

DISCUSSION PAPER

258

**MEASURING THE IMPACT OF
THE OECD GUIDELINES FOR
MULTINATIONAL COMPANIES**

**BERNARDO MUELLER
FELIPE YUDI ROVIELLO**



MEASURING THE IMPACT OF THE OECD GUIDELINES FOR MULTINATIONAL COMPANIES

BERNARDO MUELLER¹
FELIPE YUDI ROVIELLO²

1. Professor at the Department of Economics of the University of Brasilia (UnB). E-mail: <bmueller@unb.br>.

2. Master's student at the Department of Economics of UnB. E-mail: <feli-peroviello@gmail.com>.

Federal Government of Brazil

Ministry of Economy
Minister Paulo Guedes

ipea Institute for Applied
Economic Research

A public foundation affiliated to the Ministry of Economy, Ipea provides technical and institutional support to government actions – enabling the formulation of numerous public policies and programs for Brazilian development – and makes research and studies conducted by its staff available to society.

President

CARLOS VON DOELLINGER

Director of Institutional Development

MANOEL RODRIGUES JUNIOR

**Director of Studies and Policies of the State,
Institutions and Democracy**

FLÁVIA DE HOLANDA SCHMIDT

Director of Macroeconomic Studies and Policies

JOSÉ RONALDO DE CASTRO SOUZA JÚNIOR

**Director of Regional, Urban and Environmental
Studies and Policies**

NILO LUIZ SACCARO JÚNIOR

**Director of Sectoral Studies and Policies
of Innovation and Infrastructure**

ANDRÉ TORTATO RAUEN

Director of Social Studies and Policies

LENITA MARIA TURCHI

**Director of International Studies, Political
and Economic Relations**

IVAN TIAGO MACHADO OLIVEIRA

Head of Press and Communication

ANDRÉ REIS DINIZ

OMBUDSMAN: <http://www.ipea.gov.br/ouvidoria>
URL: <http://www.ipea.gov.br>

Discussion Paper

A publication to disseminate the findings of research directly or indirectly conducted by the Institute for Applied Economic Research (Ipea). Due to their relevance, they provide information to specialists and encourage contributions.

© Institute for Applied Economic Research – ipea 2021

Discussion paper / Institute for Applied Economic
Research.- Brasília : Rio de Janeiro : Ipea, 1990-

ISSN 1415-4765

1. Brazil. 2. Economic Aspects. 3. Social Aspects.
I. Institute for Applied Economic Research.

CDD 330.90.908

Ipea publications are available for free download in PDF (all) and EPUB (books and periodicals).

Access: <http://www.ipea.gov.br/portal/publicacoes>

The opinions expressed in this publication are of exclusive responsibility of the authors, not necessarily expressing the official views of the Institute for Applied Economic Research and the Ministry of Economy.

Reproduction of this text and the data contained within is allowed as long as the source is cited. Reproduction for commercial purposes is prohibited.

JEL: G3; G14; M14.

DOI: <http://dx.doi.org/10.38116/dp258>

CONTENTS

ABSTRACT

1 INTRODUCTION	6
2 THE OECD GUIDELINES FOR RESPONSIBLE BUSINESS CONDUCT	8
3 ESTIMATING THE IMPACT OF THE OECD GUIDELINES	15
4 DATA	20
5 ESTIMATION OF THE IMPACT OF THE OECD GUIDELINES ON THE DENOUNCED FIRMS	21
6 DISCUSSION AND CONCLUDING REMARKS	39
REFERENCES	40

ABSTRACT

The OECD Guidelines for the responsible business conduct of Multinational Enterprises is one of the largest corporate social responsibility programs (CSR) in the world. By February 2021, 501 allegations of misconduct by multinational companies have been brought to the attention of the National Contact Points (NCP) of its signatory countries. These cases have resulted in a varied sample of experiences which can be used to quantitatively identify the impact on a company of being denounced within this framework. This paper describes an approach to measure the impact of the OECD Guidelines on the affected enterprises. Because adherence to the Guidelines is voluntary, these do not have judicial force or official sanctioning mechanisms. Their efficacy depends on the impact that they have on companies' reputation and public image. Measuring this impact is not straightforward, yet a large literature on Corporate Social Responsibility has developed many methods to detect and quantify how the public's perception of companies' behavior and choices in areas not directly linked to their core business can be measured. In this paper we use several variants of event studies as methods to detect potential costs imposed on a multinational company of having an allegation of misconduct brought before a NCP. These methods use stock market data to detect stock price if there are oscillations around the time of the denunciation. If the companies' returns suffer due to having alleged misconduct brought before this mechanism, it is reasonable to assume that the Guidelines and NCPs are effective. The variations revolve on what is used as a counterfactual to the observed market oscillations. We consider three counterfactuals: i) a market model; ii) a sample of non-denounced companies matched along a set of covariates; and iii) the firms' returns over an interval of placebo-dates, when they were not denounced. We find clear evidence that in many cases the firms' exposure through the program led to a statistically significant decrease in returns at the time of the denunciation. These effects varied, however, across firms, sectors and countries.

Keywords: responsible business conduct; ESG; OECD; governance; event studies.

1 INTRODUCTION

In this report we seek to measure the impact of the OECD Guidelines for Responsible Business Conduct as implemented by the National Contact Points (NCPs) of the member countries. This program has been in place for over 20 years and currently counts with the adherence of 49 countries (OECD, 2016). The Guidelines have been revised several times with the latest and most important revision in 2011, leading to the growth of the program in terms of cases, coverage and types of grievances addressed. The Guidelines are a form of “soft law” that requires the denounced companies to voluntarily accept to participate in the mediation process. This relies on the companies’ perception that non-participation can affect their public image and corporate reputation, with possible consequences for things they care about such as sales, profit, employee morale and attracting talent. Most cases brought to the program are of an ESG nature (Environmental, Social and Governance – ESG), making it one of the largest Corporate Social Responsibility (CSR) programs in the world.

There is some evidence that these instruments have had an impact on companies’ behavior, such as the case of against the British mining company Vedanta accused of infringing holy ground in India (Nieuwenkamp, 2013, p. 171). A grievance was brought before the British NCP by a group of NGOs and, given the refusal of the company to engage, led to a sale of shares by institutional investors. Another example is the specific instance brought against the Royal Shell Company in 2011 by three NGOs in relation to oil spills in the Niger Delta. During the mediation process the company requested “a guarantee from the NGOs that these cases would not be part of a campaign or appear in a public domain”. The NGOs did not agree to this request of confidentiality because “campaigning is at the heart of their actions”.¹ According to Nieuwenkamp (2013, p. 174) “many investors, like the Norwegian Oil Fund or the US Investment Fund Calvert, use the OECD Guidelines for their investments and take NCP statements very seriously”.

Although several such cases can be found, there is very little systematic evidence of how effective the Guidelines as a whole really are. Several authors have tried to perform a more rigorous evaluation. Ruggie and Nelson (2015), for example, provide a survey of several cases

1. See *Oil spills in the Niger Delta* case at: <<https://bit.ly/3xyP94S>>.

DISCUSSION PAPER

and analyze how the implementation has varied over time. Weissman et al. (2016) attempt to address the magnitudes of costs and benefits of due diligence for businesses using a large literature review and proposed a survey that could be used to collect directly from the companies. Thorstensen, Cortellini and Gullo (2018) give an overview and detailed breakdown the data from the OECD specific cases database. Davarnejad (2011) presents a content analysis of 57 published final statements by NCPs reporting on specific instances. PwC (2018) undertake seven in-depth case studies focusing on strategies and instruments. In general, these evaluations find mixed results with great variability of practices and even of understanding and interpretation of what the Guidelines entail and what is the role of the NCP. The great flexibility of how grievances are addressed was generally found to have both advantages and disadvantages, though several commentators recommended the need for further systematization and dissemination of the program.

These studies provide valuable descriptions of the Guidelines' impact on the affected companies, yet there remains the need for a more quantitative assessment to complement the case studies and survey evidence. In this report we use a method which is frequently used in CSR studies, but adapt the methodology to explicitly ensure that we are capturing causal effects. Whereas in typical CSR studies the decision to participate is initiated by the firm, in the case of the Guidelines it is imposed on the firm by a third agent, such as an NGO or a union, which brings forth a grievance to a NCP. This difference might play in favor of the use of event studies as it makes it easier to pinpoint the date of the event, which is often tricky in event studies. The benefit of this methodology is that it provides a clear test statistic with confidence intervals – cumulative average abnormal returns – that can be used to rigorously test the hypotheses that the company has been negatively or positively affected by being denounced as well as by its choice to engage or not in mediation.

In event studies, the returns for a company at the time of an event whose impact on a firm one wishes to investigate, is measured to see if they differ from what would have happened had the event not taken place. The standard counterfactual used for this comparison is the market model, where the firm's normal returns are estimated as a function of the market index using a long period of data (at least 120 days) prior to the event. We follow this standard practice, but in addition we repeat the analysis with two different counterfactuals so as to allow us to

test the robustness of the results and to have greater certainty that we are actually capturing causal effects. The first alternative procedure is to match our treated firms with an equivalent number of comparable non-treated firms. This is done through propensity score matching and with the matching done on the book to market ratio, number of employees and sector. The third method we used compares the firm's returns in the event window against all placebo events one year before and one year after the actual event. The expectation is that the actual cumulative abnormal returns should differ from those for the placebo events. In conjunction, these tests provide strong evidence that the OECD Guidelines can have a significant impact on the companies that receive a complaint, suggesting that the program can be an important incentive for responsible business conduct. Not all firms are affected, however, and we identify the sectors and countries where the program seems to have the greatest impact.

The report is structured as follows. In the next section we provide a series of statistics of the countries and cases, which together describe the program and its track record. Section 3 describes the theory and methods for applying event studies and also for the matching and placebo procedures we use. Section 4 describes the stock market data used for the event studies. Results are presented in section 5 and discussed in section 6.

2 THE OECD GUIDELINES FOR RESPONSIBLE BUSINESS CONDUCT

In this section we provide an overview of the full data-set of 487 cases in the OECD Responsible Business Guidelines for Multinational Enterprises.² This analysis is purely descriptive and has been done before, for example Thorstensen, Cortellini and Gullo (2018), though here the more recent cases have been added. The purpose is simply to set out the variety of cases covered and explore the heterogeneity across their multiple dimensions. The subsequent sections describe how this data can be analyzed to quantify the impact of the Guidelines on the companies' behavior.

Figure 1 shows the breakdown of cases by themes. It shows that the three main themes of the alleged misconduct brought before the program are employment and industrial relations (23,95%), general policies (23,48%) and human rights (16,28%). These are the proportions for

2. See: <<https://bit.ly/3yCW1hF>>. This was the number of cases in October 2020.

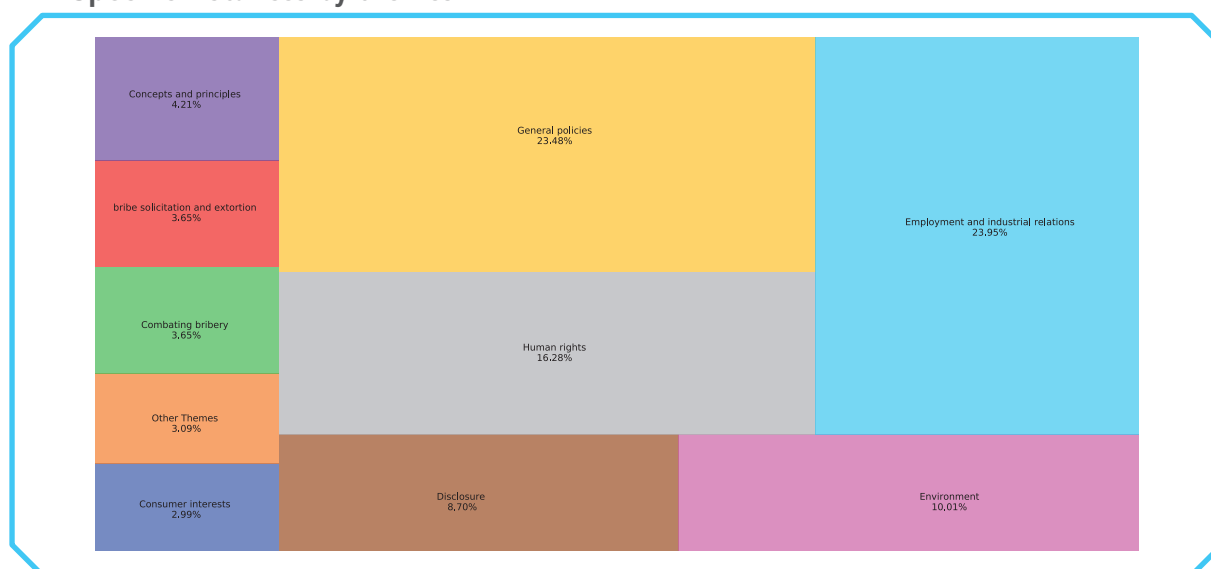
DISCUSSION PAPER

all cases, however, while recent years have seen the growth of human rights and environmental cases. Figure 2 shows the breakdown by sectors, with the top three being manufacturing (24.21%), mining and quarrying (13.61%) and other service activities (12.97%).

Another issue of interest is the incidence of cases per NCP and the incidence per host country. This data is shown in the maps in figures 3 and 4. The five most active NCPs considering all cases are the United Kingdom (55), United States (48), Netherlands (39), Brazil (31) and France (31). The top five host countries that have received more cases are the United States (36), DR Congo (35), Brazil (33), India (25), and the United Kingdom (19). Figure 5 shows a network that links NCP to host countries to show at a glance where the hotspots of action are located. The network does not distinguish which nodes are NCPs and which are host countries, but rather it gives a sense of the level (size of node) and pattern (edges) of the involvement of different countries and also of different pairs of countries.

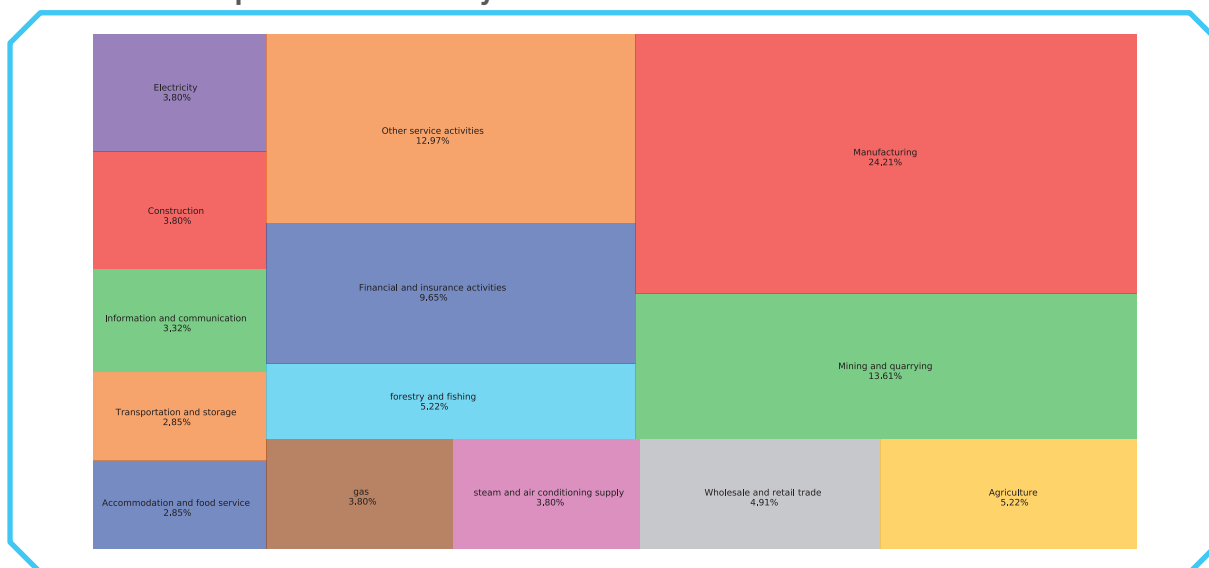
FIGURE 1

Specific instances by themes



Source: OECD. Available at: <<https://bit.ly/3wwx3ii>>.

Publisher's note: Figure whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

FIGURE 2**Number of specific instances by lead NCP**

Source: OECD. Available at: <<https://bit.ly/3wwwx3ii>>.

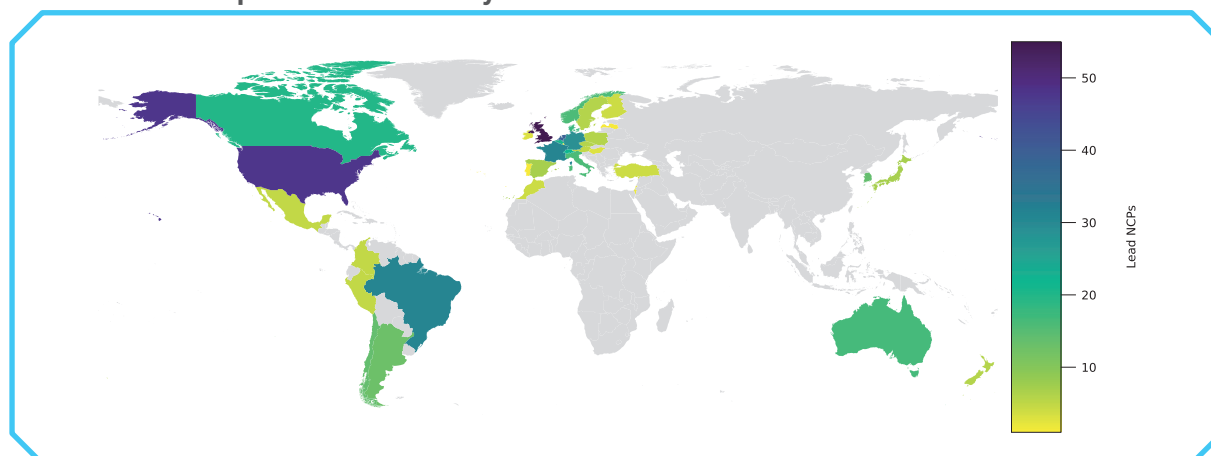
Publisher's note: Figure whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

Who are the main sources of denunciations, that is, the agents that bring the case to the program? Figure 5 shows that NGOs and trade unions are by far the most active sources, followed by individuals with less than half as many. This suggests that efforts to disseminate stakeholders' participation in the Guidelines should both further facilitate the engagement with these active groups as well as seek to engage with underrepresented groups, such as, local governments, business groups and others.

DISCUSSION PAPER

FIGURE 3

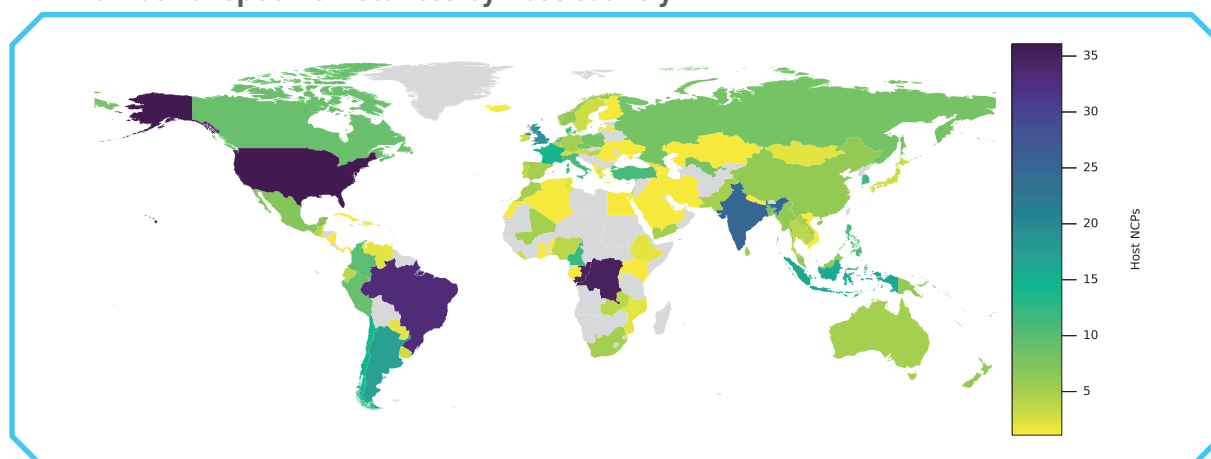
Number of specific instances by lead NCP

Source: OECD. Available at: <<https://bit.ly/3wwwx3ii>>.

Publisher's note: Figure whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

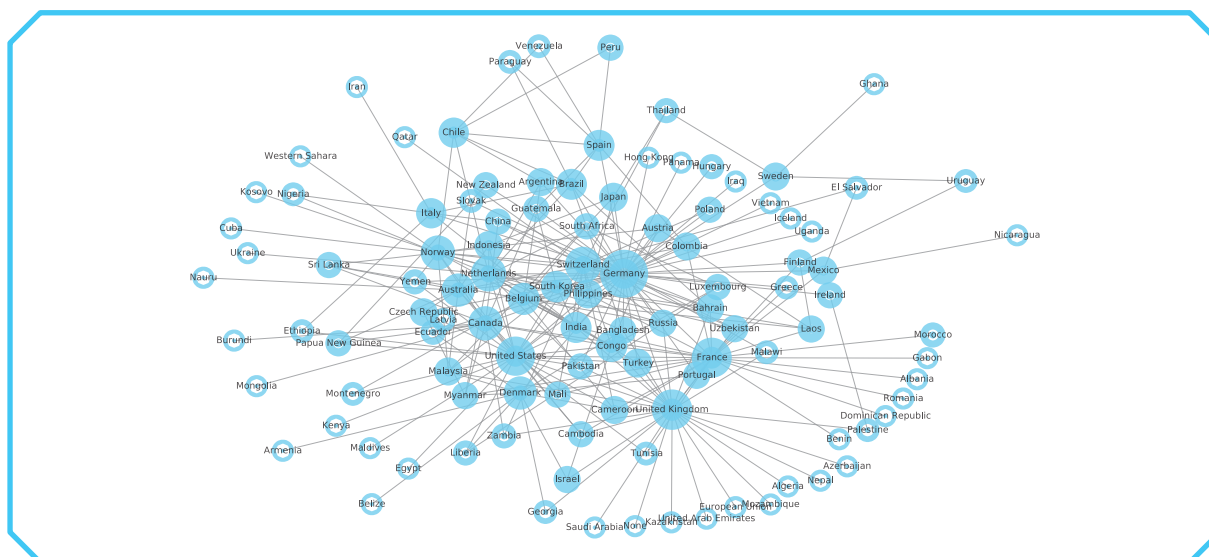
FIGURE 4

Number of specific instances by host country

Source: OECD. Available at: <<https://bit.ly/3wwwx3ii>>.

Publisher's note: Figure whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

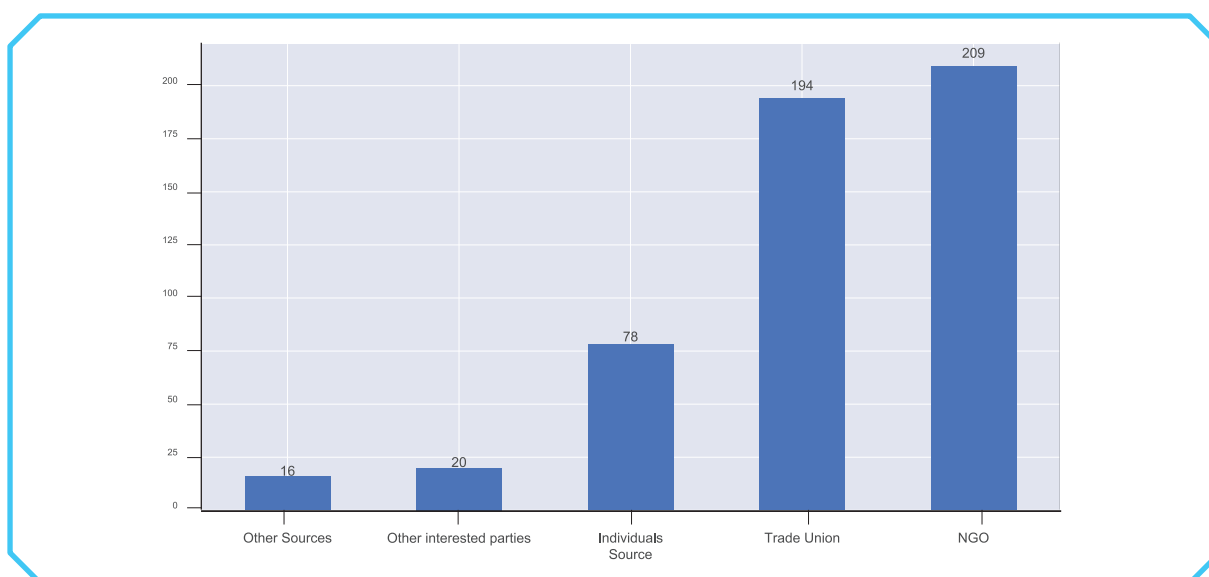
FIGURE 5
Network of NCP and host countries



Source: OECD. Available at: <<https://bit.ly/3www3ii>>.

Publisher's note: Figure whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

GRAPH 1
Sources of cases



Source: OECD. Available at: <<https://bit.ly/3www3ii>>.

Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

DISCUSSION PAPER

One way to ascertain the effectiveness of the program is to see how many of the cases were mediated and how many reached agreements. If a case was mediated this means that the company agreed to engage in the procedure. Some companies refuse to recognize the program and will simply not take part. One of the objectives of this study is to measure if such a choice imposes a cost on the company through corporate social responsibility channels. Another issue is to analyze the proportion of those cases where there was engagement to reach a consensual agreement. Here too we seek to determine whether doing so provides the company with CSR benefits. Graph 2 shows that, of the cases that have been concluded, 55% did not achieve the denounced company's engagement. Of those that did, 28% reached an agreement, 3.5% a partial agreement and 68% no agreement. Note that in some instances a case is not mediated because the issue is taken to some other mediating instance, such as the courts. Graph 3 provides a snapshot of the status of the full portfolio of cases as of November 2020. Over the life of the program a third of the cases brought to the various NCPs were not accepted. Of the 66% of cases that were accepted, 58% have been concluded and 7.4% are still in progress. This suggests that the program has been keeping a good pace in recent years. There were 19 cases in 2020 despite the global pandemic and also 19 cases in 2019.³

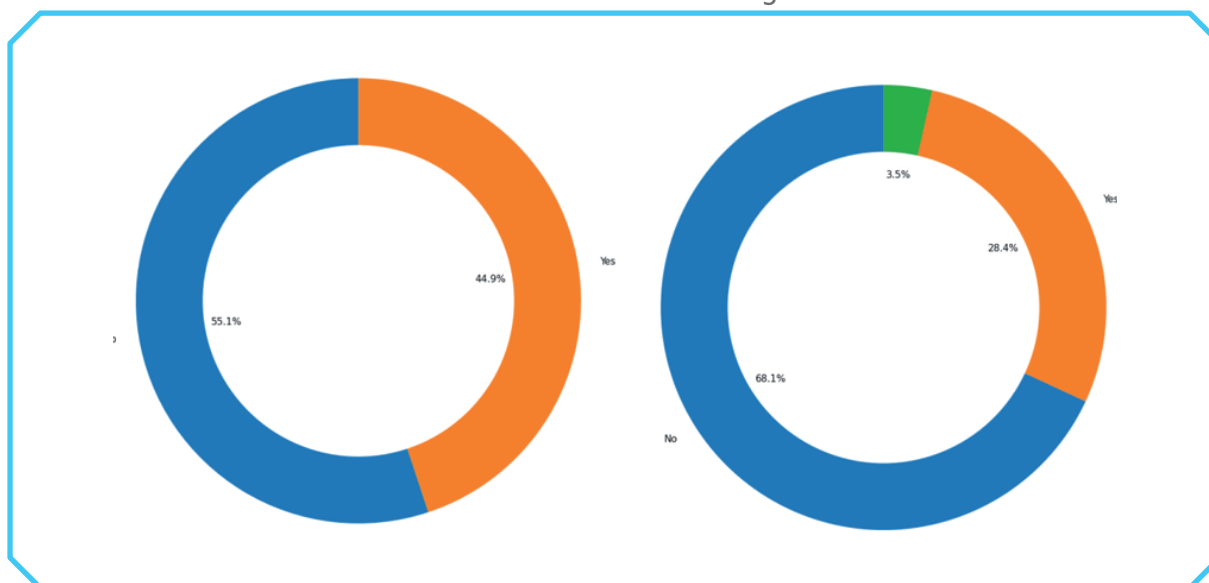
3. See Ruggie and Nelson (2015) for a comparison of the cases and their treatment before and after the 2011 Guideline revision.

GRAPH 2**Mediation and agreement**

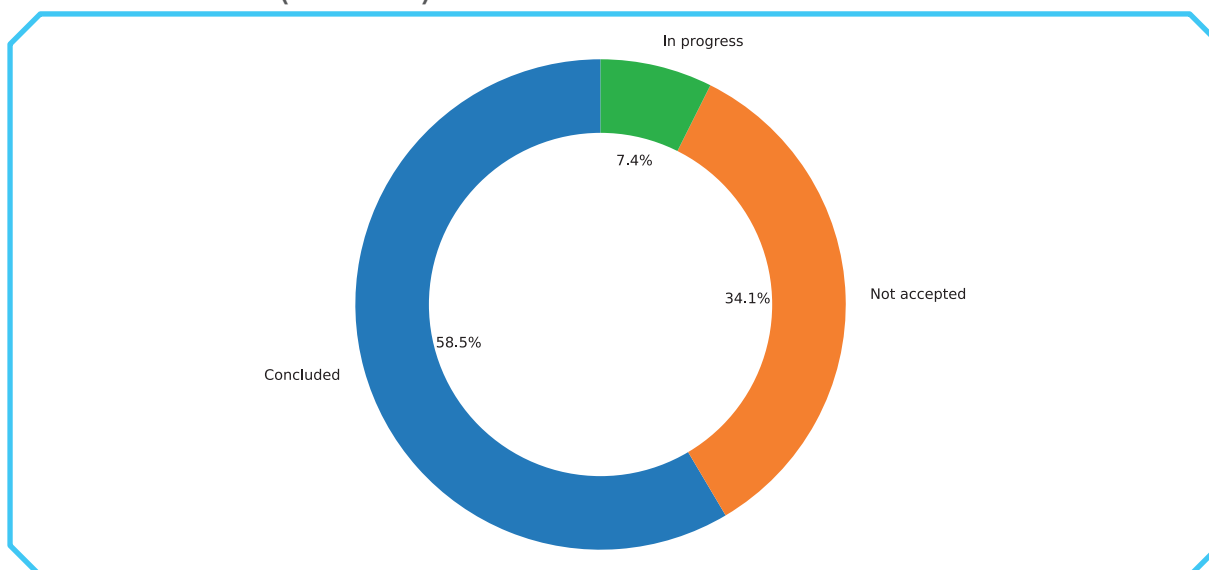
(In %)

2A – Mediation – Concluded cases

2B – Agreements – Concluded cases

Source: OECD. Available at: <<https://bit.ly/3wwwx3ii>>.

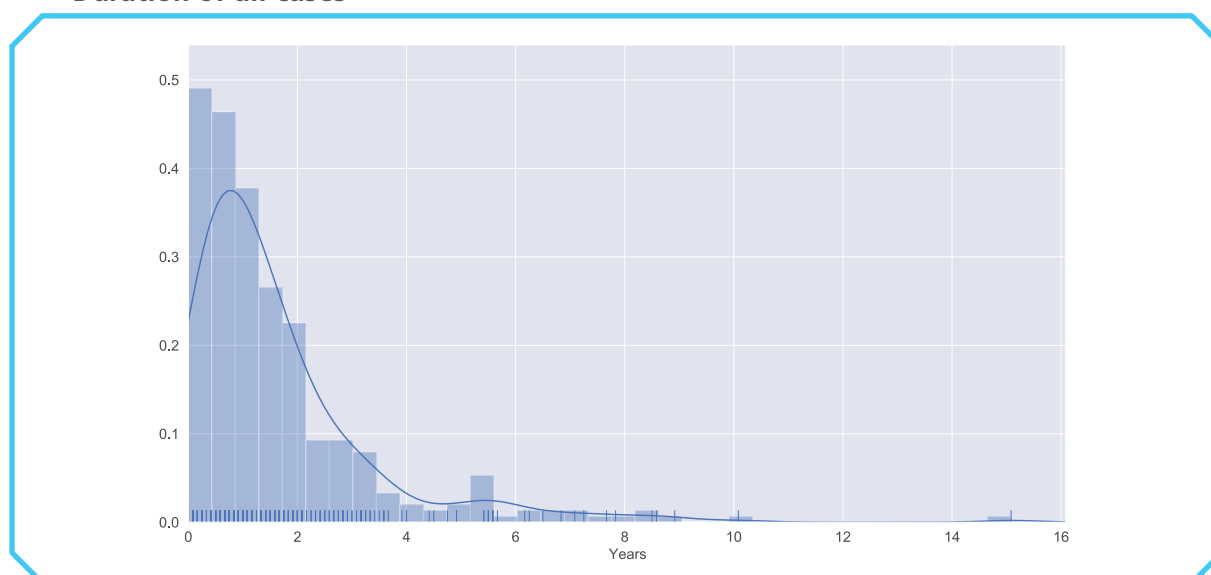
Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

GRAPH 3**Status of cases (Nov. 2020)**Source: OECD. Available at: <<https://bit.ly/3wwwx3ii>>.

Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

DISCUSSION PAPER

Finally, graph 4 provides information on the distribution of the duration of the cases. More than half of the cases are closed in less than two years. Nevertheless, it is not uncommon for cases to drag on for more than 3 or 4 years. The distribution for concluded cases shows greater duration, as it does not include those immediately discarded by the NCPs. It is important to bear in mind, however, that although many cases do drag on for long periods, “court proceedings and quasi-judicial international and regional systems can take even longer” (Ruggie and Nelson, 2015, p. 20).

GRAPH 4**Duration of all cases**

Source: OECD. Available at: <<https://bit.ly/3wwwx3ii>>.

Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

3 ESTIMATING THE IMPACT OF THE OECD GUIDELINES

3.1 Traditional event study with the market model as counterfactual

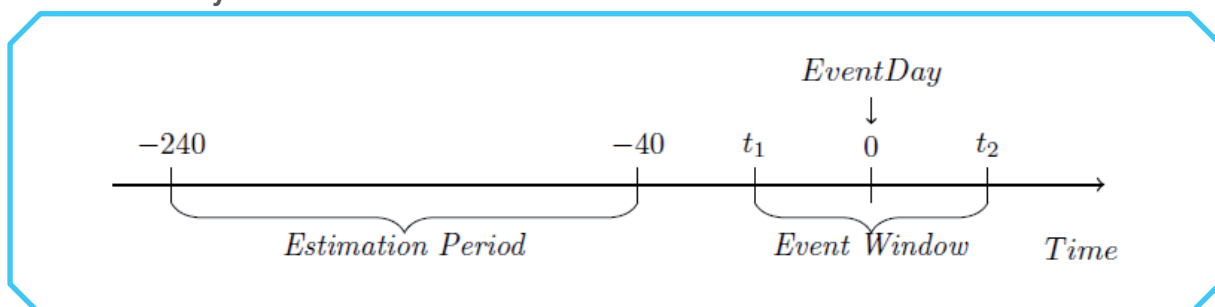
Event studies rely on the hypothesis that stock markets perfectly reflect all the available information about firms through their stock prices (Fama, 1970). If this is the case, it is possible to analyze how a particular event impacts a firm's future performance by measuring the impact

of the event on the firm's stock prices. As an example, suppose an oil company announces the discovery of a new oil field. Security markets should reflect this new discovery and stock prices for this company should increase around the day the discovery is announced. On the other hand, negative announcements, such as oil spills, should affect the stock prices negatively, leading to negative returns.

One could only look at real stock returns to analyze the impact of an event on a certain firm, but this approach can be misleading, as the stock performance of a specific firm is usually correlated to the performance of the stock market as a whole. Event studies aim at analyzing stock price movements at the time a specified event takes place by calculating abnormal returns, that are estimated by subtracting the expected returns of a stock from the observed returns. Most studies use the "market model" approach for obtaining abnormal returns. We also follow this procedure, but in addition use propensity score matching and a placebo approach to test the robustness of our original results.

In the market model, expected returns are hypothetical returns that would be observed if the event had not happened (Kothari and Warner, 2007). They can be obtained by regressing a security's return against the return of a market index using a sample of returns before (or after) the event date. After that, abnormal returns can be calculated and tested. Abnormal returns can also be summed over a specific time window and averaged across firms to create different measures of an event's impact on stock prices. Figure 6 demonstrates this procedure.

FIGURE 6
Event study



Authors' elaboration.

Publisher's note: Figure whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

DISCUSSION PAPER

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + \varepsilon_{i,t} \quad (1)$$

In equation 1, $R_{i,t}$ represents the returns from security i at day t , $R_{m,t}$ is a market index return for that same day and $\varepsilon_{i,t}$ is the error term. This study uses a window of $[-240;-40]$ days before the event as the estimation period. The expected values for the returns for firm i in day t are given by equation 2.

$$\hat{R}_{i,t} = \hat{\alpha}_i + \hat{\beta}_i R_{m,t} \quad (2)$$

Abnormal returns (AR) are simply the difference between the real returns and the expected returns:

$$AR_{i,t} = R_{i,t} - \hat{R}_{i,t} \quad (3)$$

Abnormal returns might be associated with a different day before or after the event day. Furthermore, certain events can have an impact over a specific period of time (referred here as "event window"). In order to take such factors into account, one can add up the Abnormal returns obtained in equation 3 to obtain Cumulative Abnormal Returns (CAR). Equation 4 demonstrates this procedure, where t_1 is the initial date for the event window and t_2 the final date. Commonly used event windows include two or three-day intervals, with $t_1 = -1$ and $t_2 = 0$ or $t_2 = 1$.

$$CAR(t_1, t_2)_i = \sum_{t=t_1}^{t_2} AR_{i,t} \quad (4)$$

Event studies usually contain observations over multiple firms. Cumulative Average Abnormal Returns CAAR can be calculated to measure the impact of an event that affects multiple firms.

$$CAAR = \frac{1}{n} \sum_{i=1}^n CAR(t_1, t_2)_i \quad (5)$$

After estimating the CAARs, the next step is to test the null hypothesis $H_0: CAAR = 0$. Traditionally, a t-test was used, but its low led to the development of many different tests, as well as a branch of the literature dedicated to the discussion of the most appropriate tests (Gelbach, Helland and Klick, 2013). Event study tests can be classified into two different

categories: parametric and non-parametric tests. Parametric tests rely on the assumption that returns are normally distributed, whereas non-parametric tests do not depend on the distributional properties of the returns. To control for the possibility of non-normal returns, this study uses both types of tests.

In particular, the parametric test chosen was the Boehmer test, as it adjusts for event-induced increases in the variance of abnormal returns (Boehmer, Masumeci and Poulsen, 1991). The non-parametric test chosen was the GRANKT test, or generalized rank test. Besides being distribution free and having more power on small samples, this test has other advantages, such as robustness to induced volatility by a specific event and autocorrelation on the abnormal returns (Kolari and Pynnonen, 2011).

3.2 Propensity score matching as counterfactual

Matching techniques are used to estimate the effect of a treatment (in our case the OECD program) that is not randomly assigned. They are particularly useful when the treated individuals have characteristics that determine whether or not they receive the treatment. By using matching, one can generate a counterfactual sample that acts as a control group and compare statistics from this new sample to the one that received the treatment (Heckman, Ichimura and Todd, 1997; Rosenbaum and Rubin, 1983; Reese Junior and Robins, 2017).

Matching is used here to select a synthetic sample used in a robustness test. For each company denounced after 2011, the goal is to select a non-denounced company that is similar to one in our sample. After finding a match for each company, we run an event study using the matched firms' returns and the event dates from the original firms. In this subsection, we refer to the denounced companies as "treated", and to the non-denounced companies as "untreated". The untreated companies' sample is composed of all the companies we could find in the Yahoo Finance website that are traded in the United States. For this reason, this test is only applicable to treated companies traded in US markets.

In this study, we use Propensity Score Matching (PSM) as the matching algorithm. PSM uses a logit model and predictors to obtain probabilities (scores) that tell us how similar are the

DISCUSSION PAPER

companies that were not treated to the ones that were treated. In particular, the predictors used are the number of full-time employees and the book-to-market ratio. To prevent the matching of companies from different sectors, we run a different PSM model for each sector. We also use a 3 nearest-neighbors approach, meaning that we assign 3 non-treated companies (ordered by score) to each treated company. This is important because there is no guarantee that returns for the matched company are available for the same period as they are for the treated company. Therefore, if no data is available for the first matched company, we use returns from the second one. If no data is available for the second one, the third one's returns are used.

3.3 Placebo event dates as counterfactual

The second robustness test is a placebo event study. The basic idea behind the test is to run multiple even studies for the same companies using different dates as the event day. This generates a distribution of CAARs that we use to compare to the position of the "real" CAAR (i.e. the CAAR obtained when the real event date is used). This test can be interpreted as the opposite of the previous one. In the PSM test, we were looking for a matching company to estimate CAARs using the same day treated companies were denounced. Here, we already have a match (the best match for a company is itself), and we want to know what happens to the CAARs in days other than the real event day. If the real CAAR is significant and negative, we would expect it to be in the lower tail of the CAAR distribution.

We run two versions of the placebo test. In the first one, the placebo event days for each company are all the business days one year before and one year after the real event date. Table 1 demonstrates this procedure. We then obtain 730 placebo CAARs for each estimation window, that we compare to the real CAAR (which is the one obtained when the dates from the column Date are used). The method used in the second version of the test is the same one used in the first version, but instead of using days relative to the real event date, the days we select for each company are chosen randomly from a uniform distribution. More specifically, we randomly sample, with replacement, 500 days in a time window of one year before and one year after the real event date. To avoid complications related to different business days in different countries, we only use companies traded in US markets.

TABLE 1
Placebo event dates

Ticker1	Date-365	...	Date-2	Date-1	Date	Date+1	Date+2	...	Date+365
RIO	-	...	-	-	2020/09/28	2020/09/29	2020/09/30	...	-
UBS	-	...	-	-	2020/06/22	2020/06/23	2020/06/24	...	-
TLPFY	2019/04/18	...	2020/04/15	2020/04/16	2020/04/17	-	-	...	-
VALE	2019/02/14	...	2020/02/12	2020/02/13	2020/02/14	-	-	...	-
...
ADDYY	-	...	2001/06/29	-	2001/07/01	2001/07/02	2001/07/03	...	2002/07/01
SIEGY	2000/06/05	...	-	2001/06/04	2001/06/05	2001/06/06	2001/06/07	...	2002/06/05
MAKSY	-	...	-	-	2001/05/14	2001/05/15	2001/05/16	...	2002/05/14
MAKSY	-	...	-	-	2001/04/02	2001/04/03	2001/04/04	...	2002/04/02

Source: Data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

To compare the real CAARs to the placebo ones, we first set all non-significant CAARs to 0 based on the Grankt test with a significance level of 5%. For each estimation window, we then calculate a “weak” score, that corresponds to the definition of a cumulative distribution. A weak score of 50% means that 50% of the values are less than or equal to the real CAAR. In the third step, we compute a “strict” score, that is similar to the weak one, but only takes into account CAARs that are strictly less than the real CAAR. Finally, we average the “weak” and the “strict” scores to obtain the statistic used in this test.⁴

4 DATA

The main components of an event study are security returns and an event date. In this study, the event date is the day that OCED received an allegation against a company or the day the Final Statement was published. This information is readily available in the OECD databases. The denounced companies stock returns, however, are not as easy to obtain. The OECD specific cases database contains around 500 allegations of misconduct of multinational

4. The score used in this test uses the same approach as the “mean” score in scipy’s percentileofscore method. Available at: <<https://bit.ly/2U1VuXU>>.

DISCUSSION PAPER

companies. In order to obtain all denounced company names, we first read all cases summaries and (if necessary) the Initial Statements. However, the names of the denounced companies are not always revealed due to confidentiality clauses. Furthermore, not all companies are traded in stock markets. Therefore, the dataset only contains stock returns from firms that could be identified, that were listed on the Yahoo Finance website (where the stock prices were taken from) and that were traded in stock markets at least 240 days before the OECD received an allegation against it.

Event studies usually use returns from securities traded in American markets. However, allegations brought to the OECD often include non-American companies, many of which are not listed in American stock exchanges. To investigate how different markets react to the same event, this study uses two different sets of stock returns. The first one, called ticker1, contains the stocks of all denounced firms we could find in the website Yahoo Finance for securities traded in US markets (including ADRs). The second one, called ticker2, contains all the stocks we could find in the same website for securities traded in the country where a denounced firm's headquarter is located. For example, according to Yahoo Finance, Rio Tinto's ticker for stocks listed in the NYSE is RIO. This company, however, is based in London, where the symbol for its stocks (listed in the LSE) is RIO.L. A complete list of the tickers used can be found in table 10. Events concerning ticker1 stocks add up to a total of 139 events, while ticker2 stocks contain 189 events.

Market index returns also vary according to the country in which a security is traded. For US markets, the S&P 500 was used. A complete list of the market indexes used in this study can be found in table 11.

5 ESTIMATION OF THE IMPACT OF THE OECD GUIDELINES ON THE DENOUNCED FIRMS

In this section we present the results of the tests described in section 3, which we use to quantify the impact of the OECD Guidelines for Multinational Enterprises on firms that are brought to the attention of the country NCPs. We have three different tests, which we apply to two sets of stock market prices, for two different sample periods, for the date of the event

(denunciation to the OECD) as well as for the date of the final statement when the case was closed.⁵ In addition, we estimate the impact not only for the full set of firms, but also breaking them down by continent as well as by sector. Each set of results are presented for 6 different event windows to allow for different impact dynamics. There are thus a large number of results, which we present in eight tables. The first four tables (tables 2 to 5) refer to the date of the original event and the last four (tables 6 to 9) to the date of the final statement. Even numbered tables (tables 2, 4, 6 and 8) present the market model estimations in columns 1 and 2, and the propensity score matching results in column 3. Columns 4 to 8 present the market model results (post-2011) by continent. The odd numbered tables (tables 3, 5, 7 and 9), show the results for the post-2011 market model broken down by sector. It is important to have such a detailed view of the estimated impacts because there is no reason why the impact would be the same along any of these dimensions. As we show, there is some interesting variation across continents and sectors. The results for the placebo date tests are presented subsequently.

5.1 Market model and propensity score matching results

To understand the results in tables 2 to 9 it is useful to remember the null hypothesis that is being tested. This null hypothesis is that the initiation of a case against a firm within the OECD Guidelines for Multinational Enterprises does not impact the enterprise's stock-market returns within the specified event window. A negative and statistically significant estimate in these tables thus means that this hypothesis is refuted, and we accept that the Guidelines do have an impact that can provide incentives for firms to amend and compensate for their behavior, as well as a demonstration effect for other firms to avoid similar misconduct.

The first three columns of table 2 provide a good summary of our main result. This table shows the CAARs estimated using securities traded in American markets, including ADRs (ticker1). Column Full Sample contains all events in which we could identify the denounced firm, whereas Column Post-2011 was estimated using a sub sample of firms denounced after 2011. Column Propensity Score Matching 2011 contains estimations for the matched firms (obtained using Propensity Score Matching) denounced after 2011. The estimates for all six

5. Most results are for the post-2011 sample. This is because the program underwent a large reform at this date, when it acquired its current configuration. Also, the case data is more complete after 2011.

DISCUSSION PAPER

windows in these columns are negative, and 13 of 18 are statistically significant at least at 10%. Windows $[-2;0]$, $[-2;1]$ and $[-2;2]$ in particular seem to best capture this negative impact, suggesting that the news of firms' referral to the Guidelines is often available two days before the official date in the dataset. The fact that the results are robust across two different counterfactuals (market model and PSM) provides some confidence in the results.

Columns 4 to 8 of table 2 show that the Guidelines are especially effective in the USA and Canada, as 5 of the 6 event windows display negative and significant estimates.⁶ In Europe and Asia there is also evidence of a significant negative impact, though only for windows $[-3;1]$ and $[-1;0]$, respectively. For Oceania (windows $[-1;0]$ and $[-1;1]$) and Latin America ($[-3;1]$), however, the estimated impacts are positive. This result is likely because, unlike Europe and USA & Canada, these two continents have fewer firms with complete data in the sample. If those cases are concentrated in a sector that is positively affected by the Guidelines (such as Financial Services, see below), that would explain these results.

Table 3 shows the breakdown of the market model (post-2011) event study by sector. The results can be grouped into four classes of impact. Basic Materials, Consumer Defense and Energy displayed a strong negative impact from alleged misconduct being brought to the Guidelines. This greater impact may be due to the fact that the indemnities to be paid in these are typically larger. Healthcare and Industrials also yielded negative and significant impacts, though slightly weaker than those previous sectors. The estimated impacts for Utilities and Consumer Cyclical were also negative, but not statistically different from zero. Interestingly, Financial Services exhibit a positive impact for a firm being denounced to the Guidelines. It is as if shareholders were actually pleased that the banks and financial companies involved were not following the rules of proper environmental, social and governance criteria, and instead pursued only shareholder value. It is beyond the scope of this report to investigate this interpretation. It is uncanny, however, that of all the sectors, only financial services displayed a positive estimate.

6. The division by continents was based on the country where a company is headquartered, not where it is traded. The sectors are not the same as used by the OECD, as the same instance can be classified into many different sectors. Rather, the sectoral classification is that used by Yahoo Finance.

Tables 4 and are similarly structured to tables 2 and 3, but contain securities traded in the native country of the denounced firms (ticker2), slightly increasing the sample size.⁷ Nevertheless, practically all the previous results are replicated. The estimate for the Guideline's overall impact is negative for all windows and statistically significant for $[-2,0]$. Once again, this impact is mainly due to the cases in the USA and Canada. Oceania, Asia and Latin America show evidence of a positive impact, but this is probably reflecting special cases of few firms in specific sectors. In terms of sectors, Basic Materials, Energy, and Consumer Defense show the strongest negative impact. Financial Services in this sample still displays a positive impact, even stronger than in case of ticker1. The robustness of the results in tables 2 and 3 compared to the results in tables 4 and 5, despite the differences in ticker 1 and ticker2 samples, once again provides confidence that the impacts we are estimating are real.

The results above used as the event of interest the date at which the allegation of misconduct was brought to the OECD Guidelines. Another event of interest is the date at which the event is concluded, that is, when a final statement is issued by the NCP, marking the completion of the procedure. This completion might be that a complaint was not accepted, or it might be that an agreement was or was not reached. In practice there is much variability in the format and amount of detail used by the NCPs in their final statements (Davarnejad, 2011), so we do not try to sort the impact according to the outcome. Instead, in tables 6 and 9 we repeat the same analysis we did for the opening date, to the final statement date.

We don't have strong priors as to the expected impact of the final statement. It may be that a final statement is much less conspicuous than the shock and surprise of the initial allegation, in which case no effect would be expected. It may be, however, that reaching a resolution would be seen with favor by the markets and a positive impact to returns would be forthcoming. But it is also possible that the final statement simply brings back to the limelight the fact that the firm had previously been denounced, and this could produce negative returns.

The results in tables 6 to 7, for ticker1, and tables 8 to 9, for ticker2, seem to contain some of each of these interpretations, indicating that further investigation sorting the cases

7. The propensity score estimation was not done here because we could not obtain large samples of untreated firms for different markets from Yahoo Finance.

DISCUSSION PAPER

by outcome would be warranted. In table 6 columns 1 to 3 we find both negative impacts of the final statement (windows $[-1;0]$ and $[-1;1]$) as well as positive impacts (window $[-2;2]$ and $[-3;1]$). The USA and Canada which previously displayed the most impact from the program, no longer do so. Oceania and Latin America, on the other hand, switch from the previous positive impact to negative impacts. Once again this is probably due to the small number of firms for these continent and the fact that those few cases are over-represented in the financial sector, which is now found to be negative (table 7). This switch for financial firms from a positive impact in the initial date to a negative impact in the final statement date is compatible with the interpretation that this sector is not only immune but thrives on being perceived as violating the Guidelines. Energy companies and Utilities, on the other hand, that were punished for being denounced, seem to benefit from reaching a resolution. And Industrials, for some reason, suffer a negative impact both on the initial and on the final date. Explaining these differences is beyond the scope of this report and would involve analyzing the nature of the relationship between firms in these sectors and markets, consumers, advocates, NCPs, regulators, governments, and other stakeholders. The results for final statements in ticker2 markets are similar (tables 8 and 9).

The results discussed above and presented in tables 2 to 9 are given an alternative display in graphs 5 to 8. These graphs highlight the impact of the Guidelines on individual firms. The graphs refer to ticker1 and ticker2 markets and the colors show the classification by continent or by sector. Table 10 lists the ticker names that appear in the graphs. Ticker size varies according to the magnitude of the event impact. The further from the zero-impact line, the higher the impact. The graphs show that there are both positive and negative impacts, yet in general negative impacts dominate. Also, the graphs show that our previous results are due to a subset of all firms, but by no means all firms, as many are close to the zero-impact line.

TABLE 2

CAARs for ticker1 – Full sample, propensity score matching and post-2011 estimations for continents

Window	Full sample	Post-2011	Propensity score matching 2011	Europe	USA and Canada	Oceania	Asia	Latin America
[-1:0]	-0.0016	-0.0039**	-0.0068*	-0.0015	-0.0116*	0.0060**	-0.0147*	0.0132
[-1:1]	-0.0016	-0.0029	-0.0035	-0.0018	-0.0099	0.0141**	-0.0082	0.0115
[-2:0]	-0.0027**	-0.0052**	-0.0092*	-0.0001	-0.0186***	0.0027	-0.0112	0.0063
[-2:1]	-0.0027***	-0.0042**	-0.0060*	-0.0003	-0.0169***	0.0108	-0.0046	0.0045
[-2:2]	-0.0028**	-0.0036**	-0.0122**	-0.0022	-0.0201***	0.0032	0.0000	-0.0032
[-3:1]	-0.0038*	-0.0049***	-0.0055	-0.0046*	-0.0151**	0.0034	-0.0017	0.0267*

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Obs.: 1. The asterisks refer to the results of the Grankt test, where *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Boehemer test results available from the authors.

2. Sample sizes for the window [-1;0] are respectively 139, 95, 81, 55, 23, 3, 7 and 6.

TABLE 3

CAARs for ticker1 – Post-2011 estimations for sectors

Window	Basic materials	Financial services	Industrials	Utilities	Energy	Consumer defence	Consumer cyclical	Healthcare
[-1:0]	-0.0118*	0.0091*	-0.0016	-0.0041	-0.0031	-0.0055	-0.0062	-0.0023
[-1:1]	-0.0077	0.0060	-0.0001*	-0.0032	-0.0043	-0.0041	-0.0068	-0.0036
[-2:0]	-0.0155**	0.0120	-0.0079	0.0023	-0.0029	-0.0056**	-0.0052	-0.0024
[-2:1]	-0.0114	0.0088	-0.0063	0.0031	-0.0040	-0.0042	-0.0058	-0.0037
[-2:2]	-0.0130	0.0036	-0.0052	0.0022	-0.0087*	0.0021	-0.0017	-0.0031
[-3:0]	-0.0080	0.0119	-0.0052	-0.0037	-0.0161*	-0.0073**	-0.0046	-0.0065**
[-3:1]	-0.0039	0.0088	-0.0036	-0.0029	-0.0173*	-0.0060**	-0.0051	-0.0078

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Obs.: 1. The asterisks refer to the results of the Grankt test, where *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Boehemer test results available from the authors.

2. Sample sizes for the window [-1;0] are respectively 25, 9, 9, 3, 16, 10, 9 and 2.

DISCUSSION PAPER

TABLE 4**CAARs for ticker2 – Full sample and post-2011 estimations for continents**

Window	Full sample	Post-2011	Europe	USA and Canada	Oceania	Asia	Latin America
[-1:0]	-0.0037	-0.0058	-0.0043	-0.0161**	-0.0006*	0.0020	-0.0028
[-1:1]	-0.0021	-0.0047	-0.0020	-0.0170**	0.0030	0.0022	-0.0061
[-2:0]	-0.0044**	-0.0061***	-0.0049	-0.0194***	0.0061**	0.0048*	-0.0028
[-2:1]	-0.0015	-0.0030	-0.0006	-0.0160***	0.0093**	0.0056	-0.0061
[-2:2]	-0.0013	-0.0023	0.0036	-0.0178***	0.0124**	-0.0010	-0.0111
[-3:0]	-0.0023	-0.0031	-0.0019	-0.0180***	0.0057	0.0098	0.0084**
[-3:1]	0.0003	-0.0005	0.0014	-0.0153**	0.0074	0.0106	0.0082*

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Obs.: 1. The asterisks refer to the results of the Grankt test, where *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Boehemer test results available from the authors.

2. Sample sizes for the window [-1;0] are respectively 189, 130, 67, 28, 10, 14 and 9.

TABLE 5**CAARs for ticker2 – Post-2011 estimations for sectors**

Window	Basic materials	Financial services	Industrials	Utilities	Energy	Consumer defence	Consumer cyclical	Healthcare
[-1:0]	-0.0103**	0.0021	-0.0082	0.0165**	-0.0116	-0.0048	-0.0078	-0.0053
[-1:1]	-0.0097**	0.0066**	-0.0079	0.0045	-0.0082	-0.0042	-0.0056	0.0012
[-2:0]	-0.0086*	0.0044	-0.0123*	0.0303	-0.0124	-0.0076***	-0.0127	0.0005
[-2:1]	-0.0021	0.0133**	-0.0119	0.0183	-0.0090	-0.0066**	-0.0100	0.0025
[-2:2]	-0.0077	0.0080*	-0.0141	0.0303	-0.0049	-0.0078***	-0.0058	0.0023
[-3:0]	-0.0067	0.0026*	-0.0124	0.0328	-0.0179***	-0.0026	0.0073	-0.0057
[-3:1]	-0.0027	0.0113***	-0.0118	0.0208	-0.0138***	-0.0020	0.0111	0.0007

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Obs.: 1. The asterisks refer to the results of the Grankt test, where *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Boehemer test results available from the authors.

2. Sample sizes for the window [-1;0] are respectively 32, 13, 18, 4, 20, 16, 7 and 3.

TABLE 6

CAARs for ticker1 – Final statement: full sample, propensity score matching and post-2011 estimations for continents

Window	Full sample	Post-2011	Propensity score matching 2011	Europe	USA and Canada	Oceania	Asia	Latin America
[-1:0]	-0.0002**	-0.0004**	-0.0040	0.0000	0.0007	-0.0124*	0.0023	-0.0078*
[-1:1]	-0.0004**	-0.0003**	0.0035	0.0014**	-0.0019	-0.0203***	0.0065	-0.0094*
[-2:0]	0.0027	0.0026	0.0027	0.0024***	0.0040	-0.0123*	0.0120*	-0.0006
[-2:1]	0.0025	0.0027	0.0102	0.0036	0.0015	-0.0202***	0.0162**	-0.0022
[-2:2]	0.0021**	0.0028*	0.0141	0.0020*	0.0064	-0.0253***	0.0030	0.0087
[-3:1]	0.0028	0.0034	0.0061**	0.0050**	0.0022	-0.0166	0.0073	0.0045

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Obs.: 1. The asterisks refer to the results of the Grankt test, where *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Boehemer test results available from the authors.

2. Sample sizes for the window [-1;0] are respectively 125, 106, 68, 58, 28, 4, 8 and 5.

TABLE 7

CAARs for ticker1 – Post-2011 estimations for sectors

Window	Basic materials	Financial services	Industrials	Utilities	Energy	Consumer defence	Consumer cyclical	Healthcare
[-1:0]	-0.0002	-0.0081**	-0.0074***	0.1141	0.0073**	-0.0090	-0.0034	-0.0146
[-1:1]	-0.0024	-0.0093***	-0.0071***	0.1675	0.0058*	0.0019	-0.0036	-0.0029
[-2:0]	0.0104***	-0.0100	-0.0035	0.1336	0.0076	0.0004	-0.0049	0.0034
[-2:1]	0.0082**	-0.0112*	-0.0032	0.1870	0.0060	0.0031	-0.0051	0.0150
[-2:2]	0.0084*	-0.0075	-0.0002	0.1328**	0.0004	0.0031	-0.0140***	0.0238
[-3:1]	0.0050	-0.0111	-0.0009	0.2136	0.0035	0.0035	-0.0015	0.0093

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Obs.: 1. The asterisks refer to the results of the Grankt test, where *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Boehemer test results available from the authors.

2. Sample sizes for the window [-1;0] are respectively 26, 12, 15, 2, 11, 13, 12 and 2.

DISCUSSION PAPER

TABLE 8

CAARs for ticker2 – Final statement: full sample and post-2011 estimations for continents

Window	Full sample	Post-2011	Europe	USA and Canada	Oceania	Asia	Latin America
[-1:0]	-0.0007	-0.0006	0.0011	-0.0035	0.0045	-0.0036	-0.0009
[-1:1]	0.0003	0.0014	0.0045	-0.0054	0.0025	0.0008	0.0009
[-2:0]	0.0008	0.0005	0.0011**	-0.0005	0.0024	-0.0011	0.0031
[-2:1]	0.0018	0.0026	0.0044*	-0.0022	0.0004	0.0028	0.0049
[-2:2]	0.0027*	0.0038	0.0054	0.0047	-0.0041	-0.0072	0.0117
[-3:1]	0.0017**	0.0022	0.0045**	-0.0012	-0.0025	-0.0014	0.0064

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Obs.: 1. The asterisks refer to the results of the Grankt test, where *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Boehemer test results available from the authors.

2. Sample sizes for the window [-1;0] are respectively 137, 109, 61, 27, 3, 11 and 6.

TABLE 9

CAARs for ticker2 – Final statement: post-2011 estimations for sectors

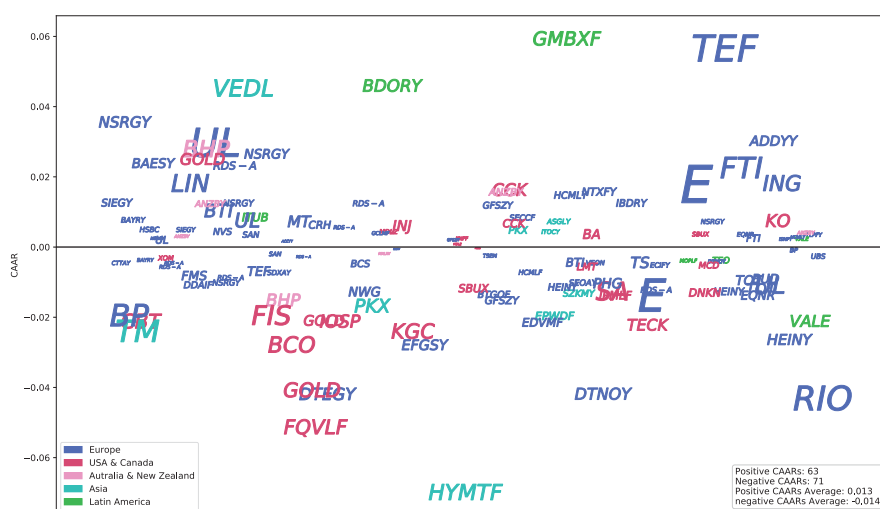
Window	Basic materials	Financial services	Industrials	Utilities	Energy	Consumer defense	Consumer cyclical	Healthcare
[-1:0]	-0.0034	-0.0062***	-0.0035	0.0496	0.0118***	0.0016	-0.0031	-0.0145
[-1:1]	-0.0040*	-0.0032***	0.0000	0.0940	0.0104**	0.0026	-0.0024	-0.0072
[-2:0]	0.0004	-0.0104***	-0.0047	0.0732	0.0130	-0.0033	-0.0013	0.0036
[-2:1]	-0.0003	-0.0074***	-0.0013*	0.1176*	0.0116	0.0008	0.0007	0.0108
[-2:2]	0.0030	-0.0063***	0.0456*	0.1099*	0.0122	0.0020	-0.0079	0.0215
[-3:1]	-0.0020	-0.0070**	-0.0028*	0.1208*	0.0165	0.0018	-0.0002	0.0072

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Obs.: 1. The asterisks refer to the results of the Grankt test, where *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Boehemer test results available from the authors.

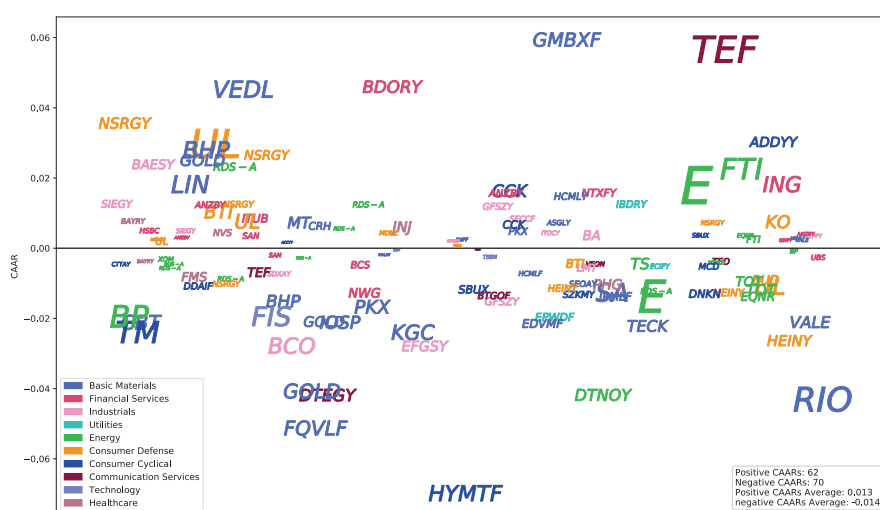
2. Sample sizes for the window [-1;0] are respectively 26, 12, 16, 3, 11, 15, 13 and 2.

GRAPH 5**Individual CAARs by continent – Ticker1**

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

GRAPH 6**Individual CAARs by sector – Ticker1**

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

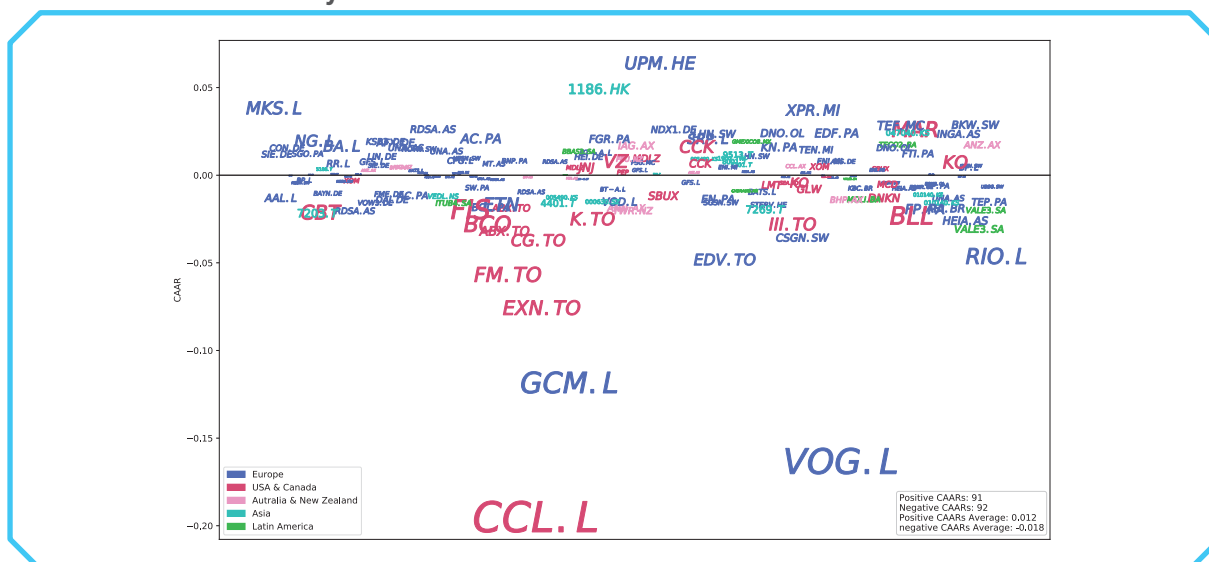
Authors' elaboration.

Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

DISCUSSION PAPER

GRAPH 7

Individual CAARs by continent – Ticker2



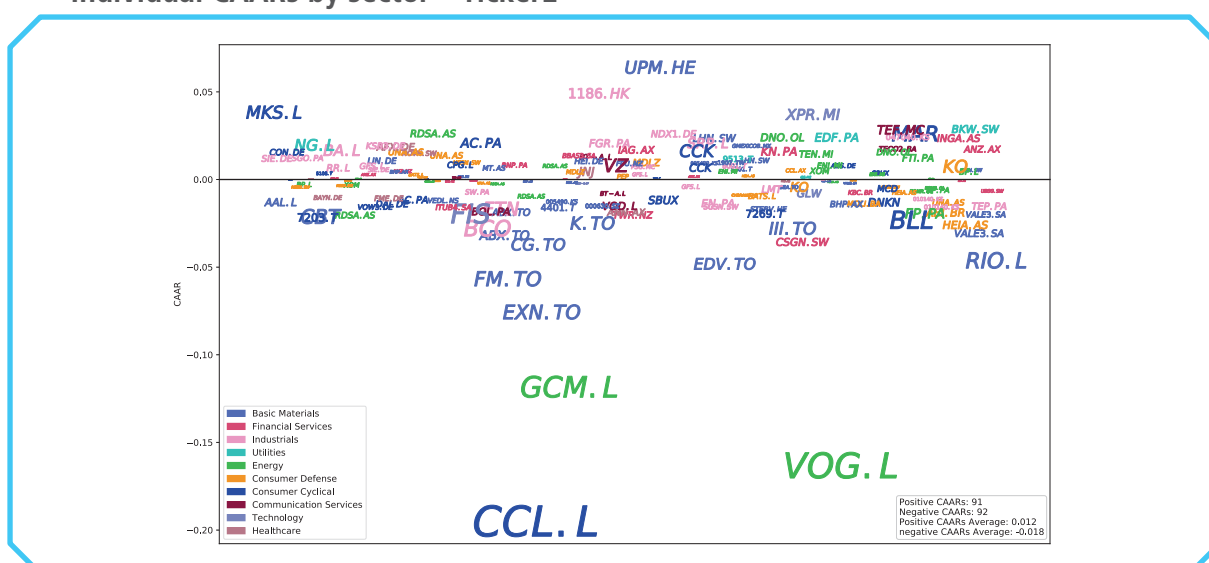
Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

GRAPH 8

Individual CAARs by sector – Ticker2



Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

TABLE 10**Yahoo finance tickers used**

Ticker1	Ticker2	Ticker1	Ticker2	Ticker1	Ticker2
ACYY	AC.PA	ING	INGA.AS	-	000630.SZ
ADDYY	ADS.DE	IOSP	010140.KS	-	-
ANZBY	ANZ.AX	IPMLF	III.TO	-	047040.KS
ANZBY	ANZ.NZ	ISNPY	ISP.MI	-	1186.HK
ASGLY	5201.T	ITOCY	8001.T	-	1907.TW
BA	ITUB	ITUB4.SA	4401.T	-	-
BAESY	BA.L	JINFF	5108.T	-	-
BASFY	BAS.DE	JNJ	JNJ	-	7201.T
BAYRY	BAYN.DE	KGC	K.TO	-	AAL.L
BCO	BCO	KO	KO	-	ACS.MC
BCS	LIN	LIN.DE	AFX.DE	-	-
BDORY	BBAS3.SA	LMT	LMT	-	ANN.AX
BHP	BHP.AX	MALRY	MIN.AX	-	BKW.SW
BP	BP.L	MCD	MCD	-	BLL
BTGOF	BT-A.L	MDLZ	MDLZ	-	BNP.PA
BTI	BATS.L	MOPLF	MOLI.BA	-	BOL.PA
BUD	ABI.BR	MT	MT.AS	-	CCL.AX
CBT	CBT	NSRGY	NESN.SW	-	CCL.L
CCK	CCK	NTXFY	KN.PA	-	CG.TO
CRH	NVS	NOVN.SW	CHDRAUIB.MX	-	CRH
CTTAY	CON.DE	NWG	CPG.L	-	-
DDAIF	DAI.DE	PHG	PHIA.AS	-	CSGN.SW
DNKN	DNKN	PKX	005490.KS	-	EN.PA
DTEGY	PUGOY	UG.PA	EXN.TO	-	-
DTNOY	DNO.OL	RDS-A	RDSA.AS	-	FBU.NZ
E	ENI.MI	RIO	RIO.L	-	GLW
ECIFY	EDF.PA	SA	SEA.TO	-	HEI.DE
EDVMF	EDV.TO	SAN	SAN.MC	-	IAG.AX
EFGSY	FGR.PA	SBUX	SBUX	-	KBC.BR
ELUXY	ELUX-B.ST	SDXAY	SW.PA	-	KSB3.DE

(Continues)

DISCUSSION PAPER

(Continued)

Ticker1	Ticker2	Ticker1	Ticker2	Ticker1	Ticker2
EPWDF	9513.T	SECCF	SRP.L	-	MAR
EQNR	EQNR.OL	SEOAY	STERV.HE	-	MKS.L
ETN	ETN	SIEGY	SIE.DE	-	ML.PA
FIS	FIS	SZKMY	7269.T	-	MYL
FMS	FME.DE	TECK	TECK-A.TO	-	NDX1.DE
FQVLF	FM.TO	TEF	TEF.MC	-	NG.L
FTI	FTI.PA	TEO	TECO2.BA	-	PEP
FWONA	TLPFY	TEP.PA	PSG.MC	-	-
GCLMF	GCM.L	TM	7203.T	-	RR.L
GFSZY	GFS.L	TOT	FP.PA	-	SGO.PA
GMBXF	GMEXICOB.MX	TS	TEN.MI	-	SGSN.SW
GOLD	ABX.TO	TSEM	TWR.NZ	-	-
HCMLF	LHN.SW	UBS	UBSG.SW	-	UPM.HE
HCMLY	LHN.SW	UL	UNA.AS	-	VOD.L
HEINY	HEIA.AS	VALE	VALE3.SA	-	VOG.L
HSBC	0005.HK	VEDL	VEDL.NS	-	VOW3.DE
HYMTF	VEON	VZ	-	-	-
IBDRY	IBE.MC	XOM	XOM	-	XPR.MI

Source: Yahoo! Finance (available at: <<https://yhoo.it/2TNr5MM>>).

Obs.: Ticker 1 – stocks traded in US markets (including ADRs). Ticker 2 – stocks traded in the denounced company's country.

TABLE 11**Yahoo finance tickers for market indexes**

Country	Index ticker
Argentina	MERV
Australia	AORD
Austria	ATX
Belgium	BFX
Brazil	BVSP
Canada	GSPTSE
Finland	OMXH25
France	FCHI
Germany	GDAXI

(Continues)

(Continued)

Country	Index ticker
Ireland	ISEQ
Israel	TA125.TA
Italy	FTSEMIB.MI
Japan	N225
South Korea	KS11
Mexico	MXX
Netherlands	AEX
Norway	OSEAX
Portugal	BVLG
Spain	IBEX
Sweden	OMX
Switzerland	SSMI
United Kingdom	FTSE
United States	GSPC
Taiwan	TWII

Source: Yahoo! Finance (available at: <<https://yhoo.it/3InuVHP>>).

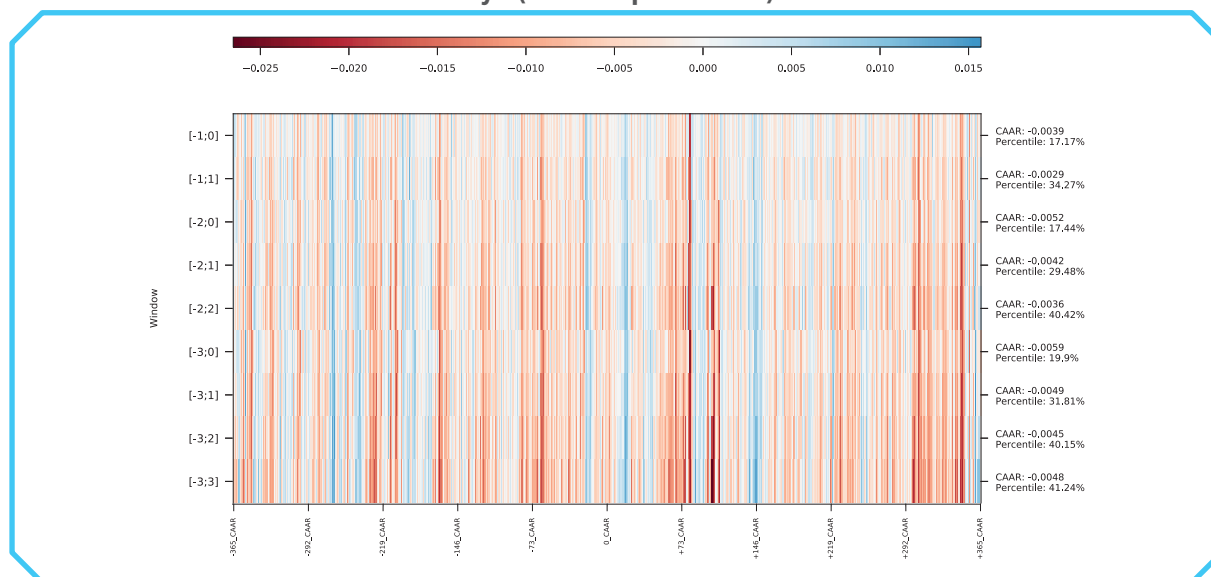
5.2 Placebo date test results

To test the robustness of the main results in the previous subsection, we now describe the results of the placebo date tests detailed in section 3. The idea is to use a different counterfactual to the firms' actual abnormal returns on the date that the misconduct allegations were registered by the NCP in the OECD Guidelines program. In the previous tests the counterfactual to a firm's returns were, first the average market return (market model), and then a matched firm's returns (PSM). Here we use as counterfactual the firm's own returns, but on placebo dates when the event did not occur. We use two sets of placebo dates. The first are the 365 days before and after the actual event, yielding a total of 730 counterfactuals. The second are 500 random dates drawn (with replacement) from the window of one year before and after the actual event. Each of these approaches yields a distribution of placebo CAARs against which we can compare the actual CAAR. These distributions have areas where the events have negative, positive and (statistically) zero CAARs, which allows us to infer the impact of the actual event.

DISCUSSION PAPER

Graph 9 shows the results for the placebo date estimation using as counterfactuals the 365 days before and after the actual event date. The date of each test, counted in relation to the actual date, are marked on the horizontal axis. Nine different event windows are displayed on the vertical axis. Estimated CAARs are colored in accordance with the scale shown at the top of the graph, with negative values darker shades of red and positive values darker shades of blue, as they increase in absolute value. It is easily visible that there are negative and positive events in the examined period. Because this analysis uses a sample of different firms each with a different event date, these placebo events are due to chance. After all, each placebo date falls on a different calendar date, so they cannot be capturing some common market shock. The question which we want to ask is if the event in ground zero (0_CAAR), which is the normalized actual event date for all firms, is statistically different from zero. We do this by determining what percentage of the placebo dates have estimated CAARs below the CAAR at the actual date. The actual CAAR and these percentages are shown for each window in the right hand side of the graph. Graph 10 repeats this procedure but sets all statistically non-significant estimated CAARS (using a Grankt test) to zero.

The CAARs for all windows in graph 9 are negative, but the lowest cumulative percentage, in window $[-1;0]$, which can be thought of as a p-value, is 17.17%. But in graph 10, where the values that are not statistically significant at 5% are set at zero, the estimated CAARs for three of the windows are below 10% of the placebo CAARs (8.82% for $[-1;0]$, 9.23% for $[-2;0]$ and 9.51% for $[-3;0]$). The interpretation for these results is that negative CAARs at or below the estimated actual-date value would only happen by chance in these percentage of cases, so we can accept that the Guidelines, on average, had a negative impact on returns.

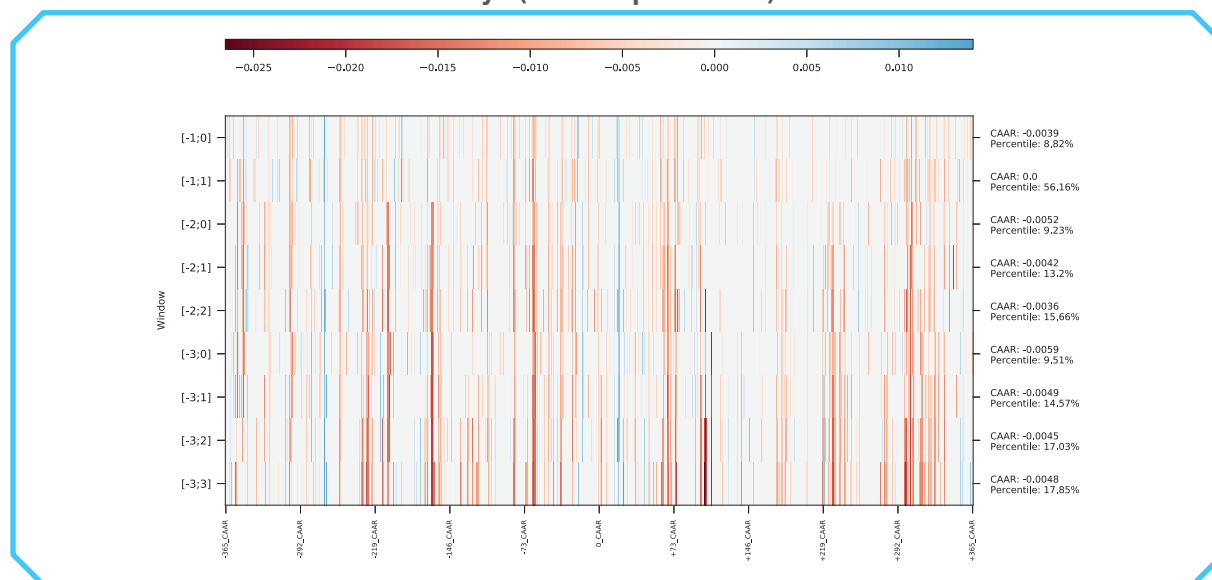
GRAPH 9**Placebo event dates \pm 365 days (Ticker1 post-2011)**

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

DISCUSSION PAPER

GRAPH 10**Placebo event dates \pm 365 days (Ticker1 post-2011)**

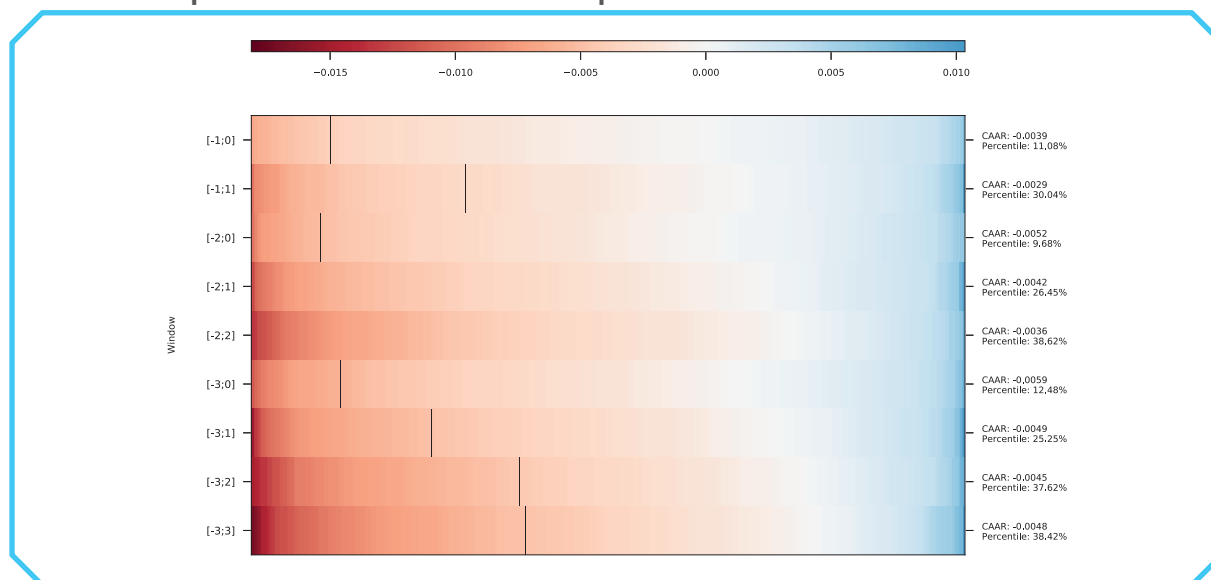
Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Obs.: Non-significant CAARs based on the Grankt test set to 0.

Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

In graph 11 and 12 the counterfactual dates are chosen randomly in the window of dates one year before and one year after the actual event date (500 draws with replacement). In both graphs the estimated CAARs have been ordered from smallest to largest and the actual CAAR is indicated with a black vertical line. As before, the difference between the graphs is that the second sets the non-statistically significant values to zero. The statistics on the right-hand side of the graphs indicate that the estimated CAARs are below the 10% mark in window [-2;0] in graph 11, and in windows [-1;0], [-2;0] and [-3;0] in graph 12. These results indicate the robustness of our previous finding that, on average, the OECD Guidelines produced negative abnormal returns on the affected companies.

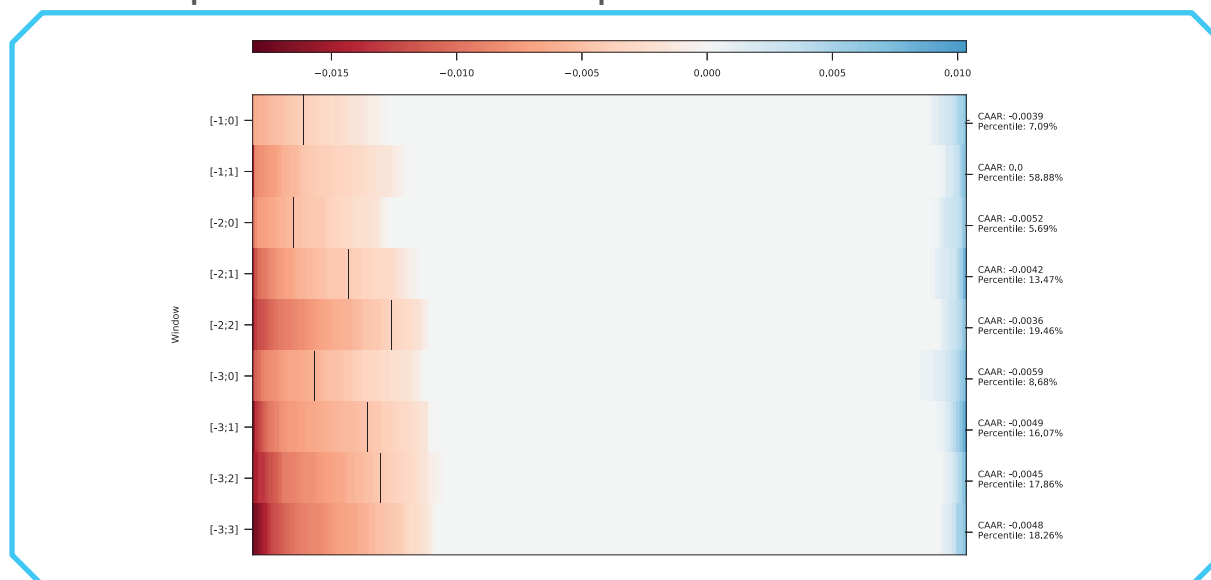
GRAPH 11**Random placebo event dates – Ticker1 post-2011**

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

DISCUSSION PAPER

GRAPH 12**Random placebo event dates – Ticker1 post-2011**

Sources: Stock market data from Yahoo Finance (available at: <<https://yhoo.it/3fq66Y9>>); and data for specific instances from the OECD Guidelines for Multinational Enterprises (available at: <<https://bit.ly/3xuAXsR>>).

Authors' elaboration.

Obs.: Non-significant CAARs based on the Grankt test set to 0.

Publisher's note: Graph whose layout and texts could not be formatted and proofread due to the technical characteristics of the original files.

6 DISCUSSION AND CONCLUDING REMARKS

For the OECD Guidelines for Multinational Enterprises to be effective it is necessary that the program be able to offer a credible threat. Companies that are denounced must feel that they have something to lose if they adopt behavior that exposes them to the program and if they do not then opt to comply with what the program classifies as responsible business conduct. The NCPs do not have police, regulatory, or any other formal legal power, so if they are to convince the companies to incur the costs of compliance, they must have some threat to persuade the company to engage in the mediation and arbitration procedure. This power must not only be unequivocal, but it must also be large, otherwise it can be shrugged off as a minor nuisance. It is thus important that the Guidelines have the OECD stamp. The OECD is one of the largest and most respectable international organizations in the world, so the program is, in this regard, on firm ground. But it must also be the case that the procedures wielded within this program actually work to impose embarrassment and reputational (and other) costs on firms

that opt not to participate. Financial markets have experienced remarkable growth in interest in ESG investment in the past 10 years (Amel-Zadeh and Serafeim, 2018). These trends should increase the willingness of companies to abide by the OECD Guidelines and to participate in the mediation process in cases where they are denounced.

This is why the results in this report are important. They show, quantitatively, that the premise on which the whole program depends, are actually sound. That is, being denounced to the program can affect shareholder value, so that firms that are called out would be better off engaging with the arbitration and mediation offered by the Guidelines. Even more importantly, other firms whose behavior violates responsible business conduct will realize that it might be in their interests to preemptively comply.

Our main result is to quantify that on average the impact for a company of being denounced is negative. The estimated cumulative abnormal return may look like a small and temporary effect, in the 0.001 to 0.007% range in table 2, for example. But these are percentages over the value of very large companies, and they get at what is most dear to shareholders, which is the returns to their shares.

Our results are also more nuanced than simply measuring the average impact of the program. We found that the program works better in the USA and Canada than in other continents, and in some sectors than in others. This information can be used to better identify what works and what does not. The program relies not only on how the NCPs differently approach each case, but on how well they are able to engage and achieve the trust of both the aggrieved social groups and organizations, as well as of the firms. While our investigation does not uncover which approaches are more effective, they serve as an important indication of where to look.

REFERENCES

AMEL-ZADEH, A.; SERAFEIM, G. Why and how investors use ESG information: evidence from a global survey. **Financial Analysts Journal**, v. 74, n. 3, p. 87-103, 2018.

DISCUSSION PAPER

BOEHMER, E.; MASUMECI, J.; POULSEN, A. B. Event-study methodology under conditions of event-induced variance. **Journal of Financial Economics**, v. 30, n. 2, p. 253-272, 1991. Retrieved from: <<https://bit.ly/3wwFeLD>>.

DAVARNEJAD, L. In the shadow of soft law: The handling of corporate social responsibility disputes under the OECD guidelines for multinational enterprises. **Journal of Dispute Resolution**, n. 2, p. 351-385, 2011.

FAMA, E. F. Efficient capital markets: a review of theory and empirical work. **The Journal of Finance**, v. 25, n. 2, p. 383-417, 1970.

GELBACH, J. B.; HELLAND, E.; KLICK, J. Valid inference in single-firm, single-event studies. **American Law and Economics Review**, v. 15, n. 2, p. 495-541, 2013.

HECKMAN, J. J.; ICHIMURA, H.; TODD, P. E. Matching as an econometric evaluation estimator: evidence from evaluating a job-training programme. **The Review of Economic Studies**, v. 64, n. 4, p. 605-654, 1997.

KOLARI, J. W.; PYNNONEN, S. Nonparametric rank tests for event studies. **Journal of Empirical Finance**, v. 18, n. 5, p. 953-971, 2011. Retrieved from: <<https://bit.ly/3wDIXIq>>.

KOTHARI, S. P.; WARNER, J. B. Econometrics of event studies. In: ECKBO, B. E. (Ed.). **Handbook of corporate finance: empirical corporate finance**. Amsterdam; Oxford: North-Holland; Elsevier, 2007. v. 1. p. 3-36.

NIEUWENKAMP, R. OECD guidelines for multinational enterprises on responsible business conduct. **The Dovenschmidt Quarterly**, n. 4, p. 171-175, 2013.

OECD – ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Implementing the OECD guidelines for multinational enterprises: the national contact points from 2000 to 2015**. Paris: OECD, 2016.

PWC – PRICEWATERHOUSECOOPERS. **Strategies for responsible business conduct**. [n.l.]: PwC, 2018. (Reference: 2018-09111/AK/mp/vd).

REESE JUNIOR, W. A.; ROBINS, R. P. Performing an event study: an exercise for finance students. **The Journal of Economic Education**, v. 48, n. 3, p. 206-215, 2017.

ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, v. 70, n. 1, p. 41-55, 1983.

RUGGIE, J. G.; NELSON, T. Human rights and the OECD guidelines for multinational enterprises: normative innovations and implementations challenges. **The Brown Journal World Affairs**, v. 22, n. 1, p. 99-127, 2015.

THORSTENSEN, V. H.; CORTELLINI, A. C. N.; GULLO, M. F. **A OCDE como fórum de governança das empresas multinacionais**. São Paulo: EESP/FGV, 2018. (Working Paper, n. 488).

WEISSMAN, E. et al. **Quantifying the costs, benefits and risks of due diligence for responsible business conduct**: framework and assessment tool for companies. [n.l.]: OECD, 2016. (OECD Report).

Ipea – Institute for Applied Economic Research

PUBLISHING DEPARTMENT

Head of the Publishing Department

Reginaldo da Silva Domingos

Assistants to the Head of the Department

Rafael Augusto Ferreira Cardoso

Samuel Elias de Souza

Supervision

Camilla de Miranda Mariath Gomes

Everson da Silva Moura

Typesetting

Aeromilson Trajano de Mesquita

Anderson Silva Reis

Cristiano Ferreira de Araújo

Danielle de Oliveira Ayres

Danilo Leite de Macedo Tavares

Jeovah Herculano Szervinsk Junior

Leonardo Hideki Higa

Cover design

Aline Cristine Torres da Silva Martins

Graphic design

Aline Cristine Torres da Silva Martins

*The manuscripts in languages other than Portuguese
published herein have not been proofread.*

Ipea Bookstore

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo

70076-900 – Brasília – DF – Brazil

Tel.: + 55 (61) 2026 5336

Email: livraria@ipea.gov.br

Ipea's mission

Enhance public policies that are essential to Brazilian development by producing and disseminating knowledge and by advising the state in its strategic decisions.



ipea

Institute for Applied
Economic Research

MINISTRY OF
ECONOMY



**PÁTRIA AMADA
BRASIL**
BRAZILIAN GOVERNMENT