggregations based on the same original estimation of year and cohort ATTs for CS, (4.3). The statistics at the bottom of the table are from the cohort-phase aggregation. CS standard errors are bootstrapped and clustered at the firm level both for TWFE and CS estimations. TWFE always includes firm and phase FEs. nyt and nt refer to not-yet treated and never treated, respectively, referring to different samples of control firms.