Benedikte Svanes Sørbye

Do sustainability commitments by companies actually lead to more sustainable practices?

An exploratory analysis of the effects on chemical releases from joining the United Nations Global Compact

Master's thesis spring 2023
Oslo Business School
Oslo Metropolitan University
MSc in economics and business administration

Abstract

The aim of this master thesis is to examine whether joining the United Nations Global Compact has a significant effect on the release levels of toxic chemicals. The UN Global Compact is a global corporate sustainability initiative with over twenty thousand participants today. To address the aim of study, I use data on chemical releases as reported by companies in the United States to the Toxics Release Inventory from 1987 to 2021.

Using an exploratory approach, I have employed several different models and subsets of the data to study the relationship between joining the UN Global Compact and chemical releases. In the main model, the variable of interest was the total chemical releases in pounds. The analyses from this model found a statistically significant and robust negative effect, indicating that joining the UN Global Compact is associated with large reductions in chemical releases by American companies. This finding was in part supported by a toxicity-weighted model. A model assessing percentage changes in chemical releases did not show significant effects.

In conclusion, the analyses of this thesis, indicated that joining the UN Global Compact was associated with reductions in chemical releases. This could reflect a true effect of the program on chemical releases, however, the data generating process does not allow for causal interpretation of the finding. Alternative explanations, especially self-selection, cannot be excluded.

Foreword

This master thesis represents the final chapter of a two-year master's degree at the Oslo Metropolitan University, terminating spring 2023. My interest for the chosen subject originated from a general interest in economics as a tool for environmental protection, but I have developed a much deeper understanding and curiosity for the role of chemicals in our economy and their effects on human and environmental health. Building on the knowledge gained from the courses Statistical methods and Econometrics, I wished to improve my command of the tools they offer with this thesis. Through the painful process of trying and failing, I have learned much, even if there is still a lot left to conquer. Working on this thesis has been both challenging and fun, I have cried, I have laughed, sometimes at the same time. To work with such a large dataset has required a steep learning curve, but I am pleased with the skillset and knowledge this process has earned me.

First and foremost, I would like to thank my supervisor Svenn Jensen for all the support. Thank you for your engaging guidance, always delivered with a sunny disposition, and thank you for keeping me on track when I teeter on the edge of a rabbit hole. Secondly, I would like to thank Fenella Carpena for taking the time to help me navigate through the winding maze of empirical research and for coming up with the research question. I would also like to thank Per Arne Tufte for making statistics downright enjoyable.

With this master thesis I take my leave of OsloMet, thanks for the knowledge!

Benedikte Svanes Sørbye

OsloMet, Mai 2023

Table of Contents

List of figures and tables

Figures

Tables

Stata Outputs

List of acronyms and terminology

- PFAS – per- and polyfluoroalkyl substances. These chemicals do not break down in the environment and bioaccumulate in humans, animals and plants. Many of them are known to cause several adverse health effects, including cancer and birth defects (European Chemicals Agency, 2023). Used to make, among other things, Teflon pans and waterproof clothing.
- PBT – Persistent Bioaccumulative Toxic chemicals. Persistent as they resist environmental degradation. Bioaccumulative as they build up in humans and wildlife. Toxic as they cause adverse health effects in humans and animals, ranging from cancer to damaged nervous systems. Used for example as flame retardants in diverse products (Toxic-Free Future, 2023).
- Dioxins – Chemical compounds that are persistent environmental pollutants. They accumulate in food, such as meat, and are highly toxic. They can cause reproductive problems, damage the immune system and cause cancer to name a few. Mainly a by-product of industrial processes (World Health Organization, 2016).
- EDC's - Endocrine-disrupting chemicals. Linked with developmental, reproductive, brain and immune problems. Found in everyday products such as cosmetics and plastics. (National Institute of Environmental Health Sciences, 2023)
- Greenwashing –Definition (Cambridge Dictionary, n.d.): "to make people believe that your company is doing more to protect the environment than it really is".
- Carcinogen – A substance capable of causing cancer.
- TRI – Toxics Release inventory
- EPA – United States Environmental Protection Agency
- UNGC – United Nations Global compact

1 Introduction

1.1 Motivation and some historical context

In the 1920's a chemist by the name of Thomas Midgley found the solution to a problem that had been plaguing the automobile industry. The issue was called the «knock», a noise that came from premature ignition and it reduced efficiency and damaged the engines. The chemist found that adding lead to the fuel solved the problem. The same chemist later came up with a solution to another problem. Refrigerators at this time were extremely unsafe and the chemist's employer, General Motors, had a refrigerator division in dire need of help. Tasked with this problem, Midgley again found a solution, a gas they would later name Freon. This gas made refrigerators stable and safe, and it also would later be used for air-conditioning and other inventions. These advancements made the company loads of money and improved the lives of their consumers. But in solving these problems, Midgley created new, and very severe, issues. Lead poisoning caused cancer, madness and affected brain development for generations to come. The gas Freon, depleted the ozone layer at an alarming rate (Harford, 2022).

Chemicals have in many ways shaped our economy. Most of our products rely on them and they have created many advancements in living standards. They are used to make the fabric of the clothes you wear, in your mobile phone battery, in the plastic box you keep your lunch in, in your shampoo, in your toothpaste and sprayed on your apple to keep parasites away. Bakelite, invented in 1907 (Science History Institute, 2017), began the era of synthetic plastics which are now a large part of everyday tools and necessary for medical machinery. Chemicals are everywhere and some tend to outstay their welcome. Lead in fuel was not removed until more than 60 years after its invention despite evidence of its harmful effects (Harford, 2022). And today, banning a chemical from use requires substantial evidence of its harmfulness, which can take decades of research to prove. In the meantime, these chemicals are free to roam, and the consequences may not immediately be apparent. But to preserve human health and the environment, something must be done. This is not only a question of morals, but also of economics; for instance, endocrine-disrupting chemicals (EDC's) are calculated to cost the US around 2% of its gross annual domestic product in human disease and disabilities (Attina et al., 2016). To put this number into perspective, every year the

United States loses 340 billion dollars, which is more than two thirds of the total Norwegian gross domestic product.

In the US, the regulations for chemical releases are not very stringent. Voluntary environmental programs (VEP's) are popularly used today as an addition or alternative to government regulations. The idea is for companies to internalise negative externalities and to take on environmental practices that go beyond legal requirements and in return, they may use the program as a signal to stakeholders (Potoski & Prakash, 2013). A voluntary environmental program can provide a credible stamp of approval, communicating the firm's sustainability efforts to stakeholders when this cannot be observed fully. Voluntary programs can in theory be an economically efficient alternative to stricter regulations, allowing the firm flexibility in improving sustainability in the way that best suits the individual firm (OECD, n.d.). The important question is, however, do voluntary programs contribute to actual improvements in the environmental performance of companies? And what role can these programs play in chemical management?

1.2 Research question

The following pages will attempt to answer the question:

"Do sustainability commitments by companies actually lead to more sustainable practices?"

Sustainability is a broad term and for this thesis I have chosen to focus on chemical releases, particularly releases of chemicals that are poisonous or harmful, in other words, toxic chemicals. Toxic chemicals can cause a wide range of effects on the human body, such as irritation to the skin or eyes, difficulty of breathing, headaches, nausea, behavioural abnormalities, genetic mutations, cancer, reproductive impairment, kidney failure, physical deformations and birth defects. They can further cause damage to the environment, destroy ecosystems and cause explosions or fire (EPA, n.d.-a). One example is the Cuyahoga River in Cleveland which would regularly catch on fire between the 1860's and the 1960's as a result of the vast amount of pollution from the manufacturing companies along the river. This was merely accepted as the price for a thriving industry and booming economy, even though some of these fires caused millions of dollars' worth of damage and even killed people

(Boissoneault, 2019). Chemical releases pose a threat to human health and to our environment, but it is also an economic issue. As chemical releases are not heavily regulated, this provides a great opportunity to study the effects of voluntary programs without high regulatory disturbance. There are different ways to determine sustainable practices in terms of chemical usage, such as how much of the chemicals are recycled or treated, or how much chemicals are used in total. In this thesis, I have chosen to specifically address total chemical releases. These are the chemicals that go straight into our air, water and earth. This measure is also used as the variable of interest in articles with similar research questions, for example, Bui and Kapon (2012) and Sam et al. (2009). Bui and Kapon (2012) study pollution prevention programs and their effects on toxic pollution, and Sam et al. (2009) analyse the TRI 33/50 program, a voluntary program from the EPA to reduce 17 targeted toxic chemicals, and its effect on releases of these targeted chemicals. For this master thesis, I have chosen the United Nations Global Compact (UNGC) as the voluntary program to be examined. Their website states that it is the "world's largest corporate sustainability initiative"(United Nations Global Compact, n.d.-c). The participant list is openly available online, and since many chemical releases not only affect the country of origin, but the whole planet, I wish to study a program with global reach. To my knowledge, there has been no previous study investigating the effects of the UN Global Compact specifically on chemical releases. The thesis thus aims to provide new insights.

1.3 Structure

The paper consists of six sections. Section one covers the introduction and explains the chosen research question while section two examines existing literature and provides all necessary background information. Section three presents and details the dataset I have been working with and section four explains the chosen methodology. Section five presents the results and analysis, while section six provides a conclusion of the paper.

2 Background and Theory

To understand how to best proceed with the research question, it is useful to get a better sense of what earlier research has already found. In addition to this, we need to understand the regulations and institutions that affect the research question. I will first introduce three subjects of importance: the United Nations Global Compact, the voluntary program of interest; the Toxics Release Inventory, the regulation that provides the dataset used; and the Clean Air Act, the most important current regulation for chemical releases. Then I will present some of the existing literature relevant to this paper.

2.1 The United Nations Global Compact

The United Nations Global Compact, hereby named UNGC or the initiative or the program, is a large global program reaching out to companies to improve in areas of human rights, labour, environment and anti-corruption. Three of the ten principles address environmental issues:

- Principle 7: Businesses should support a precautionary approach to environmental challenges
- Principle 8: Businesses should undertake initiatives to promote greater environmental responsibility
- Principle 9: Businesses should encourage the development and diffusion of environmentally friendly technologies (United Nations Global Compact, n.d.-a).

The UN Global Compact was launched in 2000 and aims to encourage corporations to take on greater social responsibility and align themselves in the battle for a more sustainable global economy (Annan, 2002). The way the program works is that the CEO, with approval from the board, sends a letter of intent to integrate the ten principles in the company's strategy and operating procedures. The company subsequently implements changes to improve in these areas and provides written reports of the efforts made annually (United Nations Global Compact, n.d.-b).

The UN Global Compact covers several areas of corporate social responsibility, but for this analysis I will only look at the environmental aspects of the initiative.

2.2 Toxics Release Inventory

The Toxics Release inventory (TRI) is a mandatory program that requires companies in chosen industries that fulfil certain requirements to annually report on how they manage certain toxic chemicals. The TRI provides a list of chemicals that must be reported on if they are used in excess of the reporting threshold. The companies report how much is released, recycled, used for energy recovery or treated of each chemical they use on the TRI's list. The threshold indicates how much of the chemical must be used or produced before it is required by law to be reported. For dioxin and dioxin-like compounds, the threshold for reporting is 0.1 grams. However for a TRI chemical not classified as Persistent Bioaccumulative Toxic chemicals (PBT) or Per- and Polyfluoroalkyl substances (PFAS), the reporting threshold can be 10,000 or 20,000 pounds depending on how the chemical is used or produced (EPA, n.d.-b). The chemicals on the TRI list do not cover all toxic chemicals used in the US but are toxic chemicals that cause cancer, chronic or significant acute health effects in humans, and/ or have significant adverse environmental effects. Both companies and individuals can petition for chemicals to be put on or taken off the list. The program does not control how much a company may use or release of a chemical but is a tool to provide transparency when handling environmental issues. Instead of regulating levels of chemical usage and release, it provides information to the public and third parties to make the company more accountable and creates incentive to improve their environmental impact. The TRI was established by the Emergency Planning and Community Right-to-Know Act (EPCRA) passed in 1986. The mission of the act was to support emergency planning and make information on toxic releases publicly available (EPA, 2023).

The information reported from all these companies comprises the TRI dataset, which is available to the public on their website and which constitutes the raw dataset used in this paper. The first reporting year is 1987, and they currently have annual data for all subsequent years until and including 2021. The data is self-reported, the TRI assists facilities in the reporting process and conduct quality control phone calls to check the accuracy of the

reported data after reporting deadlines (EPA, 2022b). The reporting requirements have changed over time, for example in terms of which industries must report, which chemicals must be reported on, reporting thresholds and additions of chemical categories. For example, in 1997 the list of which industries were required to report was expanded. In 1999 seven new PBT's were added and thresholds for existing PBT's on the list were lowered to 100 pounds. 150 PFAS chemicals were added in 2019 (EPA, 2022a). When analysing the dataset, it will be useful to remember these changes in reporting requirements and consider their possible effects.

2.3 Clean Air Act

In terms of regulations on chemicals in the US, the most important one is the Clean Air Act. The Clean Air Act is a federal law which regulates air emissions to protect human health and the environment. It requires the Environmental Protection Agency (EPA) to establish air quality standards which each state is then required to make plans for achieving in a way that can be enforced. The Act was first implemented in 1955, and then heavily revised in 1970, 1977 and 1990 to take on the increasing challenge of upholding air quality (Congressional Research Service, 2022, p. 1).

In the analysis, it may be a concern that some of the variations in chemical releases arise from regulation, and not voluntary efforts induced by joining the UNGC. However, voluntary programs can also assist companies in better adhering to existing regulations. The Clean Air Act is the most influential regulation, and approximately 60% of the observations in the dataset are from chemicals regulated under the Clean Air Act. If we take the sum of all the total chemical releases in pounds, meaning all chemicals that are released to water, air or land, almost 50% of this total sum comes from chemicals that are regulated under the Clean Air Act. I will therefore address this in the analysis.

2.4 Background literature

The existing literature is still ambiguous on the effects of voluntary environmental programs in general. Bui and Kapon (2012), uses the TRI dataset to study voluntary pollution

prevention programs and finds evidence that these programs have a significant effect on reducing average facility level toxic releases. Sam et al. (2009) also uses the TRI dataset and studies the effects of the TRI's 33/50 program, the United States Environmental Protections Agency's first voluntary program to reduce 17 targeted toxic chemicals. The article finds a significant negative effect on chemical releases of targeted chemicals from the program. Potoski and Prakash (2013) finds a significant effect on reductions in air pollution from the adoption of ISO 14001, a non-governmental organization that formulates global technical and management standards. The uptake of the program does not have a significant effect on water pollution, only air pollution, and the researchers theorise that this is due to air pollution being more visible to stakeholders, and therefore the reduction efforts are directed towards this.

There can also be a difference in program efficiency depending on when the participants join, Kube et al. (2019) finds weak evidence for reductions for early joiners, but none for late joiners for the German Eco-Management and Audit Scheme (EMAS). Gamper-Rabindran (2006) finds that when controlling for participants' self-selection into the TRI's 33/50 program, participants do not reduce health-indexed emissions in key industries compared to non-participants. Vidovic and Khanna (2007) finds that the effect other papers find from the TRI 33/50 program on chemical releases is not due to the program itself, but rather an independent trend. As the literature shows, there seems to be no conclusion on whether voluntary environmental programs have the desired effect of improving companies' environmental performance.

It is important to understand why companies join voluntary programs, as the motivations can sometimes explain what effects are found. Innes and Sam (2008) shows that companies join to reduce regulatory scrutiny; regulatory bodies shift their inspections to non-participants resulting in participants experiencing less regulatory pressure. Brouhle et al. (2009) comments on how companies join such programs in an attempt to deter stricter future regulations and Innes (2006) shows that companies join to deter consumer boycotts. Zhang and Khanna (2020) shows how public pressure and the risk of being labelled greenwashers affects companies' decision to join such programs. Greenwashers are

companies that present themselves as being more environmentally friendly than they actually are.

When assessing the United Nations Global Compact in particular, Orzes et al. (2018) provides an exhaustive overview of possible motivating factors for companies deciding to join the UNGC. These motivating factors include pressure from stakeholders such as activists, non-governmental organizations, employees, competitors, investors, customers, government, media and unions; improving reputation; network opportunities with potential to create new partnerships; reduction of costs from improved technology and less waste; improvement in productivity from increased worker welfare; improved sales from entering new markets; increase in stock price and ethical alignment with management, meaning the company itself has a preference for sustainable practices.

A major weakness of the program is that companies may join the initiative without making any changes to include the UNGC principles in their practices; this is called decoupling. This can be due to the initiative's lack of independent audits, lack of enforcement, or lenient reporting requirements (Orzes et al., 2018). Berliner and Prakash (2014) and Sethi and Schepers (2014) conclude that the UNGC does not have an effect on improving corporate social responsibility (CSR) performance for participants. Li and Wu (2020) finds that private firms do improve their environmental, social and governance (ESG) performance, but public ones do not and are more likely to engage in decoupling. The dependent variable in this paper is a monthly count of negative ESG incidents for a firm, found in media, commercial or regulatory documents. Ortas et al. (2015) finds that UNGC participation and engagement improves companies' environmental, social and governance (ESG) performance overall, here using index scores for the three areas as the dependent variables. The literature is not conclusive on the effects of the UNGC, and the existing research focuses on a broad measurement of corporate social responsibility actions with little to no research on the effects of the program on environmental performance alone.

Some researchers, such as Ortas et al. (2015), view the initiative as a great opportunity for companies to take part in creating a more sustainable corporate landscape and internalise external effects, stating that the UN's authority gives the program the gravitas and credibility necessary for success. Their paper investigating companies in Japan, France and Spain (the

countries with highest number of participants in the UNGC) finds that joining the initiative led to improved corporate environmental performance. They also found a bidirectional link between corporate social responsibility and financial performance, and companies that reported high levels of engagement also reported higher gains from the initiative on corporate performance. Other researchers, like Sethi and Schepers (2014), comment on the lack of enforcement, specific goals, and measurable outcomes to hold participants accountable to their promises. They conclude that without any assurance or supervision, lacking accountability or transparency, the initiative becomes a protective cover for polluting companies, promoting adverse selection and high numbers of free-riders who gain the benefits of the UNGC membership without improving sustainability performance.

Croson and Treich (2014) in their paper on behavioural environmental economics refer to the well-established finding that firms respond to consumer preferences for environmental goods by implementing voluntary environmental efforts to please customers. Demand for corporate environmental responsibility is supported by empirical evidence, and if firms engage in the supply, this suggest that it is in their best interest to do so. Consumers might have a preference for sustainable products, or they might desire such products due to their self-image and social preferences. If the consumers have a preference for green products, the companies may face a higher risk of being labelled greenwashers. If the consumers have social preferences for being seen as green consumers, there might be less scrutiny as the value lies in the image. The UNGC offers no information that can be used to scrutinise the companies' environmental performance, so we should expect that companies join the initiative without making improvements to entice green consumers without costly changes that stakeholders cannot evaluate.

Participation in the UNGC is found to have a high impact for company performance in securing network opportunities and improving corporate image (Cetindamar & Husoy, 2007). If the UNGC improves the corporate image independent of how the company actually performs in terms of the UNGC principles, and if the company joins the initiative to improve their reputation and attract green consumers, then we should assume that there will be no actual improvement in sustainability practices. Therefore, I would expect to find no significant effect in the analysis in this thesis.

3 Data

The raw data used in this thesis comes from two sources that are available to the public online. One is the Toxics Release Inventory Basic dataset for all years from 1987 to 2021. The other is a list of participants of the United Nations Global Compact (UNGC) from their website.

3.1 Participant List UNGC

On the UNGC's website, you can see all current participants of the initiative. Today there are over 22,000 members in total (United Nations Global Compact, 2023). To find members of interest, I only want to look at companies and small to medium enterprises in the US for all sectors. This generates a list of 806 participants, meaning there are 806 companies in the US who participate in the UNGC, and 176 of these companies fulfill the criteria requiring them to report to the TRI and are hence also found in the TRI dataset. Figure 1 shows the increase in new members over time. Table 1 is an overview of how many companies joined the initiative which year. The information was acquired January 2023, therefore the number for 2023 is not complete as marked by the asterisk.

Figure 1 - New members joining UNGC

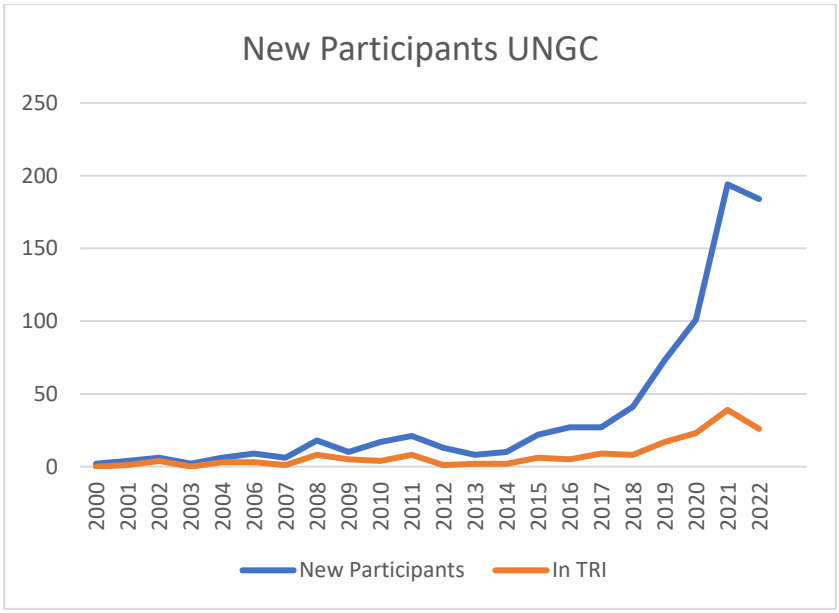


Table 1 - Participants in the UNGC

Participants joining the UN Global Compact

Year Joined	New Participants	In TRI
2000	2	0
2001	4	1
2002	6	4
2003	2	0
2004	6	3
2006	9	3
2007	6	1
2008	18	8
2009	10	5
2010	17	4
2011	21	8
2012	13	1
2013	8	2
2014	10	2
2015	22	6
2016	27	5
2017	27	9
2018	41	8
2019	73	17
2020	101	23
2021	194	39
2022	184	26
2023	5*	1*
Total	806	176

3.2 About the TRI dataset

The raw dataset from TRI is quite large spanning years 1987 to 2021, with 128 variables and almost three million observations in total. The raw data contains information at facility and chemical detail level. One observation in the raw dataset is one chemical released from a specific facility in a given year. A company can have several facilities, and each of these facilities separately report on each individual chemical they use or produce every year. In the dataset there is information on facility name and sometimes parent company name which I will use to identify the companies in the dataset. There is also information on state and industry for each facility, which I will use to compare companies that join the UNGC (the treatment group) and companies that do not join the UNGC (the control group). There is also a variable stating whether the facility is federal or not, which will be used later in the analysis.

There are variables indicating whether the chemical category is PBT, Dioxin, PFAS or other, and whether this chemical is regulated under the Clean Air Act. PBT, Dioxin and PFAS are chemical categories of higher toxicity than other chemicals in the dataset, found to have severe adverse human health effects, and I will be using these chemical classifications in the analysis. The variable stating if the chemical is regulated under the Clean Air Act will be used in part of the analysis to address the concern of variation due to regulation and not voluntary action. There are variables measuring chemical amounts in pounds, such as total recycled, total treated, total energy recovery and total releases. Total recycled is the amount of chemicals that were recycled, on and off-site. Total treated is the amount of chemicals that were treated off and on-site. Total energy recovery is how much of the non-recyclable waste is converted to usable energy. The variable for total releases is the chosen variable of interest and it is the sum of chemicals released to air, water and earth.

3.2.1 Descriptive statistics

The TRI raw data spans the years 1987 to 2021, with 2,977,192 observations in total. The observations are on facility and chemical level. To be able to perform the analysis, I must aggregate the observations to company level. When aggregated to company level, there is data for all years, but as I combine the observations from different facilities and of different

chemicals, there are now only 288,063 observations. Companies that join the UNGC within the time window for the dataset (meaning 2021 at the latest) will construct my treatment group "Treat" and all other companies are the control group "Control". Table 2 provides an overview to compare the sizes of the two groups.

Table 2 - Overview of dataset

	Sample	Control	Treat
TRI raw data			
Number of observations	2,977,192	2,742,465	234,727
Average total releases	54,689	53,086	73,417
Average number of facilities per company	10.9	10.5	24.4
Aggregated to company level			
Observations	388,063	383,964	4,099
Companies	28,270	28,151	119
Annual average total releases	419,571	379,168	4,204,212

We see here that companies in the treatment group in total constitute less than 10% of the total observations in the raw dataset, and only 1% of the observations when aggregated to company level.

Facilities in the treatment group seem to have a significantly higher average total releases, 73,417 pounds compared to 53,086 pounds in the control group. There are several possible explanations for this. It could be that companies who release more are more likely to join voluntary programs; Vidovic and Khanna (2007) found that EPA's 33/50 program attracted the most polluting firms. It could also be a result of the larger number of observations in the control group giving the large observations less weight, thus lowering the average. When the observations from the raw dataset have been aggregated to company level, summing up all chemicals and facilities per company, it gives much higher numbers. A company in the control group releases 379,168 pounds on average yearly and a company in the treatment group releases over four million pounds a year on average. A company in the control group has 10.9 facilities on average, whereas in the treatment group, a company has 24.4 facilities on average. This explains why when aggregated to company level, the difference between

the two groups in average releases increases. The companies in the treatment group are clearly larger companies, with more facilities and more chemical releases on average.

Using the raw TRI dataset, it can be useful to get an overview of what types of chemicals make up the dataset, and how do the companies in the treatment and control group compare in certain characteristics. Table 3 divides the chemical releases on three aspects, how they are classified (PBT, Dioxin or other TRI chemical), whether they are carcinogens (chemicals that cause cancer), and if they are regulated by the Clean Air Act. The mean total releases are on facility- & chemical-level.

Table 3 - Chemical composition

FROM FULL DATASET TRI, COMPARING TREATMENT AND CONTROL GROUP												
	N (observations)			Percentage of observations			Mean totalreleases			Percentage of totalreleases		
	Sample	Control	Treat	Sample	Control	Treat	Sample	Control	Treat	Sample	Control	Treat
CLASSIFICATION												
Dioxin	24,030	22,860	1,170	0.81%	0.83%	0.50%	0.18	0.13	1.15	0.00%	0.00%	0.00%
PBT	323,604	306,393	17,211	10.87%	11.17%	7.33%	43,582	44,171	33,112	8.65%	9.25%	3.31%
TRI	2,629,558	2,413,212	216,346	88.32%	87.99%	92.17%	56,556	54,721	77,021	91.41%	90.41%	97.09%
CARCINOGEN												
No	2,255,573	2,071,140	184,433	75.76%	75.52%	78.57%	67,661	65,832	88,202	93.87%	93.15%	94.77%
Yes	721,619	671,325	50,294	24.24%	24.48%	21.43%	14,141	13,762	19,202	6.26%	6.33%	5.62%
CLEAN AIR ACT CHEMICAL												
No	1,146,537	1,053,574	92,963	38.51%	38.42%	39.60%	76,981	74,400	88,202	54.17%	53.70%	57.44%
Yes	1,830,655	1,688,891	141,764	61.49%	61.58%	60.40%	40,727	39,790	19,202	45.77%	46.03%	42.79%
Totalreleases	2,977,192	2,742,465	234,727	100%	92%	8%	54,689	53,086	73,417	100%	90%	11%

From this graph, it seems the treatment group and the control group are somewhat similar in their chemical compositions. The largest difference being that facilities in the treatment group seem to have a smaller proportion of PBT releases, 3% versus 9 % in the control group. Note that although carcinogens are almost a fourth of the observations (for facilities in both groups), they only constitute around 6% of the total releases. Chemicals regulated under the Clean Air Act make up over 60% of the raw observations, but they make up less than half of the total releases. Dioxin constitutes a very small part of the dataset, less than 1% of the observations, and with very small values, nearly 0% of the total releases. The treatment group constitutes 8% of the observations, but 11% of the total releases.

Table 4 compares the control and treatment group on how the chemicals are handled, either released to air, water or land, or managed through energy recovery, recycling or treatment.

Table 4 - Chemical management

	Sample	Control	Treat
	Average in pounds of chemical		
Released:			
Air	16,664	16,113	23,102
Water	5,744	5,578	7,680
Underground	3,024	2,630	7,632
Landfills	612	581	977
Managed:			
Energy Recovery	116,170	106,431	230,021
Recycling	32,291	29,948	59,677
Treatment	89,018	84,518	141,889

If converted to percentages of total releases, the control group and treatment group only differ significantly for chemicals released in how much of the releases are underground; for the control group, this is 10% and for the treatment group this is 5%. If I convert how the chemicals are managed to percentages, the two groups look similar as well.

Next, I study where the facilities are located in the United States. When looking at how many facilities are in which states as a percentage of total facilities, the treatment group and the control group are quite similar. The three states with the highest share of facilities are California, Ohio and Texas, and are presented in the table below.

Table 5 - State composition of facilities

	Sample		Control		Treat	
State	Facilities	Percentage	Facilities	Percentage	Facilities	Percentage
California	4,524	7.75%	4,370	7.83%	154	6.05%
Ohio	3,446	5.90%	3,282	5.88%	164	6.44%
Texas	4021	6.89%	3809	6.82%	212	8.32%

The largest difference between the two groups is for California. California has almost 8% of all facilities in the control group, but 6% of the facilities in the treatment group. This is not a

large difference, and so the two groups can be deemed similar in terms of where they are located. What is interesting to note however is the disparity between the treatment group and the control group in terms of how much of the chemicals are released in each state. Table 6 includes selected states which present interesting findings with percentages of the total releases from facilities in these states.

Table 6 - State composition of total releases

State	Sample	Control	Treat
Alaska	10.67%	11.92%	0.00%
Arizona	3.58%	2.21%	15.12%
California	4.61%	5.11%	0.37%
Texas	7.79%	6.75%	16.80%

The treatment group and the control group have some differences in terms of the amounts they release in each state, the table above shows the states with the largest differences between the groups. Comparing the percentage of facilities and percentage of total releases in these states show some interesting details. Alaska doesn't even have half a percent of all the facilities, but the state has almost 12% of the total chemical releases in the control group. In California, the companies in the treatment group have 6% of their facilities, but only 0.4% of the total releases from the treatment group occurs in this state. In Arizona, the treatment group has only 1.5% of their facilities, but 15% of their total releases. In Texas the treatment group has 8% of their facilities, but 17% of their total chemical releases. These findings could possibly be due to local regulations in the states, or different consumer demographics. It is easy to imagine that in California there is a higher prevalence of green consumers, so facilities that operate in this state and join the initiative do not wish to risk being labelled greenwashers, and therefore release relatively less here. Alaska, Arizona and Texas are all conservative states, and are often less concerned about environmental issues. However, I do not, and cannot, know why I find differences between the treatment group and the control group in terms of how much they release in each state. It could have something to do with selection bias. I cannot control for geographic location in the analysis

as many companies operate in multiple states, but it can be of use to be mindful of this when analysing the results and evaluating the validity of the findings.

Finally, I compare the treatment group and the control group based on which industries are responsible for the chemical releases. It is important to understand what companies exist in the dataset, both to know what kind of industries release chemicals, and in terms of who are required to report to the TRI. The numbers in the table below are percentages of total releases from an industry, and I have excluded industries with very low shares.

Table 7 - Industry composition of total releases in raw data

Industry Sector	Sample	Control	Treat
Metal Mining	25.52%	26.10%	20.52%
Chemicals	20.80%	17.74%	46.63%
Primary Metals	12.58%	12.95%	9.07%
Electric Utilities	10.86%	12.05%	0.12%
Paper	6.56%	7.19%	1.38%
Other	4.29%	4.71%	0.77%
Hazardous Waste	2.87%	3.21%	0.00%
Food	2.51%	1.82%	8.37%
Petroleum	2.43%	2.66%	0.45%
Transportation Equipment	2.16%	1.73%	5.80%
Plastics and Rubber	1.94%	2.03%	1.21%

In the control group, the largest contributor for chemical releases is Metal Mining, constituting 26% of the total releases in this group. The largest share of total releases in the treatment group belongs to Chemicals, which makes up almost 50% of all the releases in this group. Chemicals are the next largest contributor for the control group, at 18%. Electric Utilities stand for 10% of the releases in the control group, but not even half a percent in the treatment group. These differences might be related to a possible selection bias, asking the question of who joins the UNGC and do they differ significantly from other companies. There is interesting research showing that proximity to the final consumer in the value chain can affect how well a voluntary environmental program works (Li & Wu, 2020), so the type of

industry could possibly affect the estimated effect. For example, the food industry holds 16% of the facilities in the treatment group but constitutes only 8% of the releases, and in the control group it has 5% of the facilities and 2% of the releases. Could there be self-selection into the treatment group based on the industry's proximity to the final consumer? I cannot control for industries in the analysis, as many companies operate in multiple industry sectors, but it can be worthwhile to remember the differences between the two groups in terms of industry composition, and how this can affect the results.

In conclusion to the descriptive statistics, the treatment group and the control group are similar in most of the provided metrics but differ in some. Most noteworthy is the size, companies in the treatment group seem to be larger companies with more facilities per company. Next, the differences in industry composition could provide a challenge in terms of selection bias. If differences are constant over time, I should still be able to estimate the effect of joining the initiative, but the information will be considered when analysing the results.

3.3 Zeroes in the dataset

In the raw dataset and the company-level aggregated dataset, there are many observations with total releases equal to zero. This is a known problem with the TRI dataset, as pointed out by Khanna (2019), without a clear solution. Of the 2,977,192 observations in the raw dataset where observations are at facility and chemical detail level, 634,193 of them report zero total releases (50,681 of these observations are from the treated group), and this amounts to 21% of the raw observations (for control and treatment group). In the dataset where observations are aggregated to company level, there are 64,374 observations with zero total releases, constituting 16,6% of the dataset. 168 of these observations are in the treatment group, meaning only 4% of the observations in this group report zero total releases. For the control group it is 16.7% of the observations. This is one reason for why the average of the control group is lower than the treatment group.

On closer inspection of these zeroes, it is important to note that zero total releases does not mean that the company does not use or produce chemicals, but that they do not release any

directly to air, water, land or underground deposits. The chemicals can be treated, recycled or used for energy recovery onsite or sent offsite. Total production is the sum of total releases, total quantity recycled, total quantity treated, and total energy recovery, off and on-site (Khanna, 2019). There are however some companies that report zeroes for all these variables, meaning that they are claiming zero total production of chemicals. In the raw dataset 431,042 of the observations report zero for all variables measuring chemicals, meaning there is zero total production. The question then is why on earth they are reporting to the TRI. Companies that are in an industry covered by the TRI, have 10 or more full-time employees and who manufactures, processes or otherwise uses a TRI-listed chemical in quantities larger than the given thresholds, must report to the TRI. The companies must fulfil *all three criteria*, meaning there should be no observations with zero total production of chemicals.

Khanna (2019) recommends deleting all zeroes, but this might remove some information of interest. For example, if a company is not releasing anything one year, but all chemicals are being treated or recycled, this would be an improvement for the environment if they previously would release some of them. This would be variation that is of interest to the study. It is less clear to me what should be done about the companies that report zero production. As I cannot know what the underlying mechanisms are and how to interpret these zeroes, I have chosen to not exclude them in the analysis.

To address the issue of zeroes, robustness checks will be run using two methods. One is to leave out the observations that report zero total releases. The next is to exclude companies that always report zero total releases.

3.4 Other challenges

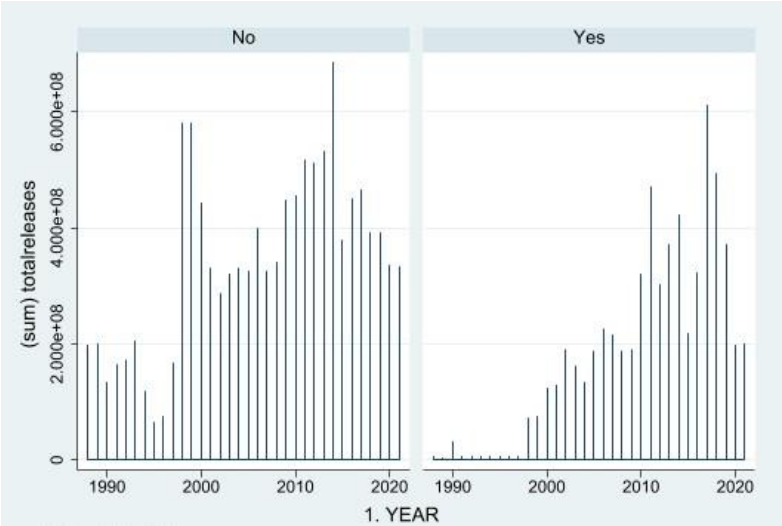
A challenge when preparing the dataset is properly aggregating observations to company level, as companies may change or misspell their names, resulting in a misleading lower average. I will address this in the paragraph about preparing the dataset. Another issue with the dataset is that some companies disappear over time. Here we cannot know if they simply closed up shop, if they changed chemicals to ones not required to report or if their

releases are now below reporting levels. A company that joined the initiative might have been so successful in their pursuit of lowering chemical releases that they disappear from the dataset, and this improvement is lost to the analysis.

The TRI data is self-reported, which can cause concern about its accuracy. The EPA runs quality controls on a regular basis to assure accuracy, but it is limited how well they can control the provided information. Companies would have incentives to report lower emissions, for example due to public scrutiny and if it cannot be properly monitored, it is possible that they might be reporting lower emissions than actual emissions. In a voluntary program from the US Department of Energy, researchers found that non-participants in the program decreased their emissions over time, but participants in the program actually increased emissions over time while reporting reductions (Kim & Lyon, 2011).

Changes in the reporting requirements create a challenge for the study. Variation in total releases in the dataset as a result from changing reporting requirements is of no interest in this study. These changes can explain some fluctuations in the dataset, the largest effect seen is from the changes in reporting requirements for PBT chemicals in 1998. Seven new PBT chemicals were added and all PBT chemicals already on the list got a lower reporting threshold of only 100 pounds. As we can see from the graph below, total releases from chemicals classified as PBT's increase from this point. ("No" = chemicals that are not PBT, "YES" = chemicals that are PBT).

Figure 2 - Sum total releases over time divided by PBT-category



To correct for the changes in reporting requirements, articles such as Bui and Kapon (2012) choose to only look at a time window and chemicals where the reporting requirements have been constant. This is not possible for this thesis, I need a large time window as companies join the program over a span of 20 years, and I also need years prior to compare. Limiting to industries and chemicals where the reporting requirements have been stable could remove useful information and the treatment group is already quite small. I will continue with all chemicals and all years but be mindful of the changes when analysing the data and interpreting results. Changes in reporting requirements would likely affect the treatment group and the control group similarly, therefore I do not consider these a large concern.

3.5 Preparing the dataset

To make the dataset ready for analysis, some cleaning measures must be implemented. Observations with Dioxin are reported in grams, so I convert these to pounds. As I wish to look at the total releases at company level and not facility level, I create a new variable called "name" which gathers information first from the standard parent company name, a TRI standardized variable for the parent company name. If this variable does not contain any information the "name"-variable will gather information from the parent company name as reported by the facility. If the observation does not report the parent company name, the "name"-variable gathers the information from the facility name. Some of these names have information in parenthesis, such as COMPANY A (formerly COMPANY AB). I remove the parenthesis and the information in it to make the names comparable and so that I can use the Stata command "reclink" later, which does not work if there are unfinished parentheses in the dataset. To clean the names further, I also delete special characters, leading or trailing spaces, and terms such as incorporated or INC/CORP. To match the information from the participant list to the master dataset from TRI, I need to use fuzzy matching, and I have chosen the "reclink" command in Stata. This algorithm seems to give the most reliable name matching results, and with fewer errors than the "matchit" command. The fuzzy matching is necessary as I do not have any company ID's in the UNGC list, and names can sometimes be written in slightly different ways, f.ex. WATER HOUSE vs WATERHOUSE. Even using 99% as a demand for matching degree, I will still get some erroneous matches, and I correct these

manually in a compressed list and then merge it back with the full TRI panel data. 176 companies in the TRI raw dataset are matched with companies that have joined the UNGC, and 119 of these joined in the years included in the TRI dataset, thus creating our treatment group (the treatment group consists of companies where I have any observations from the company after it joined the UNGC). A known weakness with the dataset is that due to misspellings, changes or differences in the written company name that go beyond the cleaning measures implemented, total releases for a company might be separated into two or more entities. This would drive average total releases down. As the dataset is so large, I cannot manually check and correct these errors, and the coding for this is beyond the scope of this paper, but I make note of the error and keep it in mind when analysing the results. When merged and collapsed this creates a simple dataset aggregated to company level, the Full Collapsed Dataset, with year, name, total releases, and year when the company joined the UNGC. I create one dummy variable which equals 1 when the company is in the treatment group, and 0 otherwise. I then create another dummy variable, active_treat, which equals 1 when the company is in the treatment group and the year is in or after year joined. This means that it will equal 0 for companies in the treatment group in years prior to participation, and it will always equal 0 for the control group. This variable will estimate the effect of participation in the UNGC on total chemical releases.

4 Method

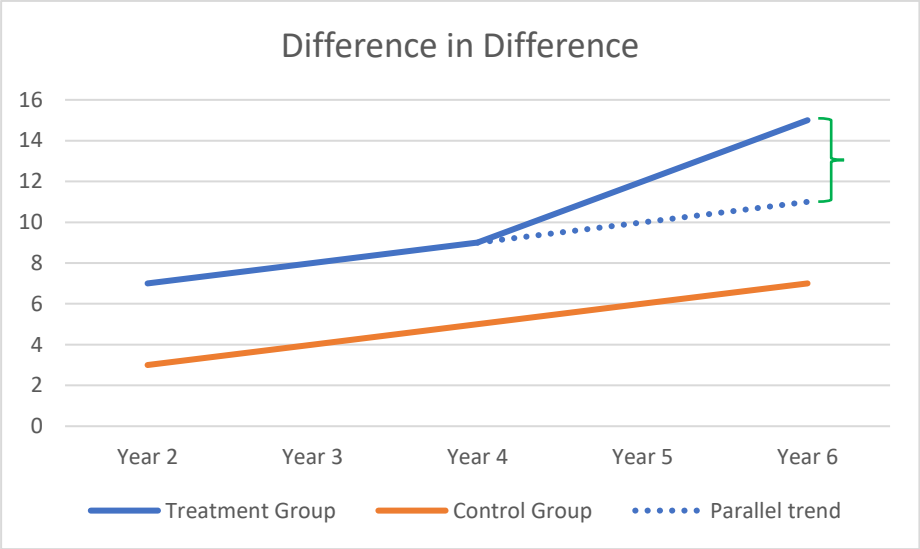
To answer the research question, I will use a method called difference in difference (DiD). In a natural experiment there should ideally be as-if randomised treatment allocation. As companies join voluntarily, I must consider that the study will likely suffer from self-selection issues. To attack the large amount of quantitative data, I turn to the power of econometrics.

4.1 Difference in Difference

The simplest explanation of a Difference in Difference model is a graphical one. The figure below shows an instance where treatment occurs in year four, and one can see that the trajectory for the treatment group changes from this point. The dotted blue line indicates

the expected trajectory should the treatment not occur. Instead of merely looking to the difference between the control group and the treatment group in year six to estimate the effect, we look to the change in the differences between these two groups, as indicated by the green bracket.

Figure 3 - Difference in Difference graphic explanation



The Difference in Difference (DiD) method adjusts for differences between the treatment group and the control group. It compares the changes in outcomes before and after treatment, and so controls for the differences in total releases for the two groups before treatment (Stock & Watson, 2020). This method is widely used in economics and is suitable for the question at hand as it is commonly used in impact evaluation studies. Instrument variable regression is also often used, but I cannot use this method as I do not have a suitable instrument in this dataset. As the treatment, joining the UNGC, occurs at different times for different companies, I must use Staggered Difference in Difference. I will not be including control variables beyond time and firm fixed effects, as there are no suitable ones available in the dataset. I would've preferred to control for state and industry, but several companies operate in multiple states and industries. Time fixed effects will control for variables that vary over time but are constant across companies, and firm fixed effects will control for variables that vary across companies but are constant over time. Possible omitted variable bias will thus only occur from variables that vary across both time and firms (Stock &

Watson, 2020). In natural experiments, we want the treatment to be as-if randomly assigned, but there is often selection bias. In this case, companies join voluntarily, meaning they self-select into the treatment group. Therefore, we might suffer from selection bias as the companies that choose to join the initiative might have characteristics that differ from that of other companies not choosing to join. In the DiD method the two groups do not need to be the same as long as the differences between them are constant. There are two main assumptions for the DiD method, the Parallel Trend Assumption and the Stable Unit Treatment Value Assumption (SUTVA). I will address each of these assumptions individually.

4.2 Stable Unit Treatment Value Assumption

The Stable Unit Treatment Value Assumption (SUTVA) requires there to be no spillover effects, meaning that the treatment status of one company should not affect the outcome of another company (Bloome, 2009). Some existing literature (Zhou et al., 2020) find spillover effects for voluntary environmental programs and argue that the reason why researchers often find voluntary programs to be ineffective is simply due to neglected spillover effects. If the UNGC creates spillover effects for companies that do not join the initiative, this assumption will not hold. To assess this, it is necessary to look at spillovers geographically or within industries, but as large companies often operate in several states and industries, controlling for such spillover effects is beyond the scope of this paper. It is also not evident that this program would create spillover effects as the companies that join are spread out in terms of geography and industry. I will assume that the assumption holds as I do not have proof otherwise for the UNGC in particular, but I will be mindful of any evidence suggesting the assumption might not hold.

4.3 Parallel trends assumption

The parallel trends assumption states that in the absence of treatment, the control group and the treatment group should have similar trends. This means that the differences between them should be constant and that they would've developed in parallel if no treatment ever occurred. As we cannot measure what would've happened without

treatment, unless we could find a way to visit parallel universes, we can look to the time before treatment occurs. If the two groups have parallel trends in the time before treatment, we assume that this would have continued without treatment. There is no test for the parallel trend assumption, but there are some possible methods to assess if there seems to be parallel trends before treatment (Cunningham, 2021).

4.4 Natural logarithm of variable of interest

In the analysis, one of the models will be using the natural logarithm of total releases as the variable of interest. This specification has the advantage that coefficients can be interpreted in terms of percentages and can also address issues with outliers. In this model the parallel trends assumption assumes a parallel trend in percentage changes.

4.5 Event Study

In the thesis I will also use a method called Event Study. This design is an extension of the standard DiD model and has two benefits. The first is that the model provides some information that can be used to assess if there are parallel trends before treatment. The second is that this model allows for dynamic effects from the treatment. The difference in difference estimator measures the average effect for the whole timespan post treatment, and the event study shows the cumulative effect for each time period after treatment. It is easy to imagine that the effect from joining the initiative would change over time; a company could get progressively better at managing their releases, not only improve (or get worse) in a single step after joining the initiative, but implementing changes gradually, learning and correcting over time. In the model the time variable is standardized. The time when the treatment occurs becomes event time 0, the year prior becomes -1 and the year after 1. This means we also estimate coefficients for the time points before treatment. These coefficients can then be used to assess parallel trends; they should either be zero or statistically insignificant, indicating that the control group and the treatment group have trends that are not significantly different from one another.

5 Analysis

I have chosen an exploratory approach for this thesis, and I will be using several models and several subsets of the data to examine the relationship between joining the UNGC and chemical releases. There is not one model that can be assumed to best represent the relationship, the exploratory approach allows me to investigate several possibilities, theorise over the possible mechanisms, and address specific concerns. Throughout the analysis I will use cluster-robust standard errors to take into account the probable issues of heteroskedasticity and serial correlation. A company's chemical releases one year are likely to be correlated with the previous year's releases, and a quick Breusch-Pagan test confirms my suspicion of heteroskedasticity.

Stata Output 1 – Breusch-Pagan test

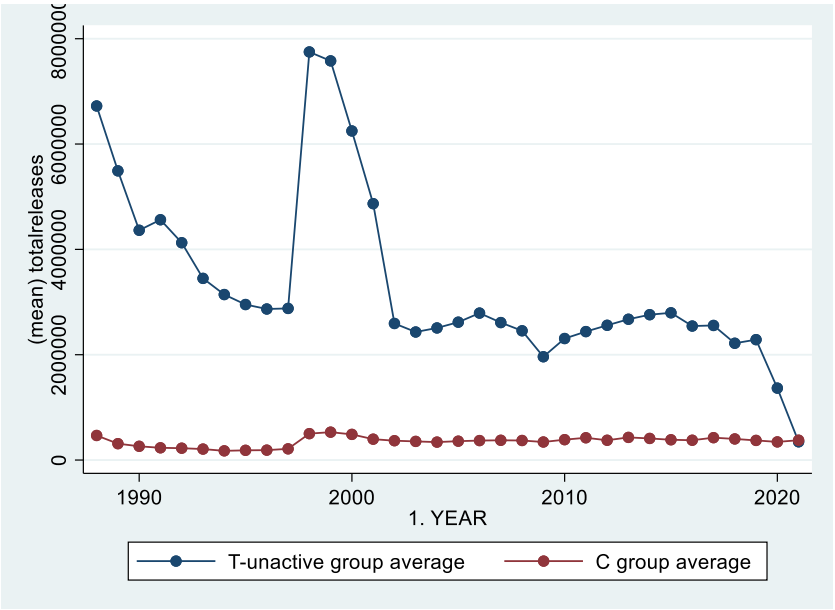
```
Breusch–Pagan/Cook–Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of totalreleases

H0: Constant variance

    chi2(1) = 153.95
Prob > chi2 = 0.0000
```

Since the p-value is less than 0.05, I can reject the null hypothesis which states that there is constant variance among the residuals and conclude that heteroscedasticity is present in the data.

5.1 Parallel Trends

To examine if there are parallel trends prior to treatment, I discard observations that are from the treatment group after the treatment has occurred and make a graph with mean total releases for each group over time (Schechter, 2020). Due to a large outlier in 1987, I have excluded this year in the graph to better see the variations. This gives the following graph in figure 4.

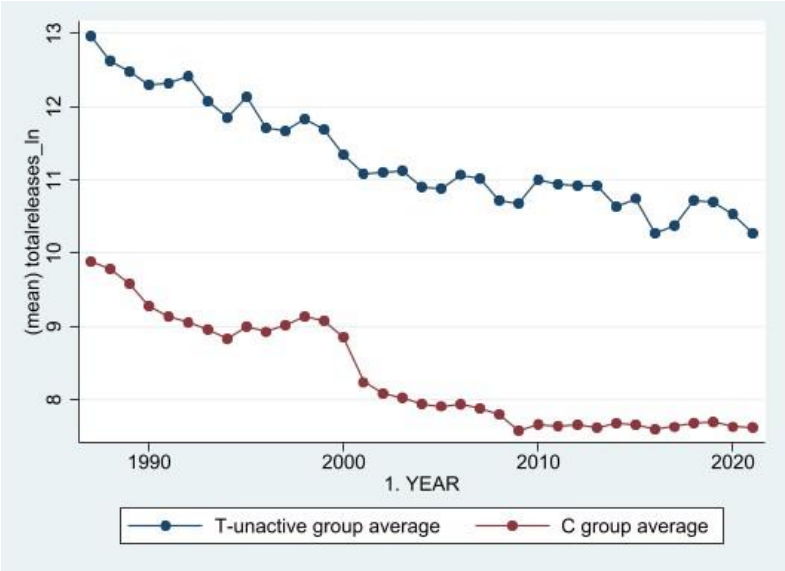
Figure 4 - Average total releases over time



On the x-axis we have years, and on the y-axis is the average total releases. The control group is represented by the red line, and the treatment group is represented by the blue line. At first glimpse, these two lines do not look strictly parallel. But they do not either look alarmingly un-parallel. The line for the treatment group is a lot more volatile, possibly due to the lower number of observations. With many observations the variations would be evened out, explaining the steady line for the control group in contrast. Both lines go up in 1998, probably as a result of the changes in reporting requirements that year for PBT chemicals, and this indicates that the two groups had a similar reaction to this external change. This method has many limitations in assessing the parallel trends, one of them being the small number of companies in the treatment group compared to the large number of companies in the control group. As I remove the observations for companies in the treatment group in years after joining the initiative, there remain even fewer observations for these years to calculate an average. It is likely that the two lines would be somewhat different as the companies in the treatment group seem to be larger, but this is fine as long as the differences are constant over time. The graph does not give a conclusive answer to the parallel trends assumption, but as it does not clearly dispute that the assumption holds, I will continue as if it holds, noting any evidence to the contrary.

One model in the analysis will be using the natural logarithm of total releases as the dependent variable. In this model the parallel trends assumption would mean that the parallel trends would be in terms of percentage changes in total releases, not in the actual release levels, but in changes from year to year. Below is a graph using the same method as above to construct a visual aid to assess parallel trends when using the natural logarithm of total releases.

Figure 5 - Average natural logarithm of total releases over time

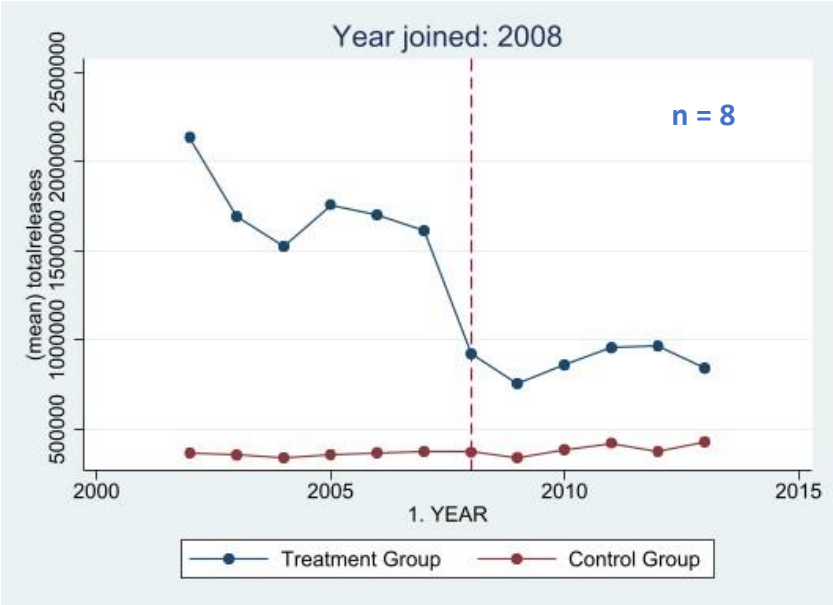


Using the natural logarithm of total releases compresses the variations in the graph, and both lines seem to have a clear downwards trend. They are not completely parallel, and some information, like the reaction to the changed reporting requirements in 1998 is less clear in this graph. The method of using such a graph to determine parallel trends is not perfect, so I cannot conclude anything on whether the assumption holds when using the natural logarithm of total releases as the variable of interest. However, it does not prove that the assumption does not hold, and I will continue. It is worth noting that in terms of parallel trends, the assumption cannot hold for both the natural logarithm of total releases and real total releases in pounds. Either the trends are parallel in real values or percentage changes, or none of the two.

5.2 Graphical examination

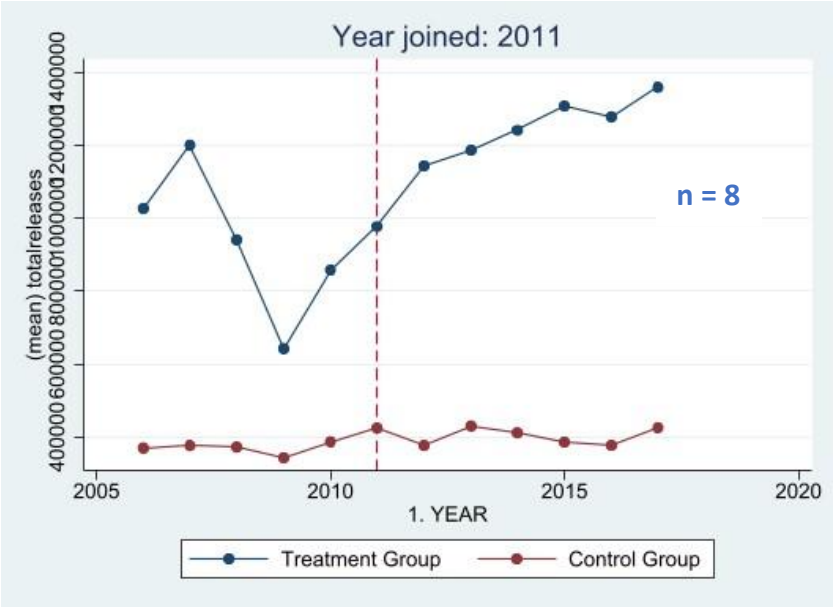
To look for parallel trends and to check if there is a graphically visible treatment effect, I divide the treatment group into smaller groups depending on which year they joined the initiative. Then I make a graph for each of these groups comparing them to the control group of companies that never join the UNGC. The red line represents the control group, and the blue line is the treatment group. The red vertical dashed line indicates the year that the treated companies joined the initiative. In an ideal world I would see a clear change in the trend for the treated companies from the treatment point, and preferably a similar trend between the control and treated group before treatment. Note that the x-axis showing years have different ranges, this is so that I can "zoom" in on a time window around the treatment event. Dividing the treatment group into groups depending on year joined gives some very small groups, while the control group consists of 28,094 companies in each instance. I have only included four of the graphs with similar numbers of companies in the treatment group to compare. The rest can be found in the appendix. The number of companies in the treatment group for each graph will be marked by n in blue.

Figure 6 - Average total releases with treatment year 2008



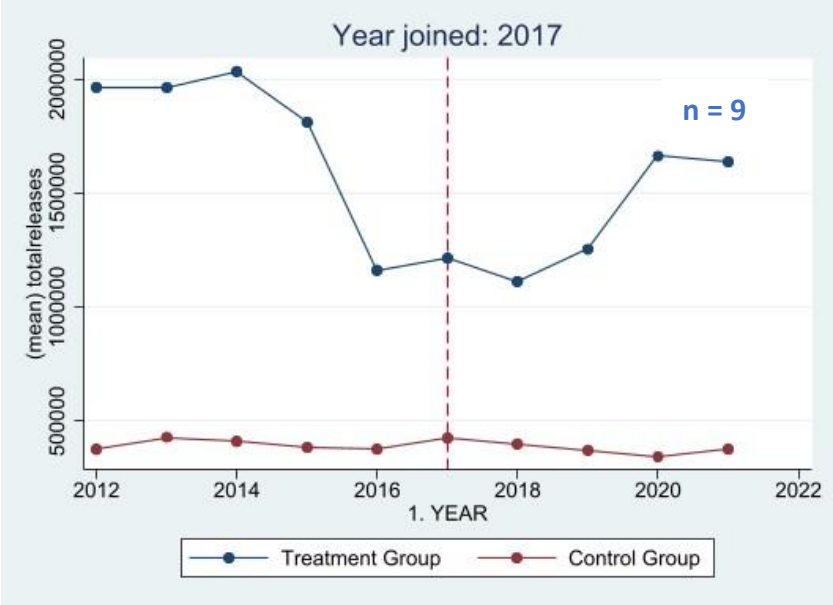
This graph looks promising with lower average releases after treatment in 2008. There is not a clear parallel trend before the treatment.

Figure 7 - Average total releases with treatment year 2011



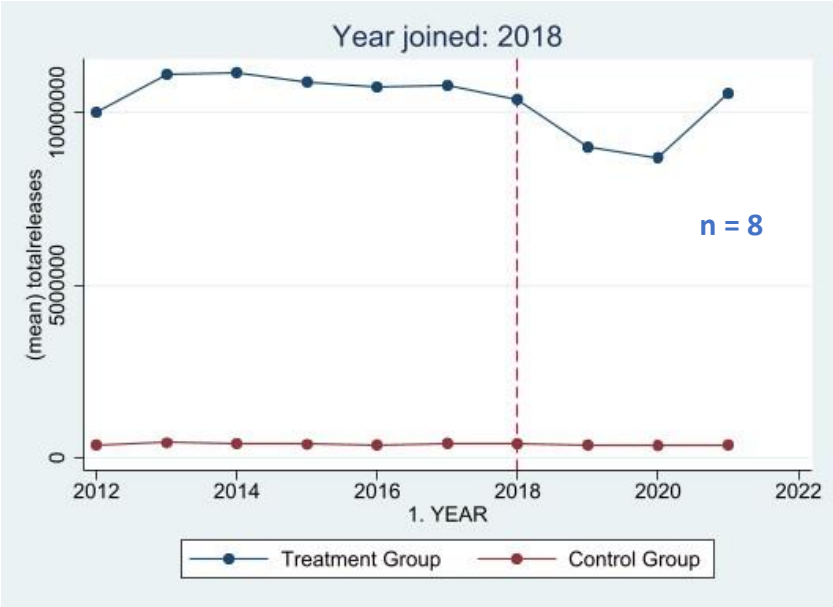
In this graph there is an increase in average total releases after treatment in 2011. There does not appear to be parallel trends before either.

Figure 8 - Average total releases with treatment year 2017



In this graph there is a reduction occurring before treatment, again after treatment, and with a following increase leading up to a level lower than a few years before treatment.

Figure 9 - Average total releases with treatment year 2018



A promising graph, where the groups seem quite parallel before treatment, and the treatment group has a reduction in average total releases after treatment. The increase in year 2021 could possibly be a "return to normal", sometimes the effects of treatment disappear over time.

The four graphs above are inconclusive. As the number of companies in each of the year treatment groups are so small, I cannot use these graphs to reliably conclude if there is a treatment effect or whether there are parallel trends before treatment. Luckily, I can use econometrics to find associations invisible to the naked eye.

5.3 Choosing control group

Choosing the right control group is important, and the goal is to have a control group that is as similar to the treatment group as possible. In the first part of the thesis, the control group will be all possible companies in the TRI dataset. This takes all available information into account, and I do not have to worry about possibly introducing new error sources due to not knowing what is chosen away. There is, however, concern that this control group is not adequately similar to the treatment group. Therefore, I will also perform the analysis with two different definitions of the control group, one where the control group is restricted

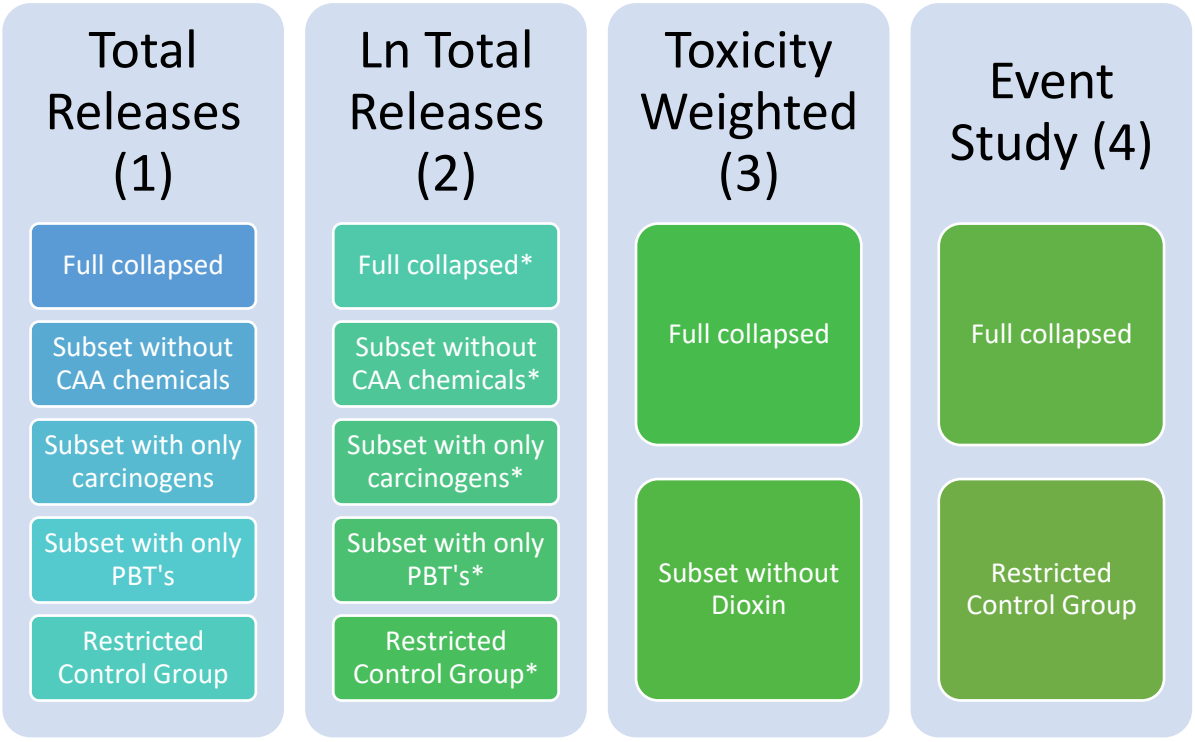
based on specific criteria and one where the control group is made up of companies that join the UNGC after 2021.

Other studies may use different approaches to construct a suitable control group, for example using financial information such as number of employees and annual sales revenue to match companies in the control group to the companies in the treatment group in terms of size (Li & Wu, 2020). It would've been ideal to construct a control group using such metrics, but it will not be possible for this paper due to time constraints as this information is not in the chosen dataset.

5.4 Regression models:

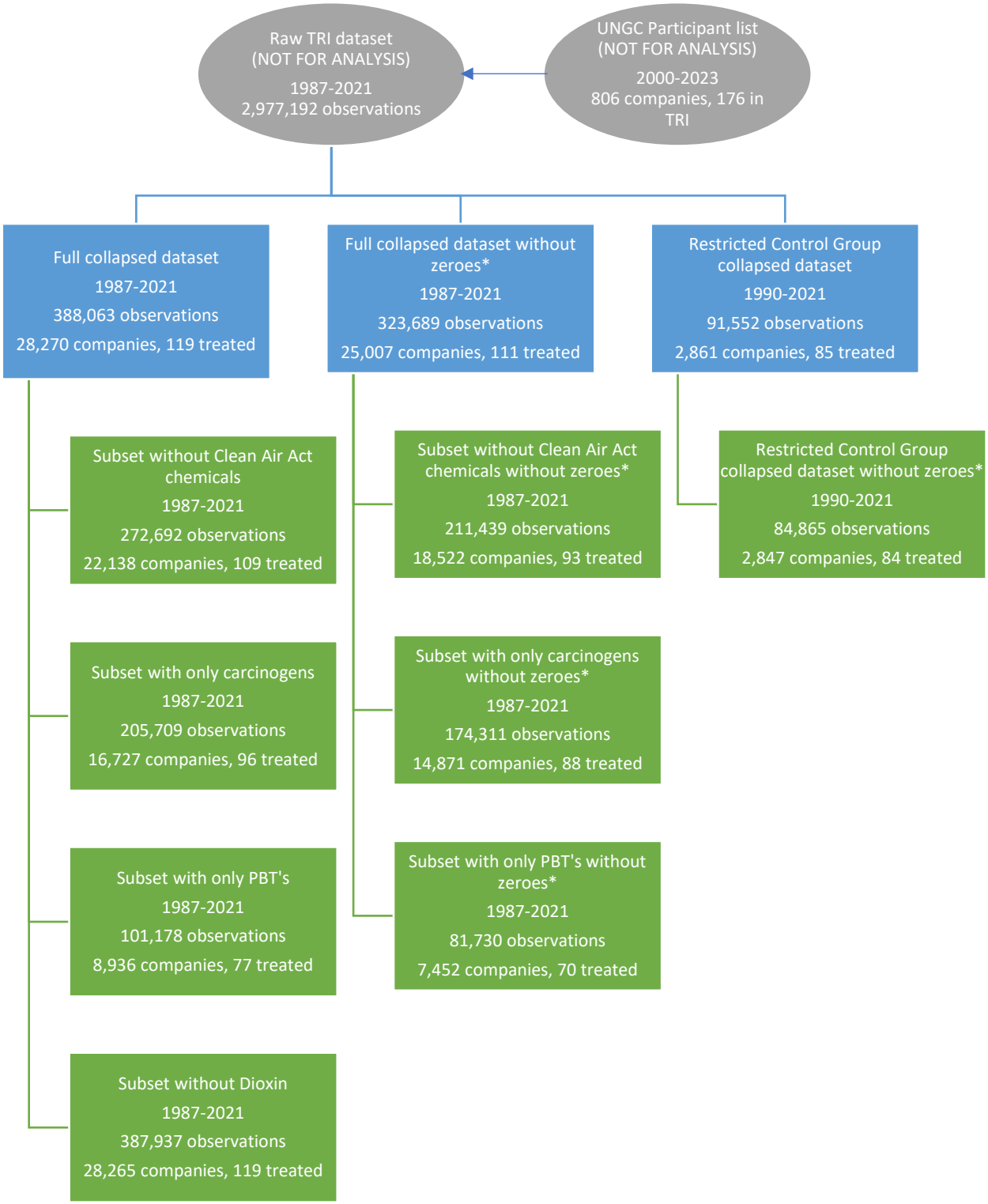
I will be using four different models, the first three models are standard difference in difference estimators, and the fourth is an event study. Figure 10 below shows an overview of the four models and the coloured boxes indicate which datasets will be used for each model. Figure 11 provides an overview of the datasets and sub-datasets used for analysis. The main model, Model (1), measures the outcome in total chemical releases in pounds. Model (2) measures the outcome with the natural logarithm of total releases, looking at percentage change in total releases. I would have preferred to use a model that takes impact of chemical releases into consideration, but I do not have a satisfactory approach for such a model, and instead use an imperfect toxicity-weighted model, Model (3), to address this. Model (4) is an event study, a method that allows for dynamic changes in total releases. I have chosen Model (1) as my main model as this estimates an effect in real values of chemical releases in pounds, which in my opinion reflects the real-world implications of the effect in the simplest possible way. This is also how companies report to the TRI, in pounds of chemicals released.

Figure 10 - Overview of Models



*Without observations with zero total releases.

Figure 11 - Overview of datasets for analysis



*Without observations with zero total releases

5.4.1 Total Releases (1):

In Model (1), the unit for the variable of interest is the aggregated total chemical releases in pounds. The mathematical model is as follows:

$$totalreleases_{it} = \beta_0 + \beta_1 * active\_treat_{it} + \alpha_i + \delta_t + u_{it}$$

The dependent variable *totalreleases* is total pounds of chemical releases for company *i* in year *t*. The independent variable, *active_treat,* is a dummy variable that equals 1 for companies in the treatment group for years after joining the initiative and 0 otherwise. It is this variable's coefficient, β_1, that measures the effect of the treatment, treatment here meaning joining the UNGC. u_{it} is a hopefully well-behaved error term with conditional mean of 0. α is the firm fixed effects for firm *i*, and δ is time fixed effects for time *t*.

5.4.1.1 Full collapsed

I will begin with performing the regression for Model (1) on the full collapsed dataset. This gives the following output:

Stata Output 2 – Model (1) on full collapsed dataset

```
Fixed-effects (within) regression               Number of obs      =       388,063
Group variable: name_n                          Number of groups   =        28,270

R-squared:                                      Obs per group:
     Within  = 0.0010                                         min =             1
     Between = 0.0003                                         avg =          13.7
     Overall = 0.0002                                         max =            35

                                                F(35,28269)        =          4.08
corr(u_i, Xb) = -0.0277                         Prob > F           =        0.0000

                              (Std. err. adjusted for 28,270 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
totalreleases|  Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat|   -3267673    1120205    -2.92   0.004     -5463328    -1072018
```

The result indicates that joining the UNGC is associated with a reduction in total releases of 3.3 million pounds, a large effect considering the average annual total chemical releases in the treatment group is 4 million. This effect is statistically significant at the 1% level. This does not necessarily mean that joining the initiative causes this reduction, but that the

reduction and participation are correlated. The R-squared is quite small but this is to be expected when not including any control variables, as there are probably many factors that affect total releases, and I will not give this metric much attention as we continue. This first result is a bit surprising, showing a quite large and statistically significant effect from joining the initiative. This is not what I would've expected to find, and there are multiple questions that arise from such a finding. Do the assumptions for the model hold? Is the estimated effect spurious, meaning there is a third factor affecting the dependent and independent variable? If the effect is real, which way does causation run, do companies release less because they join or do they join because they release less?

To further examine the effects of joining the initiative I perform the Total Releases regression, Model (1), on several subsets of the data. This method using the same regression but for several subsets of the data is similar to the approach employed by Chintrakarn and Millimet (2006). This approach allows me to look at specific subsets and compare with the effect found above, and to look for possible explanations for the findings. Similar results could support the primary finding.

5.4.1.2 Subset without Clean Air Act chemicals

I begin by running the Model (1) regression on a subset of the dataset where I have removed all observations of chemicals regulated under the Clean Air Act. The Clean Air Act is the most influential regulation for chemicals, in the raw dataset approximately 60% of the observations are chemicals regulated under the Clean Air Act, and these observations constitute almost 50% of the total chemical releases. As these chemicals are regulated, variations in the releases of these might not be due to voluntary efforts. With this subset, excluding all observations of chemicals that are regulated by the Clean Air Act, the regression will estimate the effect of joining the UNGC on releases of chemicals that are not regulated by the Clean Air Act. This then removes the concern of variation due to regulation, not voluntary action.

Stata Output 3 – Model (1) on subset without clean air act chemicals

```
. xtreg totalreleases i.active_treat i.year, fe cluster(name_n)

Fixed-effects (within) regression               Number of obs     =    272,692
Group variable: name_n                          Number of groups  =     22,138

R-squared:                                      Obs per group:
     Within  = 0.0010                                         min =          1
     Between = 0.0001                                         avg =       12.3
     Overall = 0.0005                                         max =         35

                                                F(35,22137)       =       2.87
corr(u_i, Xb) = -0.0263                         Prob > F          =     0.0000

                              (Std. err. adjusted for 22,138 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
totalreleases|  Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat|   -2007204   983290.5    -2.04   0.041     -3934523   -79884.35
```

The outcome shows a negative coefficient that is statistically significant at the 5%-level, supporting the first finding. Joining the UNGC is here correlated with a reduction of 2 million pounds in toxic releases from chemicals not regulated by the Clean Air Act. This effect is somewhat smaller than the effect estimated above on the full dataset, but still a large effect. The effect is half the size of the average annual total releases, 4 million, in the treatment group, and this estimated effect is only from chemicals unregulated by the Clean Air Act which constitute a bit over 50% of the total chemical releases.

Even when attempting to remove variation that might come from regulatory changes, it seems like there is a large significant negative effect from joining the initiative on chemical releases. The initiative could also have an effect on regulated chemicals beyond the demands of the Clean Air Act, and therefore only this one subset will exclude these chemicals.

My next move is to look at categories of chemicals. Different chemical categories hold different levels of threat to human health and the environment. By looking at specific categories, I might learn more about the development for particularly toxic substances. The total aggregated releases does not consider the toxicity level and the level of impact, so the following part can possibly show where the effect comes from.

5.4.1.3 Subset with only carcinogens

For separating on chemical categories, I first make a subset of the dataset that only includes the chemicals that are known carcinogens. Carcinogens constitute almost a fourth of the raw data observations but only constitute around 6% of the total chemical releases, so this dataset is considerably smaller. I would hope that any company who joins a voluntary program would be concerned not only about the level of emissions, but also about how much they are releasing of chemicals known to cause cancer in humans. Running the Model (1) regression on this dataset produces the following outcome:

Stata Output 4 – Model (1) on subset with only carcinogens

```
. xtreg totalreleases i.active_treat i.year, fe cluster(name_n)

Fixed-effects (within) regression               Number of obs     =    205,709
Group variable: name_n                          Number of groups  =     16,727

R-squared:                                      Obs per group:
     Within  = 0.0083                                         min =          1
     Between = 0.0002                                         avg =       12.3
     Overall = 0.0012                                         max =         35

                                                F(35,16726)       =       5.66
corr(u_i, Xb) = -0.0568                         Prob > F          =     0.0000

                              (Std. err. adjusted for 16,727 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
totalreleases| Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat|  -302627.7   109372.3    -2.77   0.006      -517009   -88246.48
```

Again, this result supports the findings in the first and second analysis. There is a negative coefficient, statistically significant at the 1% level, indicating that there is a correlation between joining the UNGC and an average reduction of 300 thousand pounds of total releases of carcinogenic chemicals. It is not surprising that there should be an effect, as cancer is a topic that receives quite a lot of media and research attention. If a company wishes to improve their social standing by joining the initiative, then addressing areas that receive a lot of attention makes sense. The estimated effect is a lot smaller than the earlier estimates, but this effect is from chemicals that only constitute 6% of the total chemical releases, so it is still a large effect.

5.4.1.4 Subset with only PBT's

Secondly, I make a dataset that only contains the observations from chemicals that are of the category PBT. PBT's are particularly nasty chemicals that do not break down but accumulate in humans and nature and they are very toxic and cause a long list of adverse health effects. If companies are serious about reducing their negative impact, I would hopefully see a reduction in the releases of these chemicals in particular. Below is the result of the Model (1) regression run on this subset:

Stata Output 5 – Model (1) on subset with only PBT chemicals

```
Fixed-effects (within) regression               Number of obs      =       101,178
Group variable: name_n                          Number of groups   =         8,936

R-squared:                                      Obs per group:
     Within  = 0.0009                                         min =             1
     Between = 0.0000                                         avg =          11.3
     Overall = 0.0001                                         max =            35

                                                F(35,8935)         =          1.41
corr(u_i, Xb) = -0.0104                         Prob > F           =        0.0534

                                (Std. err. adjusted for 8,936 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
totalreleases|  Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat|   -37429.4   96239.41    -0.39   0.697    -226080.7    151221.9
```

The results show a negative coefficient, but now it is not statistically significant even at the 10%-level. One important thing to note before interpreting this result are the changes in reporting requirements for PBT chemicals. PBT chemicals have large changes in their reported quantities from the year 1997 to 1998, and this probably comes from a change in reporting requirements made in 1999 (companies report in 1999 for 1998 so this change comes into effect in the 1998 dataset). Seven new PBT chemicals were listed and the reporting thresholds for existing PBT chemicals on the list were lowered to 100 pounds. These changes likely generate noise and make it difficult for the regression to estimate the effect. An explanation for the lack of statistical significance could also be the lack of statistical power, as this subset has much fewer observations than the previous subsets. Only 3% of the total releases in the treatment group are from PBT chemicals, giving the regression little information to make an estimate with. Finally, the result could also mean that there is no statistically significant change in the release of PBT chemicals correlated with joining the UNGC.

5.4.1.5 Restricted Control Group

Next, I use regression Model (1) on a restricted dataset. The analyses performed this far have used all available companies in the TRI dataset as the control group. This can create some concern about how similar the control group really is to the treatment group. To address this issue, I have found three restrictions to implement to create a control group that should be more similar to the treatment group. The restrictions I have concluded to use follow in three steps.

First, I remove observations from federal facilities as these are likely to differ in characteristics from the treatment group. There are no companies in the treatment group that have federal facilities, and federal facilities cannot join the initiative. Below is the outcome of running the Model (1) regression on the full collapsed dataset without federal facilities.

Stata Output 6 – Model (1) on dataset without federal facilities

```
. xtreg totalreleases i.active_treat i.year, fe cluster(name_n)

Fixed-effects (within) regression               Number of obs     =    387,598
Group variable: name_n                          Number of groups  =     28,233

R-squared:                                      Obs per group:
     Within  = 0.0010                                         min =          1
     Between = 0.0003                                         avg =       13.7
     Overall = 0.0002                                         max =         35

                                                F(35,28232)       =       4.17
corr(u_i, Xb) = -0.0279                         Prob > F          =     0.0000

                            (Std. err. adjusted for 28,233 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
totalreleases | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat |   -3266928    1120182    -2.92   0.004     -5462539    -1071317
```

There is almost no change in the estimate from the first analysis, it is still negative and statistically significant. The coefficient has gone down a little, but there are slightly fewer observations. In this analysis there are 387,598 observations and 28,233 companies, 119 of these in the treatment group.

The next step is to remove the first three years of the dataset. I remove the first three reporting years, 1987, 1988 and 1989, as these might be more prone to errors being the first few years of the TRI reporting program, and after three years, I would assume that the companies have become more accustomed to the rules of the program and therefore reporting to be more accurate.

If I run the Model (1) analysis again on this dataset, now excluding federal facilities and the three first reporting years, I get these results:

Stata Output 7 – Model (1) on dataset without federal facilities and first three reporting years

```
. xtreg totalreleases i.active_treat i.year, fe cluster(name_n)

Fixed-effects (within) regression               Number of obs     =    349,471
Group variable: name_n                          Number of groups  =     26,237

R-squared:                                      Obs per group:
     Within  = 0.0009                                         min =          1
     Between = 0.0006                                         avg =       13.3
     Overall = 0.0000                                         max =         32

                                                F(32,26236)       =       4.30
corr(u_i, Xb) = -0.0244                         Prob > F          =     0.0000

                              (Std. err. adjusted for 26,237 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
totalreleases|      Coef.   Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat|  -1859034   552588.7    -3.36   0.001    -2942138   -775930.2
```

There is still a negative and statistically significant effect, but it can be interesting to note that the effect is now smaller. Joining the UNGC is here correlated with a reduction in total releases of 1.8 million pounds, still a large effect. The first reporting year, 1987, of the TRI program contains some large outliers. The question could be whether these measurements are overestimated (measuring chemical releases is not an exact science) in the first year, and companies correct their measurement errors the next year. Or it could be that the first year is a more honest report of releases, and that due to public scrutiny they adjust their numbers afterwards but still release the higher amounts. Or could it be that this is a correct measurement for the first year, and that either due to public scrutiny or regulation the companies actually reduce their releases for the next year? The years removed are all many years before any company joins the UNGC, and so it is a little baffling that there should be

such a large difference in the estimated effect. But as the DiD estimator compares the average from before and after treatment, these large outliers might have a large impact on the average before treatment, and as such affect the estimated treatment effect.

The last restriction imposed to create the restricted control group is to restrict the dataset to only include balanced companies, which means only keeping companies with observations for all remaining 32 years. Using only the balanced dataset, like Bui and Kapon (2012) do for their analysis, removes many companies from the dataset, but what remains is full information for the companies that are left. This leaves 2,861 companies in total, from an earlier 28,270 companies, now with 85 companies in the treatment group. Removing the treated companies that do not have information for the entire window can remove some concerns about irregularities such as companies that stop reporting before joining or stop reporting right after joining the initiative. I cannot know what changes caused them to stop reporting; it could be that they use less chemicals and therefore are not required to report, or it could be that they went out of business, or changed chemicals to something that is not on the list; the possibilities are many and does not provide the information I seek. This restriction creates a much smaller dataset, and running Model (1) on the restricted dataset, now without federal facilities, the first three reporting years and only including balanced companies, provides the following result:

Stata Output 8 - Model (1) on restricted control group dataset

```
. xtreg totalreleases i.active_treat i.year, fe cluster(name_n)

Fixed-effects (within) regression               Number of obs     =     91,552
Group variable: name_n                          Number of groups  =      2,861

R-squared:                                      Obs per group:
     Within  = 0.0021                                         min =         32
     Between = 0.0055                                         avg =       32.0
     Overall = 0.0001                                         max =         32

                                                F(32,2860)        =       2.86
corr(u_i, Xb) = -0.0316                         Prob > F          =     0.0000

                                (Std. err. adjusted for 2,861 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
totalreleases|      Coef.   Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat|  -2240716   686421.6    -3.26   0.001    -3586647   -894784.4
```

Joining the initiative is associated with a reduction in total releases of 2.2 million pounds. This is statistically significant at a 0.1% level. The coefficient shows a large reduction, more than half of the average annual chemical releases in the treatment group and implies a positive effect from joining the initiative. The effect estimated is based on a smaller treatment group which could affect the size of the estimate, but the results support earlier findings.

5.4.1.6 Alternative definition of control group

Another way of testing the found effect is to redefine the control group in the full collapsed dataset. There are companies in the dataset that join the UNGC after 2021, meaning after the dataset ends. Instead of using all available companies in the dataset, I can use this group of late joiners as the control group to see if there might be something that separates the companies that join the initiative from all other companies. Using the late joiners as the control group to the early joiners treated between 2000-2021, gives 176 companies, 119 of them in the treatment group. If there is something that uniquely identifies companies that join the initiative from all other companies, then this form of the analysis will be comparing like to like, as they all voluntarily join the UNGC. I already suspect that there are self-selection issues since companies voluntarily join the initiative, and this approach allows me to investigate if self-selection drives the results found above. If I find a statistically significant negative effect in the results below, it could lay these concerns to rest.

Stata Output 9 – Model (1) with alternative control group

```
. xtreg totalreleases i.active_treat i.year if treat == 1, fe cluster(name_n)

Fixed-effects (within) regression               Number of obs      =      4,794
Group variable: name_n                          Number of groups   =        176

R-squared:                                      Obs per group:
     Within  = 0.0177                                         min =          1
     Between = 0.0016                                         avg =       27.2
     Overall = 0.0119                                         max =         35

                                                F(35,175)          =       1.31
corr(u_i, Xb) = -0.0228                         Prob > F           =     0.1291

                                (Std. err. adjusted for 176 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
totalreleases|  Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat|   -1584924    1321786    -1.20   0.232    -4193617     1023768
```

The result is a negative coefficient, but it is not statistically significant. The number of observations and companies in the control group is much lower here, and this lack of statistical power could explain why I do not get a statistically significant coefficient. It could also be that this result indicates that I have a problem with selection bias or spurious correlation. Certain types of companies might be more likely to join the global compact, so when comparing early to late joiners, I might not see an effect, because the effect does not come from joining the UNGC. Rather there is a third factor for these firms that might lead them to join the initiative and release less, or companies that choose to join the initiative are companies that either way would reduce their chemical releases.

5.4.1.7 Robustness checks

All analyses with the regression from Model (1) seem to have a statistically significant coefficient, except for when using the subset with only PBT chemicals or the late joiners as control group, which both have much lower statistical power. To check the robustness of these findings, there are three concerns that should be investigated to further support or question the results. The issue of observations with zero total releases has not yet been addressed in the analysis. Therefore, I will perform the Model (1) regression on the full dataset, like the first analysis, but with some specifications regarding reported zeroes. First, I run the regression on the full dataset, but excluding companies that always report zero total releases (this removes 7 companies from the treated group and 3,260 from the control

group). The results from this regression are presented in column (1) in Stata output 10. Secondly, I run the regression on the full dataset without observations with zero total releases (this removes 64,374 observations, 168 from the treatment group). The results are presented in column (2) in Stata output 10. A slightly different concern is that the dataset contains large outliers, and that these might be driving the results. To address this, I run the regression on the full dataset without the top 5% observations, see column (3) in the output below. I have gathered the results of the regressions in a table, note that the number of observations in each regression is different from each other and from the full collapsed dataset. It is not meant for comparison in size of the coefficient, but to simply check if the results are still negative and statistically significant.

Stata Output 10 – Model (1) robustness checks

	(1) totalrelea~s	(2) totalrelea~s	(3) totalrelea~s
1.active_t~t	-3295742.6**	-3450867.6**	-23820.0*
	(1130875.0)	(1188263.2)	(11449.3)
_cons	2009038.4***	2246831.7***	65499.9***
	(511817.4)	(567757.4)	(987.8)
N	367691	323689	368659

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

The results from all three tests still show a statistically significant negative coefficient. Thus supporting the earlier findings, and giving some rest to concerns regarding the treatment of observations with zero values and outliers. Note that the effect is significantly smaller when removing the top five percent, but then again, I am not questioning whether the effect of joining the UNGC is equal for all companies, and companies that release large quantities will have the greatest opportunity to make large reductions.

5.4.1.8 Testing for spurious correlation

With more time I would've liked to investigate whether the correlation found could be spurious by adopting fake treatment dates for the treated companies. Some preliminary attempts using fake adoption dates reveal that it is possible that the estimated effect is spurious (see appendix). This could mean that joining the UNGC is not causing the

companies to release less chemicals, but there might be an underlying factor that influences both the decision to join and release levels.

5.4.2 Ln Total Releases (2):

The next model for analysis, Model (2), uses the natural logarithm of total releases as the dependent variable. This is often used in economics, partly for the ease of interpretation as the coefficients can be interpreted as percentage changes, and partly to reduce the effects of outliers and give a more symmetric distribution of the variables. It can also address concerns of heteroskedasticity. The main question is whether this is a better functional form for the regression or not. The mathematical model is as follows:

$$ln(totalreleases)_{it} = \beta_0 + \beta_1 * active\_treat_{it} + \alpha_i + \delta_t + u_{it}$$

The dependent variable is *ln(totalreleases)* which is the natural logarithm of total pounds of chemical releases for company *i* in year *t*. *active_treat*, the independent variable, is a dummy variable that equals 1 for companies in the treatment group for years after joining the initiative. It is this variable's coefficient, β_1, that measures the effect of treatment. *u* is a hopefully well-behaved error term with conditional mean of 0. α is the firm fixed effects for firm *i*, and δ is time fixed effects for time *t*.

5.4.2.1 Full collapsed

I begin by running the new regression from Model (2) on the full collapsed dataset. As it is not possible to take the natural logarithm of zero, these observations are thrown out of all analyses using the natural logarithm of total releases. To be able to compare with Model (1), I also run Model (1) without observations with zero total releases to compare with Model (2). Below is the output from Model (2) on the full collapsed dataset.

Stata Output 11 – Model (2) on full collapsed dataset

```
. xtreg ln_totalreleases i.active_treat i.year, fe cluster(name_n)

Fixed-effects (within) regression               Number of obs     =    323,689
Group variable: name_n                          Number of groups  =     25,007

R-squared:                                      Obs per group:
     Within  = 0.0965                                         min =          1
     Between = 0.0531                                         avg =       12.9
     Overall = 0.0410                                         max =         35

                                                F(35,25006)       =     119.40
corr(u_i, Xb) = -0.0014                         Prob > F          =     0.0000

                              (Std. err. adjusted for 25,007 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
ln_totalrele~s |      Coef.   Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat |   .0074117   .1901769     0.04   0.969    -.3653461    .3801696
```

Here the result is a positive coefficient, stating that joining the UNGC is correlated with an increase of 0.7% in total releases. This is however not statistically significant. Comparing to Model (1) gives the following table:

Stata Output 12 – Comparison model (1) & (2)

	(1) ln_totalre~s	(2) totalrelea~s
1.active_t~t	0.00741	-3450867.6**
	(0.190)	(1188263.2)
_cons	9.914***	2246831.7***
	(0.0263)	(567757.4)
N	323689	323689

From Model (1) to Model (2), the coefficient goes from a negative effect that is statistically significant, to a positive effect that is not. What could be the explanation for this? If the natural logarithm of total releases is the correct functional form, then the effect seen in Model (1) could be driven by large observations and outliers. However, when testing Model (1) without the 5% largest observations, it still yielded a significant negative effect. It could also be that Model (2) is not the correct specification, or that the variations in the values are so large that it is difficult to estimate a fixed percentage of the changes.

To see how this alternative specification affects the outcome, I will continue with Model (2) for the other subsets used for Model (1).

5.4.2.2 *Subset without Clean Air Act chemicals*

I run the Model (2) regression on a subset without chemicals that are regulated by the Clean Air Act:

Stata Output 13 – Model (2) on subset without clean air act chemicals

```
Fixed-effects (within) regression               Number of obs      =    211,439
Group variable: name_n                          Number of groups   =     18,522

R-squared:                                      Obs per group:
     Within  = 0.0740                                         min =          1
     Between = 0.0708                                         avg =       11.4
     Overall = 0.0364                                         max =         35

                                                F(35,18521)        =      76.58
corr(u_i, Xb) = 0.0140                          Prob > F           =     0.0000

                              (Std. err. adjusted for 18,522 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
ln_totalrele~s |      Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat |   .0468628   .1817321     0.26   0.797    -.3093488    .4030744
------------------------------------------------------------------------------
```

Stata Output 14 - Comparison model (1) & (2)

	(1) totalrelea~s	(2) ln_totalre~s
1.active_t~t	-2212332.1*	0.0469
	(1104260.3)	(0.182)
_cons	2734558.0**	9.326***
	(837751.3)	(0.0348)
N	211439	211439

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Again, the result is a positive coefficient saying joining the UNGC is correlated with an increase of 5% in total releases of chemicals not regulated by the Clean Air Act, but it is not statistically significant. A positive coefficient is the opposite effect of what we would want from a voluntary program, but it is not statistically significant. This result indicates that there is no correlation between joining the initiative and level of chemical releases. The table compares Model (1) and Model (2) using this dataset, and again, using the same data but different model specification provides opposite results.

5.4.2.3 *Subset with only carcinogens*

Below is the output from Model (2) run on the dataset including only known carcinogens:

Stata Output 15 – Model (2) on subset with only carcinogens

```
. xtreg ln_totalreleases i.active_treat i.year, fe cluster(name_n)

Fixed-effects (within) regression               Number of obs     =    174,311
Group variable: name_n                          Number of groups  =     14,871

R-squared:                                      Obs per group:
     Within  = 0.0901                                         min =          1
     Between = 0.1390                                         avg =       11.7
     Overall = 0.0856                                         max =         35

                                                F(35,14870)       =      57.72
corr(u_i, Xb) = 0.1137                          Prob > F          =     0.0000

                              (Std. err. adjusted for 14,871 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
ln_totalrele~s |      Coef.   Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat |  -.1131009   .2379002    -0.48   0.634    -.5794146    .3532127
```

Stata Output 16 - Comparison model (1) & (2)

	(1) totalrelea~s	(2) ln_totalre~s
1.active_t~t	-314231.0**	-0.113
	(115647.5)	(0.238)
_cons	156410.4***	8.565***
	(10166.0)	(0.0446)
N	174311	174311

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Now there is a negative coefficient, indicating that joining the UNGC is correlated with an 11%-reduction of carcinogenic chemical releases, but it is not statistically significant. Here it is interesting to note that the coefficient is negative, as opposed to the two estimated coefficients above.

5.4.2.4 Subset with only PBT's

The next subset is to use the dataset containing only observations of PBT chemicals:

Stata Output 17 – Model (2) on subset with only PBT chemicals

```
Fixed-effects (within) regression               Number of obs      =     81,730
Group variable: name_n                          Number of groups   =      7,452

R-squared:                                      Obs per group:
     Within  = 0.0469                                         min =          1
     Between = 0.0660                                         avg =       11.0
     Overall = 0.0505                                         max =         35

                                                F(35,7451)         =      20.87
corr(u_i, Xb) = 0.1128                          Prob > F           =     0.0000

                              (Std. err. adjusted for 7,452 clusters in name_n)
------------------------------------------------------------------------------
ln_totalrele~s | Coefficient  Robust std. err.      t    P>|t|     [95% conf. interval]
---------------+--------------------------------------------------------------
1.active_treat |   .0427274   .3032378     0.14   0.888    -.5517043     .637159
```

Stata Output 18 - Comparison model (1) & (2)

```
------------------------------------------------
                      (1)             (2)
             totalrelea~s    ln_totalre~s
------------------------------------------------
1.active_t~t     -48763.2          0.0427
                (111344.7)        (0.303)

_cons            -31249.5          5.964***
                (142867.1)        (0.101)
------------------------------------------------
N                    81730           81730
------------------------------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Here there is a positive coefficient, indicating that joining the initiative is correlated to an increase in releases of 4%, but it is not statistically significant. In this regression there is much fewer companies and much less information to estimate an effect with, remember that only 3% of the total releases from the companies in the treatment group are of this chemical category. The changes in reporting requirements for this category also likely generates noise.

5.4.2.5 Restricted Control Group

And last for Model (2) I use the restricted control group dataset, a balanced dataset excluding federal facilities and reporting years 1987-1989. These restrictions are made to address concerns about the control group in the full dataset, and to provide a more similar control group. I run Model (2) on this restricted dataset and the outcome is as follows:

Stata Output 19 – Model (2) on restricted control group

```
Fixed-effects (within) regression               Number of obs     =     84,865
Group variable: name_n                          Number of groups  =      2,847

R-squared:                                      Obs per group:
     Within  = 0.0731                                         min =          1
     Between = 0.0248                                         avg =       29.8
     Overall = 0.0151                                         max =         32

                                                F(32,2846)        =      31.09
corr(u_i, Xb) = -0.0140                         Prob > F          =     0.0000

                                (Std. err. adjusted for 2,847 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
ln_totalrele~s |      Coef.   Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat |  -.0172762   .1691708    -0.10   0.919    -.348986    .3144336
```

Stata Output 20 - Comparison model (1) & (2)

```
------------------------------------------------
                      (1)             (2)
             ln_totalre~s    totalrelea~s
------------------------------------------------
1.active_t~t       -0.0173      -2240233.7**
                   (0.169)      (690754.8)

_cons                10.55***    1135424.2***
                  (0.0391)       (99733.0)
------------------------------------------------
N                    84865           84865
------------------------------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Now there is a negative coefficient; joining the initiative is correlated with a reduction of 1.7% in total releases. The coefficient is still not statistically significant.

In summation for Model (2) using the natural logarithm of total releases as the dependent variable, there seems to be no statistically significant relationship between joining the UNGC and percentage changes in chemical releases. These findings are in opposition to the findings from Model (1). Some possible explanations could be that one of the models is a better fit, and therefore the other is simply wrong. However, there is no clear evidence as to which would be the better model, or what better models the relationship. It seems unlikely to me that the effect from joining the UNGC on chemical releases would be better estimated in percentage terms. Imagine a large corporation, with high levels of chemical releases. If they go from releasing 10,000 pounds, to 9,000 pounds, this is a reduction of 1,000 pounds, or 10%. A small company reducing its releases by 10% would have a much smaller total effect, say for example they produce 1000 pounds, but then reduce to 900 pounds, saving the environment from 100 pounds. If both of these companies instead change to a non-toxic

chemical, the impact from each would be even more different. The real value of chemical releases in pounds account for such differences, whereas the percentage effect would not. However, if each facility for each company change to a chemical where 30% less of the chemical is needed for the same level of production, then this effect comes out to be of similar percentages for the different companies. I cannot draw any conclusions as to which of the models is the best one, and both models are possible. For the purpose of this paper, I would argue that the interesting effect is the overall real reduction as this shows the actual effect of how much chemicals are going directly into our world. Larger companies with higher releases also have greater possibility for larger reductions, and compressing these numbers might take away much of the effect, especially since the companies in the treatment group seem to be generally larger in size.

In regards to a possible positive coefficient, like I have seen (however insignificant) in several of the results from Model (2), Zhang and Khanna (2020) show that there is an instance where companies would increase releases by participating in a voluntary program. These "free-riders" are companies that release a little below the threshold imposed from the program prior to participation and who increase releases after joining relative to companies that do not join. However, this is unlikely in the case of the UNGC as the program does not have specific thresholds, and so free-riders could not increase "up until" the threshold after joining.

5.4.3 Toxicity-weighted (3):

To answer the research question of whether sustainability commitments actually lead to more sustainable practices, I wish to determine if joining the UNGC has real world effects that benefit the environment and human health. The chemicals in the dataset are of different levels of toxicity, meaning that releases of different chemical categories do not have the same impact pound for pound, or percentage for percentage. A company desiring to reduce their negative externalities would hopefully take these different impact levels into consideration when looking at their sustainability profile. The following model will attempt to take the different levels of toxicity of the chemicals into account. In this next model the dependent variable will be a toxicity weighted aggregation of total releases that could better

reflect the toxicity levels and the real impact of an effect. To create the toxicity weighted unit, I use the reporting thresholds for the different chemicals as this gives me a way to measure the chemicals against each other; the more toxic the chemical, the lower the threshold. Normal TRI chemicals can have reporting thresholds at 10,000 pounds while dioxin and dioxin-like compounds have reporting thresholds at 0.1 gram. PBT and PFAS chemicals have reporting thresholds at 100 or 10 pounds. I use the information these thresholds provide to weigh the chemical releases by their toxicity and create a new toxicity-weighted unit to measure releases. The amount of a chemical released in pounds is divided by the reporting threshold for that specific chemical, and this is the new toxicity-weighted unit.

The model is now:

$$weightedtotalreleases_{it} = \beta_0 + \beta_1 * active\_treat_{it} + \alpha_i + \delta_t + u_{it}$$

The dependent variable, *weightedtotalreleases*, is a toxicity-weighted unit of total releases for company *i* at time *t*. The independent variable *active_treat* is a dummy variable that equals 1 for companies in the treatment group for years after joining the initiative. It is always zero for companies in the control group, and for all years before joining UNGC in the treatment group. It is this variable's coefficient, β_1, that measures the effect of treatment. *u* is a hopefully well-behaved error term with conditional mean of 0. α is the firm fixed effects for firm *i*, and δ is time fixed effects for time *t*.

5.4.3.1 Full collapsed

Running this regression from Model (3) on the full collapsed dataset with the new toxicity-weighted unit gives the following outcome:

Stata Output 21 – Model (3) on toxicity-weighted dataset

```
. xtreg weightedtotalreleases i.active_treat i.year, fe cluster(name_n)

Fixed-effects (within) regression               Number of obs     =    388,063
Group variable: name_n                          Number of groups  =     28,270

R-squared:                                      Obs per group:
     Within  = 0.0003                                         min =          1
     Between = 0.0000                                         avg =       13.7
     Overall = 0.0001                                         max =         35

                                                F(35,28269)       =       3.46
corr(u_i, Xb) = -0.0099                         Prob > F          =     0.0000

                              (Std. err. adjusted for 28,270 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
weightedtota~s |      Coefficient   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat |      -1725.77   1331.035    -1.30   0.195    -4334.662    883.1224
```

There is a negative coefficient; joining the initiative is correlated with a reduction of 1.7 thousand units of the toxicity-weighted total releases. This effect is not statistically significant, even at the 10%-level.

In this analysis the theory is that a company serious about their environment engagement would attempt not only to release less chemicals, but also to focus on lowering their use of highly toxic chemicals. However, dependent on their motivation for joining the initiative, if social pressure is a motivating factor, they might be more inclined to make changes that are more visible, in other words reduce releases of chemicals they use in large quantities, which are often less toxic ones.

Zhang and Khanna (2020) claim that companies who join voluntary programs and see a positive effect often show improvements in air emissions and not water emissions and they theorize that this is due to social pressure creating incentives to reduce visible emissions. If the driving force for companies joining the UNGC is social pressure we might expect to see large reductions in less poisonous chemicals, as this is more visible to the public and makes a more dramatic numerical change. This could explain why a toxicity weighted model would not yield statistically significant results.

5.4.3.2 Subset without Dioxin

Dioxin is the most toxic chemical category and the reporting threshold for it is 0.1 grams, giving it extremely much weight in this third model. Dioxins constitute less than 1% of the raw data observations and less than 1% of the total chemical releases, but with the toxicity-

weighted unit Dioxins constitute approximately 30% of the weighted total releases. Therefore, it seems prudent to run the same regression from Model (3) on a subset without observations of Dioxins. I do not have a suitable alternative to weigh this chemical category against the other chemicals, and removing these observations removes very little information from the dataset. Below is the result of running Model (3) on a subset of the toxicity-weighted dataset excluding dioxins.

Stata Output 22 – Model (3) on subset without Dioxin

```
. xtreg weightedtotalreleases i.active_treat i.year, fe cluster(name_n)

Fixed-effects (within) regression               Number of obs     =    387,937
Group variable: name_n                          Number of groups  =     28,265

R-squared:                                      Obs per group:
     Within  = 0.0002                                         min =          1
     Between = 0.0001                                         avg =       13.7
     Overall = 0.0000                                         max =         35

                                                F(35,28264)       =       3.64
corr(u_i, Xb) = -0.0081                         Prob > F          =     0.0000

                            (Std. err. adjusted for 28,265 clusters in name_n)
------------------------------------------------------------------------------
             |               Robust
weightedtota~s | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.active_treat |  -370.6785   122.9502    -3.01   0.003    -611.6668   -129.6902
```

Here there is a negative and statistically significant effect at the 1%-level. Joining the initiative is correlated to a reduction of 370 units of toxicity-weighted total releases. A likely explanation for this result being statistically significant, when the previous result was not, is that when dioxins receive a large weight, it generates noise and so the regression cannot estimate the effect. Dioxins represent a very small part of the dataset, and it is likely not the driving force behind an effect.

Having less than 1% of the total releases in real values representing 30% of the toxicity-weighted total releases unit might be giving it too much weight in relation to the other chemicals. Dioxins might not get the same focus from stakeholders as they are often of small quantities, and if we again look to the research commenting that visible emission is more likely to be scrutinized by the public (Zhang & Khanna, 2020), then a few grams, however deadly, is unlikely to win the cover story in a world where rivers suddenly catch on fire.

The results from Model (3) indicate that when removing the observations of Dioxin releases, there is a statistically significant negative correlation between joining the UNGC and

chemical releases weighted by their toxicity. This conclusion is in line with the findings from Model (1); joining the UNGC seems to be associated with a favourable effect of reducing chemical releases. Not only do companies that join the program reduce less in real values, but they also reduce their impact level. The difference in results from the two regressions in Model (3) could be the result of an over-weighting of Dioxins in the first attempt. It could also mean that there hasn't been a change in Dioxin releases, and therefore when these observations get a large weight, it removes any statistical significance. Dioxins are mainly a by-product of a manufacturing process, and so it might be more difficult to reduce these chemical releases.

Using the thresholds to weigh the chemicals might not the best possible method to account for toxicity. Much existing research using the TRI dataset seem to use the simple aggregated pounds of total releases as variable of interest, and some use different weighting methods that are beyond the scope of this thesis. There is no standard weighting method and the one employed in this thesis is an imperfect proxy, but it still yields some interesting findings.

5.4.4 Event Study (4):

The fourth and final model is an event study. The event study is a model for dynamic effects, meaning that the model does not only provide the average for the time before versus after treatment, but allows for the effect of treatment to vary over time. In the event study, only companies that join the initiative (including companies that join after the dataset, as late as 2023) are included in the regression, so these regressions are run for a much smaller dataset. For each of these companies the time variable is standardised so that when the treatment occurs the time variable equals 0, the year prior it equals -1 and the year after treatment it equals 1, etc. Then I find a time window that will include multiple periods before and after treatment for as many companies as possible. For the chosen window I create event time dummy variables for each of the event times, which takes on the value 1 for the specific event time and zero otherwise.

5.4.4.1 Full collapsed

For the full collapsed dataset, I have found the most suitable time window to be from 3 years prior to 2 years after treatment. I then create individual event time dummies for each of these event times. The variable *event_m3* equals 1 when this observation is from the event time three years prior (-3) to treatment and zero otherwise. *Event_m2* equals 1 when the observation is from the event time two years prior (-2) to treatment and zero otherwise. And so on. In the end I am left with six event time dummy variables, looking like this:

Table 8 - Event time dummy variables

Event time	event_m3	event_m2	event_m1	event_m0	event_p1	event_p2
-3	1	0	0	0	0	0
-2	0	1	0	0	0	0
-1	0	0	1	0	0	0
0	0	0	0	1	0	0
1	0	0	0	0	1	0
2	0	0	0	0	0	1

The mathematical model is as follows:

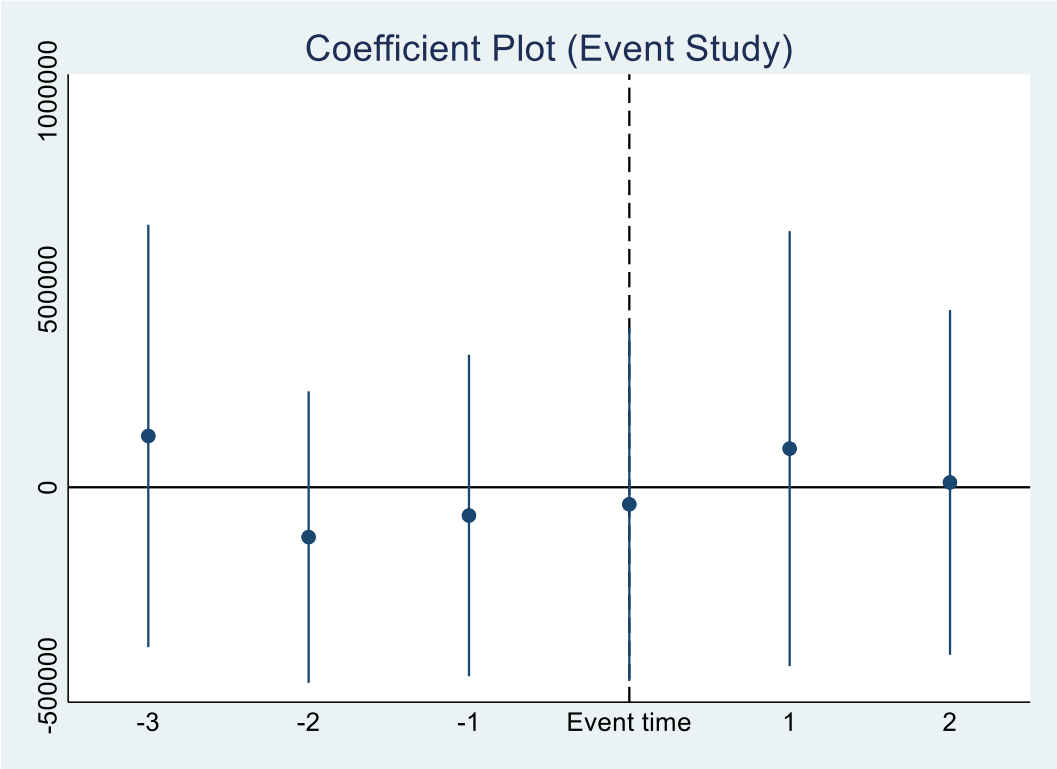
$$totalreleases_{it} = \beta_0 + \sum_{m=-M}^{P} \beta_m Z_{i,t-m} + \alpha_i + \delta_t + u_{it}$$

Totalreleases is total pounds of chemical releases for company *i* in event time *t*. *Z* are the event time dummies for company *i* in event time *t* minus *m*, *m* indicating which event time dummy, ranging from *M* past periods to *P* future periods. β_m estimates the coefficients for these dummies, estimating the dynamic treatment effects. *u* is a hopefully well-behaved error term with conditional mean of 0. α is the firm fixed effects for firm *i*, and δ is time fixed effects for time *t*.

In the regression, only a balanced dataset is used. 65 companies have the entire chosen window, and the periods when they are not treated serve as controls. One benefit of using

event studies is the possibility of graphical interpretation of the results. Below is a coefficient plot, where one can see the estimated coefficients with their confidence intervals.

Figure 12 - Coefficient plot from event study using full dataset

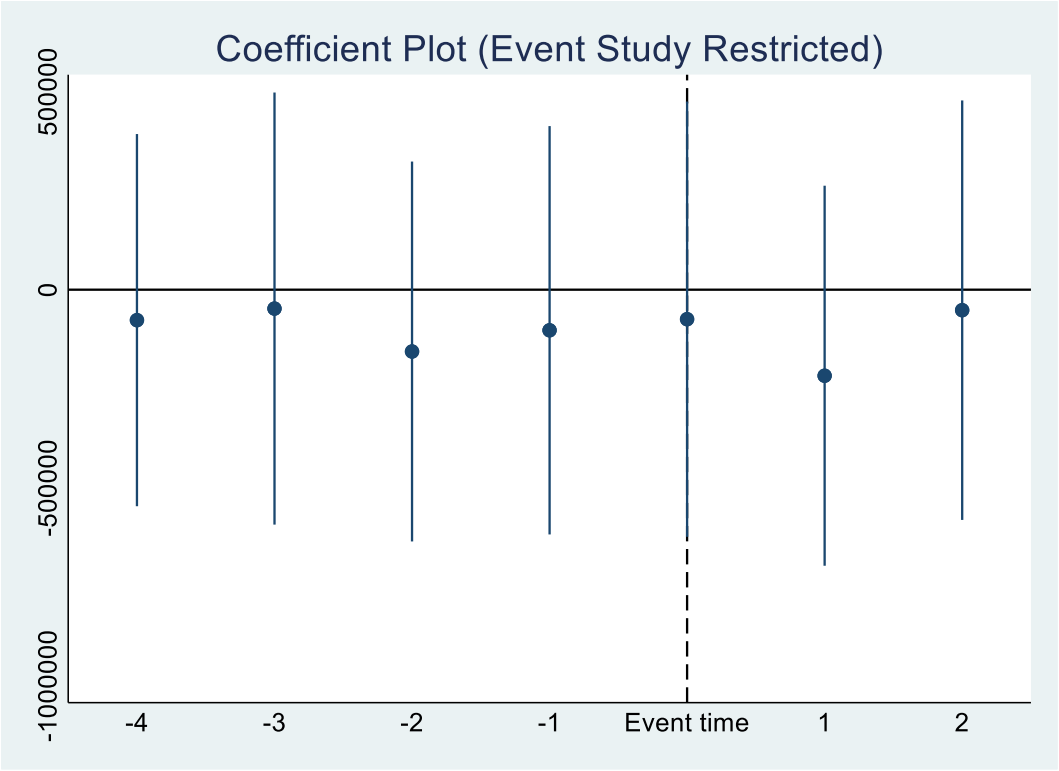


All coefficients in the results are statistically insignificant. The results would suggest that there is an increase in total releases after treatment compared to the period for up to two years before. The first three coefficients address the time before treatment and as they are statistically insignificant, it means that the control group and treatment group are not statistically different from each other, and this could support the parallel trends assumption. However, in this model, the control group and the treatment group are the same companies, so this is not surprising. This model has a small group size and few observations, which limits its statistical power. The interesting thing to note are the negative coefficients right before treatment versus the positive coefficients after. If this was statistically significant it could maybe indicate that companies reduce releases before joining the initiative. The positive coefficients after treatment could mean that after joining, they release more as they experience less scrutiny.

5.4.4.2 *Restricted Control Group*

To compare, I run the same event study model, but on the restricted control group dataset. For this dataset, I have included one more event time dummy for 4 years prior to the treatment as it does not change how many companies have the full window but might give more information about the trends before treatment. The event time window is now -4 to 2, and 50 companies have this entire time window.

Figure 13 - Coefficient plot from event study using restricted dataset

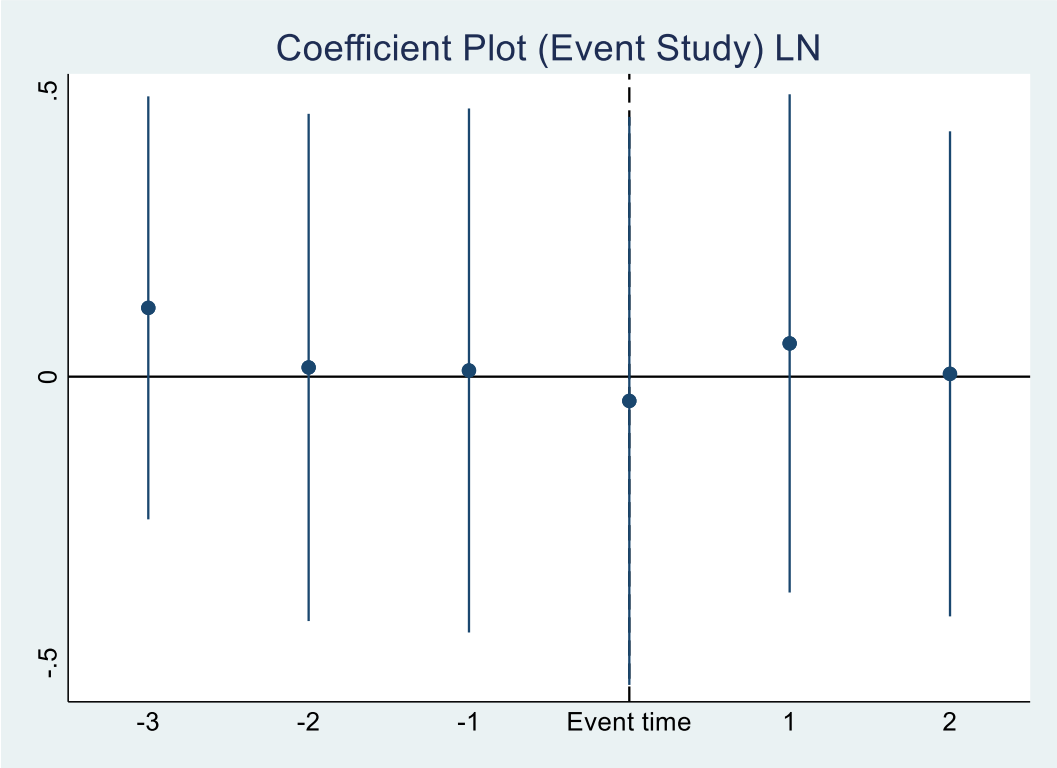


All the coefficients are negative, but statistically insignificant. Again, the number of observations and companies in the dataset used for the regression are small, only 50 companies. The coefficients before treatment are negative, and if this was statistically significant, it could possibly indicate that companies reduce releases for some other reason and join the initiative when they already have lower releases. The coefficient for the first year after treatment could indicate a larger reduction after treatment, but it is not statistically significant.

5.4.4.3 Ln Total releases

Next, I look at the event study model using the natural logarithm of total releases as the dependent variable. This shows the percentage changes over time.

Figure 14 - Coefficient plot from event study using the natural logarithm of total releases



There are no statistically significant coefficients, but they are all positive except for the coefficient for the event time when treatment occurs. There are now only 64 companies in the analysis.

The Event Study, Model (4), does not yield any statistically significant coefficients. This could be due to the lack of statistical power as the number of companies are so small, or it could support a hypothesis that there is an issue with self-selection and that when only looking at companies that self-select into the program there is no effect from joining the initiative. This could hold true if there is something that is unique for the companies that join versus all other companies, something that makes them voluntarily join the program and reduce releases, not necessarily in that order. In the Event Study only a part of the available information is used. For the standard DiD I use all the information for all years before treatment and all years after treatment. In the event study model, only the information for

the chosen time window is used, therefore discarding much information, but possibly leaving it less vulnerable to omitted variables that change over time and across firms.

6 Conclusion

In this thesis I have addressed the question "Do sustainability commitments by companies actually lead to more sustainable practices?". In an attempt to answer this question, I have studied companies' membership in the United Nations Global Compact, a global initiative which calls on corporations to take on social and environmental responsibility, in relation to chemical releases as reported by companies to the Toxics Release Inventory. I have employed an exploratory approach, using several models and sub-datasets to explore the relationship between joining the initiative – which is voluntary and not subject to any control measure, and a company's chemical releases – where reporting is mandatory. Surprisingly, the analyses mainly revealed a statistically significant negative relationship, meaning that participation in the UNGC was associated with reductions in chemical releases.

In theory, I would not expect to find a significant effect. I would expect companies to focus on maximizing profits, and to join the initiative only to gain benefits such as improved corporate image and network opportunities, without implementing changes that in any case would not be verified by the UNGC. Some companies might have an actual preference for sustainability, but I expect most companies to join a voluntary program for reasons relating to their financial performance. Improving environmental performance can provide a competitive advantage for the company and improve the financial performance through energy efficiency or from consumers willing to pay a green premium on sustainable goods. A major reason for companies to join a voluntary program is to appeal to green consumers. If consumers have a preference for sustainable goods, joining a voluntary program can attract these customers, but companies also risk being labelled greenwashers if they do not actually improve their environmental performance. If the consumers have a social preference for green products, on the other hand, then the risk of being labelled greenwashers is smaller, as the consumer is more concerned with the image than the actual environmental performance. With regard to the UNGC, there is no possibility for the consumers to check

that participants of the program actually improve their sustainability performance. Therefore, the risk for participants of being labelled greenwashers is relatively low, and one could expect a high degree of free-riders who gain the benefit of an improved corporate image without actually implementing any changes.

The results from the main model, Model (1), used total chemical releases in pounds as the dependent variable and showed significant reductions related to joining the UNGC. This finding was robust and showed statistically significant results for almost all subsets of the data that were used. Exceptions included analyses on smaller datasets with lower statistical power, namely the subset of data that only included PBT chemicals, and the analysis in which I used companies that joined the UNGC after 2021 as an alternative control group. In Model (2), I analysed the natural logarithm of the total chemical releases as the dependent variable, looking at the percentage changes in chemical releases. The analyses from this model gave no consistent results, they did not show statistically significant coefficients and estimates varied between positive and negative values depending on which subset was used. Model (3) used a toxicity-weighted unit of chemical releases as the dependent variable. The results from this model showed a statistically significant negative effect when excluding dioxins, which supports the findings from Model (1). The results from Model (4), an event study modelling dynamic effects, showed no significant effects but analyses were based on much smaller datasets, with low statistical power.

The toxicity-weighted model, Model (3) is an imperfect solution to address the real impact of the chemical releases, and in an ideal world, I would have preferred to find a more suitable method than using the reporting thresholds to weigh the chemicals. The results from the event study and the analysis using the companies that join the initiative after 2021 as an alternative control group could indicate that the effect found is spurious, and that there is some third factor affecting a company's decision to join the UNGC and its decision on how much to release. These aspects could also be evidence of issues regarding self-selection. In the descriptive statistics there is also possible evidence for selection issues in terms of state and industry composition of chemical releases, where the treatment group differs significantly from the control group.

The results from the event study could indicate that companies reduce their releases before joining the initiative, meaning that it is not joining the initiative that leads the company to reduce chemical releases. If total chemical releases affect the company's decision to join the initiative, there is a problem with simultaneous or reciprocal causality which threatens the study's internal validity, and the estimates will be biased as the independent variable is correlated with the error term. In terms of the study's reliability, the main concern is how truthful companies are in their reporting to the TRI and whether we should expect some degree of measurement error in the variable of interest, namely the total releases. Furthermore, the models may suffer from omitted variable bias. Many factors will affect a company's release levels, and variables that vary across both time and companies are not accounted for. Finally, it is questionable whether the results from this study can be generalised, since other countries operate under different regulations and different cultural attitudes, therefore findings from the US might not apply to other countries.

Something to consider in future research on the effects of participating in the UN Global compact, is to consider the definition of "treatment". One challenge with the Difference in Difference method, the method used in this thesis, can be that an entity assigned to the treatment group does not receive the treatment or does not complete the treatment. The way the models are run, the definition of treatment is merely joining the initiative by sending the letter and getting the company name on the participants list. Therefore, this possible challenge is not a concern in this thesis. But how would it look if we defined treatment as actually participating and engaging in the initiative? If companies that engage the most also report more positive effects for the business from the initiative, would it not be reasonable to think that engagement levels affect their environmental development? And if we could define engagement level, would this make a difference to the results?

In the future it would also be prudent to attain financial data on the companies to construct a control group with similar metrics to companies in the treatment group. Finally, it would be useful to control for other existing regulations in the United States, as this study only takes the Clean Air Act into consideration.

In conclusion, this thesis provided results showing that joining the United Nations Global Compact was associated with large reductions in chemical releases. The different models

used for the thesis provided mixed results, and I cannot draw any strong conclusions on which model best represents the relationship between joining the UNGC and chemical releases. Results from the preferred model using the real value of total chemical releases in pounds points in favour of the initiative. More research is needed to conclude if participation in the UNGC causes the companies to reduce chemical releases, and what mechanisms drive the effect.

References

Annan, K. (2002). The Global Compact. Corporare Leadership in the World Economy. *Symphonya. Emerging Issues in Management*. https://doi.org/10.4468/2002.2.02annan

Attina, T. M., Hauser, R., Sathyanarayana, S., Hunt, P. A., Bourguignon, J.-P., Myers, J. P., DiGangi, J., Zoeller, R. T., & Trasande, L. (2016). Exposure to endocrine-disrupting chemicals in the USA: a population-based disease burden and cost analysis. *The Lancet Diabetes & Endocrinology*, *4*(12), 996-1003. https://doi.org/10.1016/S2213-8587(16)30275-3

Berliner, D., & Prakash, A. (2014). The United Nations Global Compact: An Institutionalist Perspective. *Journal of Business Ethics*, *122*(2), 217-223. https://doi.org/10.1007/s10551-014-2217-5

Bloome, D. (2009). Violations of SUTVA. *Social Science Sttistics Blog*. https://blogs.iq.harvard.edu/violations_of_s

Boissoneault, L. (2019). The Cuyahoga River Caught Fire at Least a Dozen Times, but No One Cared Until 1969. *Smithsonian Magazine*. https://www.smithsonianmag.com/history/cuyahoga-river-caught-fire-least-dozen-times-no-one-cared-until-1969-180972444/

Brouhle, K., Griffiths, C., & Wolverton, A. (2009). Evaluating the role of EPA policy levers: An examination of a voluntary program and regulatory threat in the metal-finishing industry. *Journal of environmental economics and management*, *57*(2), 166-181. https://doi.org/https://doi.org/10.1016/j.jeem.2008.07.006

Bui, L. T. M., & Kapon, S. (2012). The impact of voluntary programs on polluting behavior: Evidence from pollution prevention programs and toxic releases. *Journal of environmental economics and management*, *64*(1), 31-44. https://doi.org/10.1016/j.jeem.2012.01.002

Cambridge Dictionary. (n.d., n.d.). *Greenwash*. Cambridge Dictionary. Retrieved 05.05.2023 from https://dictionary.cambridge.org/dictionary/english/greenwash

Cetindamar, D., & Husoy, K. (2007). Corporate Social Responsibility Practices and Environmentally Responsible Behavior: The Case of The United Nations Global Compact. *Journal of Business Ethics*, *76*, 163-176. https://doi.org/10.1007/s10551-006-9265-4

Chintrakarn, P., & Millimet, D. L. (2006). The environmental consequences of trade: Evidence from subnational trade flows. *Journal of environmental economics and management*, *52*(1), 430-453. https://doi.org/10.1016/j.jeem.2006.03.001 (Journal of Environmental Economics and Management)

Congressional Research Service. (2022). *Clean Air Act: A Summary of the Act and Its Major Requirements* (RL30853). https://crsreports.congress.gov/product/pdf/RL/RL30853

Croson, R., & Treich, N. (2014). Behavioral Environmental Economics: Promises and Challenges. *Environmental and Resource Economics*, *58*(3), 335-351. https://doi.org/10.1007/s10640-014-9783-y

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press. https://doi.org/10.2307/j.ctv1c29t27

EPA. (2022a, 30 June 2022). *History of the Toxics Release Inventory (TRI) Program*. United States Environmental Protection Agency. Retrieved 05.05.2023 from https://www.epa.gov/toxics-release-inventory-tri-program/history-toxics-release-inventory-tri-program

EPA. (2022b, 28 July 2022). *TRI Data Quality*. United States Environmental Protection Agency. Retrieved 05.05.2023 from https://www.epa.gov/toxics-release-inventory-tri-program/tri-data-quality

EPA. (2023, 25 Jan 2023). *What is the Toxics Release Inventory?* United States Environmental Protection Agency. Retrieved 05.05.2023 from https://www.epa.gov/toxics-release-inventory-tri-program/what-toxics-release-inventory

EPA. (n.d.-a, 03 Jan 2023). *Health and Ecological Hazards Caused by Hazardous Substances*. United States Environmental Protection Agency. Retrieved 13.05.2023 from https://www.epa.gov/emergency-response/health-and-ecological-hazards-caused-hazardous-substances

EPA. (n.d.-b, 08 May 2023). *Reporting Forms and Instructions*. United States Environmental Protection Agency. Retrieved 11.05.2023 from https://guideme.epa.gov/ords/guideme_ext/f?p=guideme:rfi:::::rfi:2_4

European Chemicals Agency. (2023, n.d.). *Per- and polyfluoroalkyl substances (PFAS)*. European Chemicals Agency. Retrieved 05.05.2023 from https://echa.europa.eu/hot-topics/perfluoroalkyl-chemicals-pfas

Gamper-Rabindran, S. (2006). Did the EPA's voluntary industrial toxics program reduce emissions? A GIS analysis of distributional impacts and by-media analysis of substitution. *Journal of environmental economics and management*, *52*(1), 391-410. https://doi.org/https://doi.org/10.1016/j.jeem.2005.12.001

Harford, T. (2022). Cautionary Tales [Audio Podcast Episode]. In *The Inventor Who Almost Ended the World*. PUSHKIN. https://www.pushkin.fm/podcasts/cautionary-tales/the-inventor-who-almost-ended-the-world

Innes, R. (2006). A Theory of Consumer Boycotts Under Symmetric Information and Imperfect Competition. *The Economic Journal*, *116*(511), 355-381. https://doi.org/10.1111/j.1468-0297.2006.01084.x

Innes, R., & Sam, Abdoul G. (2008). Voluntary Pollution Reductions and the Enforcement of Environmental Law: An Empirical Study of the 33/50 Program. *The Journal of Law and Economics*, *51*(2), 271-296. https://doi.org/10.1086/589659

Khanna, N. (2019). *A Narrative on Two Weaknesses of the TRI for Research Purposes*. Cornell University. https://dyson.cornell.edu/wp-content/uploads/sites/5/2019/05/Cornell-Dyson-wp1905.pdf

Kim, E.-H., & Lyon, T. P. (2011). Strategic environmental disclosure: Evidence from the DOE's voluntary greenhouse gas registry. *Journal of environmental economics and management*, *61*(3), 311-326. https://doi.org/https://doi.org/10.1016/j.jeem.2010.11.001

Kube, R., von Graevenitz, K., Löschel, A., & Massier, P. (2019). Do voluntary environmental programs reduce emissions? EMAS in the German manufacturing sector. *Energy economics*, *84*, 104558. https://doi.org/10.1016/j.eneco.2019.104558

Li, J., & Wu, D. (2020). Do Corporate Social Responsibility Engagements Lead to Real Environmental, Social, and Governance Impact? *Management science*, *66*(6), 2564-2588. https://doi.org/10.1287/mnsc.2019.3324

National Institute of Environmental Health Sciences. (2023, 9 Mar 2023). *Endocrine Disruptors*. National Institute of Environmental Health Sciences. Retrieved 05.05.2023 from https://www.niehs.nih.gov/health/topics/agents/endocrine/index.cfm

OECD. (n.d., n.d.). *Voluntary Approaches for Environmental Policy*. Retrieved 12.05.2023 from https://www.oecd.org/env/tools-evaluation/voluntaryapproachesforenvironmentalpolicy.htm

Ortas, E., Álvarez, I., & Garayar, A. (2015). The Environmental, Social, Governance, and Financial Performance Effects on Companies that Adopt the United Nations Global Compact. *Sustainability*, *7*(2), 1932-1956. https://www.mdpi.com/2071-1050/7/2/1932

Orzes, G., Moretto, A. M., Ebrahimpour, M., Sartor, M., Moro, M., & Rossi, M. (2018). United Nations Global Compact: Literature review and theory-based research agenda. *Journal of cleaner production*, *177*, 633-654. https://doi.org/10.1016/j.jclepro.2017.12.230

Potoski, M., & Prakash, A. (2013). Do Voluntary Programs Reduce Pollution? Examining ISO 14001's Effectiveness across Countries. *Policy Stud J*, *41*(2), 273-294. https://doi.org/10.1111/psj.12017

Sam, A. G., Khanna, M., & Innes, R. (2009). Voluntary Pollution Reduction Programs, Environmental Management, and Environmental Performance: An Empirical Study. *Land Economics*, *85*(4), 692-711. https://doi.org/10.3368/le.85.4.692

Schechter, C. (2020). Staggered Difference in Difference. In. Statalist.

Science History Institute. (2017, 1 Dec 2017). *Leo Hendrik Baekeland*. Science History Institute. Retrieved 05.05.2023 from https://www.sciencehistory.org/historical-profile/leo-hendrik-baekeland

Sethi, S. P., & Schepers, D. H. (2014). United Nations Global Compact: The Promise–Performance Gap. *Journal of Business Ethics*, *122*(2), 193-208. https://doi.org/10.1007/s10551-013-1629-y

Stock, J. H., & Watson, M. W. (2020). *Introduction to econometrics* (Fourth edition.; Global edition. ed.). Pearson.

Toxic-Free Future. (2023, n.d.). *Persistent, Bioaccumulative and Toxic Chemicals (PBTs)*. Toxic-Free Future. Retrieved 05.05.2023 from https://toxicfreefuture.org/toxic-chemicals/persistent-bioaccumulative-and-toxic-chemicals-pbts/

United Nations Global Compact. (2023, 11 May 2023). *Our Participants*. United Nations Global Compact. https://unglobalcompact.org/what-is-gc/participants

United Nations Global Compact. (n.d.-a, n.d.). *The Ten Principles of the UN Global Compact*. United Nations GLobal Compact. Retrieved 05.05.2023 from https://unglobalcompact.org/what-is-gc/mission/principles

United Nations Global Compact. (n.d.-b, n.d.). *What's the commitment?* United Nations Global Compact. Retrieved 05.05.2023 from https://unglobalcompact.org/participation/join/commitment

United Nations Global Compact. (n.d.-c, n.d.). *The world's largest corporate sustainability initiative*. United Nations Global Compact. Retrieved 11.05.2023 from https://unglobalcompact.org/what-is-gc

Vidovic, M., & Khanna, N. (2007). Can voluntary pollution prevention programs fulfill their promises? Further evidence from the EPA's 33/50 Program. *Journal of environmental economics and management*, *53*(2), 180-195. https://doi.org/https://doi.org/10.1016/j.jeem.2006.10.001

World Health Organization. (2016, 4 Oct 2016). *Dioxins and their effects on human health*. World Health Organization. Retrieved 05.05.2023 from https://www.who.int/news-room/fact-sheets/detail/dioxins-and-their-effects-on-human-health

Zhang, R., & Khanna, N. (2020). *Public Pressure, Greenwashing and Heterogeneous Effects of Voluntary Pollution Abatement*.

Zhou, R., Bi, X., & Segerson, K. (2020). Evaluating Voluntary Environmental Programs with Spillover Effects. *Journal of the Association of Environmental and Resource Economists*, *7*(1), 145-180. https://doi.org/10.1086/705828