Università degli Studi di Parma

Dottorato di ricerca in

"Economia e Management dell'Innovazione e della Sostenibilità"

Ciclo XXIV

in CO-TUTELA con

Università degli Studi di Ferrara

Essays in Policy Evaluation

Coordinatore: Chiar.mo Prof. Stefano Azzali

Tutore: Chiar.mo Prof. Leonzio Giuseppe Rizzo

Dottoranda: Luisa Loiacono

Anni Accademici 2018/2019 - 2020/2021

Contents

Introduction

Motivation and background

Over the last two years, the spread of COVID-19 led countries around the world to adopt stringency measures with the purpose of limiting the spread of the virus and relieve the burden on the healthcare system. The containment measures deeply changed people's behaviour and mobility, which, in turn, resulted in heterogeneous impacts on different economic sectors.

As people respond differently to these containment measures, in the first part of this dissertation (Chapter 1) we analyse the determinants of the heterogeneous responses of citizens around the world. We employ a staggered difference-in-differences (diff-in-diff) design, where we analyse how stringency measures affect citizens' mobility across different countries. As we want to measure the role of people's knowledge about the spread of COVID-19, we disentangle this effect from the role of the policy measure itself. In our analysis, we find that stringency measures have a higher impact on mobility the more people learn about COVID-19 and that this finding is driven by countries with low levels of trust in institutions and low levels of education.

Because of the nature of the containment measures related to the COVID-19 pandemic, their impact is dramatically heterogeneous across industries and sectors. For example, the sector of the contact intensive services[1] is suffering severe losses: in the second quarter of 2020 it reached a 25% downward deviation compared to its pre-pandemic trends (European Commission, 2021). Quite the opposite, other sectors such as IT sector, delivery services, e-commerce and healthcare [2], are taking advantage of the pandemic. Because of this heterogeneous response, the introduction of temporary taxes is being discussed by both international and national policy makers. These taxes are aimed at funding the recovery and sustaining those who are facing an economic disadvantage. At first, the debate only focused on those sectors that are benefiting from the ongoing pandemic. However, lately, the introduction of a temporary extra tax is also setting in because of the recent sharp increase in gas and electricity prices in Europe. In the second part of this dissertation (Chapter 2) we analyse the possible drawbacks of a non-permanent tax by estimating the impact of a

[1] Example contact intensive services are trade, transport and accommodation, arts and entertainment.

[2] See 'Prospering in the pandemic: winners and losers of the Covid era', Financial Times. https://www.ft.com/content/8075a9c5-3c43-48a5-b507-5b8f5904f443

temporary tax adopted in Italy in 2008, the so-called "Robin Hood" tax. This tax is a surcharge applying to Italian firms operating in the energy sector whose revenues are above a discrete threshold (25 million euros). Since the threat of negative trickle-down responses often make policy makers reluctant to adopt extra taxes, we study how big corporations responded to tax hikes raised in extraordinary times. In particular, we investigate whether firms shifted the surtax onto labor or onto capital. By using both a diff-in-diff and a regression discontinuity, we provide compelling evidence that, for big firms, the tax did not hurt investments and that the tax burden was not shifted onto workers.

As already mentioned, the energy sector is currently benefiting from high prices of gas and electricity, as in 2008, when the "Robin Hood" tax was firstly introduced. The various energy prices are strongly connected, therefore electricity, gas, carbon and carbon emission prices usually follow parallel paths. Nowadays, the gas wholesale spot market price is leading the increase in energy prices and the determinants are multifactorial. The driving issue is that demand started increasing because most pandemic related containment measures ended while supply is shrinking because exporting countries, upon which Europe strongly relies on, are not increasing the volumes accordingly. Moreover, the carbon emissions price from the European Union Emission Trading System almost doubled with the respect to the beginning of 2021 because of political choices recently adopted aimed at increasing the market price to encourage companies' emissions cut. In addition, the EU climate goals are phasing out conventional power plants but the technology and the market configuration are not ready yet to fully rely on renewable energy sources (RES). In fact, in absence of the proper market configuration, grid inter-connectors and storage capacity, the intermittent nature of RES raises big challenges for the energy market structure. According to the European Network of Transmission System Operators for Electricity (ENTSO-E, 2021) the Transmission System Operators[3] need to deal with the increasing share of RES by removing the gap between market outcomes and the actual grid capacities, designing optimal bidding zones, increasing transparency and developing congestion management markets. In the third part of this dissertation (Chapter 3), we analyse one of the determinants of the energy market prices mentioned above: the bidding zone configuration. We analyse the market re-configuration that took place in

[3] A Transmission System Operator (TSO) is an entity independent from the electricity market that transmits electricity at high voltage levels.

Sweden in 2011 and examine whether the prices and the cross-zonal and cross-country flows changed due to the policy intervention. We use a diff-in-diff analysis to compare the average price in Sweden with a theoretical market price, the system price, that works as a control variable. The system price is a proper control becasue, by definition, it is not affected by the market configuration as for its calculation capacities are set to infinity. With a regression discontinuity in time and an instrumented variable approach, we analyse how the policy intervention impacted on cross-zonal and cross-country flows. We find an increase in Swedish prices, a decrease in net flows from the North to the South of the country and an increase in exports towards neighbouring countries (namely Finland) suggesting that the re-configuration was effective in adjusting market mechanisms.

Policy evaluation

In the three chapters of this dissertation, we deal with policy evaluation problems where we identify the effect of some public interventions on outcomes of interest and, by using appropriate assumptions and estimation strategies we are able to draw causal inferences. In all our cases, the policy intervention is defined as the treatment and it assigns the treatment status to the treated units. Given that Y is the outcome variable, each observation is associated with the treatment (T) equal to one when the unit i is treated and zero otherwise, we get:

$$\begin{aligned} Y_i &= (Y_i^1 - Y_i^0)T_i + Y_i^0 \\ &= \alpha + \beta_i T_i + u_i \end{aligned} \tag{1}$$

where the impact of the treatment is associated with β_i and is equal to $Y_i^1 - Y_i^0$. The model in Equation 1 is known as the Rubin Causal Model (Imbens & Rubin, 2010). The Rubin Causal Model fundamental problem[4] is that one of the possible potential outcomes is always an omitted observation, i.e., we observe any unit only under the assigned treatment, and we can not observe the same unit under the other treatment. As we can only observe units in one of the two possible treatment status in a given moment, we use the counterfactual approach to infer causality. Inference models to estimate causality include randomized trials, matching, instrumental variables,

[4] This is known as the 'Fundamental Problem of Causal Inference' defined by Holland (1986): it is impossible to observe the value of Y_i^1 and Y_i^0 on the same unit i, therefore it is impossible to observe the effect of the treatment on the same unit i.

regression discontinuity designs and diff-in-diff (Angrist & Pischke, 2014). In this dissertation, we use parametric models such as diff-in-diff, regression discontinuity and instrumental variables to estimate the casual effects of the policy interventions we analyse.

In Chapter 1 we use a staggered diff-in-diff approach meaning that the treated units do not adopt the treatment at the same point in time. Our case relies on the adoption of containment measures (also called 'stringency index' throughout the dissertation) across countries in the world during the year 2021. For each country our treatment variable is equal to one after the stringency index is implemented and zero otherwise. In the model we also interact the staggered treatment variable with the treatment intensity as the stringency index is a continuous variable. In the case of the diff-in-diff, for the coefficient of the treatment to infer causality the parallel trends assumption must hold. This assumption tests whether during the pre-treatment period the outcome variable in the two groups, treated and untreated, follows parallel trends. The intuition behind this test is that, in absence of treatment, the two groups would have behaved the same also after the treatment intervention. In our analysis we test whether the outcome variable (mobility) follows a parallel trend in the pre-intervention. We verify via the Autor test (Autor, 2003) that the parallel trend assumption is satisfied therefore we are able to make inference from our model and measure to what extent the stringency index has an impact on the mobility.

In Chapter 2 we measure the firms' reaction to the introduction of a surtax by looking at multiple outcome variables and we estimate the effect of the policy intervention by using both a diff-in-diff and a regression discontinuity. The treatment status depends on the surtax eligibility that relies on the turnover being above or below a discrete threshold. We measure to what extent the introduction of the "Robin Hood" tax has an impact on pre-determined balance-sheet outcome variables. Firstly, we use a diff-in-diff approach where the control group consists of firms below the threshold. Secondly, we double check our results by using a cross-sectional regression discontinuity, a quasi-experimental method that assumes that treatment assignment is based on a deterministic function of a continuous variable, also called the 'running variable'. The units whose running variable is above the given threshold are the treated group, while the ones below it are the control group. The main assumption of the RD is that the treated and the control groups are very similar in terms of unobserved characteristics and therefore the estimated difference is given by the treatment intervention.

Finally, in Chapter 3, we estimate the casual effect of the electricity market reconfiguration in Sweden by estimating its impact on both prices and cross-zonal and cross-country flows. We apply both a diff-in-diff and a regression discontinuity in time (RDiT). We divide the analysis into two parts: in the first one we estimate the Swedish price with a diff-in-diff approach using a theoretical price as control; in the second one, we apply a RDiT to measure the impact of the policy intervention on cross-country and cross-zonal flows. We are able to estimate the price effect with a diff-in-diff as we can use the system price as control, however, given the nature of the electricity markets, neighbouring countries can not work as counterfactuals for the cross-zonal and cross-country flows. The RDiT we apply is different from a typical regression discontinuity because treatment begins at a particular threshold in time and it can not exploit cross-sectional variation. The two main implications of this design are: firstly, given the lack of cross-sectional variations, we must include observations far from the threshold; secondly, we must consider that we are dealing with a time-series.

References

(1) Angrist, J. D., Pischke, J.S., (2014). "Mastering 'Metrics: The Path from Cause to Effect", *Economics Books*, Princeton University Press, edition 1, number 10363.

(2) Autor, D.H. (2003). "Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing", *Journal of Labor Economics*, 21, 1–42, doi: 10.1086/344122.

(3) European Commission. (2021). "The sectoral impact of the COVID-19 crisis", Technical note for the Eurogroup.

(4) European Network of Transmission System Operators for Electricity. (2021). "Options for the design of European Electricity Markets in 2030".

(5) Hausman, C., Rapson, D.S. (2017). "Regression Discontinuity in Time: Considerations for Empirical Applications", *NBER Working Papers*, 23602, National Bureau of Economic Research, Inc.

(6) Holland, P. (1986). "Statistics and Causal Inference", (with discussion), *Journal of the American Statistical Association*, 81, 945-970.

(7) Imbens, G. W., Rubin, D. B. (2010). "Rubin Causal Model", *Microeconometrics* edited by Durlauf, S. N., Blume, L. E., Palgrave Macmillan UK, 229-241.

Chapter one

1 Pandemic knowledge and regulation effectiveness: evidence from the COVID-19

1.1 Introduction

This chapter is set among the growing literature on the economic effects of the COVID-19 pandemic. The World Health Organization announced the international outbreak of the COVID-19 infection on January 30, 2020, and declared it a global pandemic on March 11, 2020 (Cucinotta and Canelli, 2020). By the first half of 2020 the pandemic had reached almost all countries around the world and while during the summer the first wave seemed to had ended, during the fall of the same year a second wave started spreading. As we speak, on January 2022, more than 349 million of COVID-19 infected cases have been reported, with nearly 5.9 million deaths[5]. The COVID-19 pandemic has had devastating effects not only on citizens' health and on health care systems, but it has also had a strong economic impact on countries around the world. However, countries responses are heterogeneous and depend on multiple factors such as their exposure to the virus, the health care and economic systems.

National governments have been adopting containment measures which restrict the movement of individuals and impose various forms of social distancing. These measures significantly differ in both their timing and their intensity and while some countries adopt strong containment measures at early stages of the COVID-19 outbreak, others adopt less strict measures at later stages (Ferraresi et al., 2020).

These measures are put in place to reduce mobility and enforce social distancing to limit the spread of the virus. The ultimate goal of the stringency measures is to lighten the burden on the healthcare systems (especially on intensive care units) with the aim of decreasing the mortality rates. Nonetheless, the achievement of this purpose depends on factors that go beyond the design and timing of the containment measures themselves. In fact, as we investigate throughout this chapter, citizens' responses play a pivotal role in determining the effectiveness of the adopted measures.

The adoption of the containment measures has been the subject of strong disagree-

[5]Daily coronavirus disease (COVID-19) reports are available on the World Health Organization's webpage (https://covid19.who.int/). The actual number of infected cases is likely to be significantly higher as asymptomatic carriers of the infection are not detected.

ment amongst media, politicians, scholars and the public opinion[6]. In fact, several protests against the containment measures have taken place in multiple countries in Europe[7], in the US[8] and elsewhere. These protests were driven by the dramatic economic drawbacks of the government responses to the COVID-19 but also by their legality since they strongly limited individual freedoms.

The analysis presented here is based on the intuition that individual responses are related to citizens' perceived risks about the likelihood of contagion and of severe health outcomes. Those risks are influenced by the information that citizens have about the pandemic, which they obtain by multiple sources such as individual interactions, media and social networks. The recent literature has covered the topic of the role of media in influencing individual perceptions. For example, Mastrorocco and Minale (2018) find an effect of news media on crime perceptions. They use a diff-in-diff design that compares individual perceptions of those with different ranges of available TV channels. Moreover, scholars have investigated how perception and knowledge affect individual behavior in the case of political elections (Martinand and Yurukoglu, 2017) and crime (Shi, 2009; Velásquez et al., 2020; Spenkuch, 2018).

In this chapter we estimate the effects of stringency measures on citizens' mobility with respect to their knowledge about the pandemic. We proxy this knowledge with the relative amount of Google searches about COVID-19, a daily and country-specific measure of citizens' search for information about the pandemic.

With respect to the COVID-19 pandemic, recent literature has begun to investigate the determinants of the effectiveness of stringency measures, identifying variables such as expectations for the duration of self-isolation, trust in science (Briscese et al., 2020), political affiliation (Allcott et al., 2020; Painter and Qiu, 2020), social responsibility and social trust (Oosterhoff and Palmer, 2020), and the trust in policymakers' ability to handle the crisis (Bargain and Aminjonov, 2020; Brodeur et al., 2020; Farzanegan and Hofmann, 2021). But, to the best of our knowledge, there is no empirical analysis of the relationship between stringency measures and mobility which explicitly incorporates the knowledge about COVID-19.

[6] In the UK, for example, the restrictions that underpin the COVID-19 stringency measures have been recently challenged as being unlawful and disproportionate, breaching freedoms protected by the European Convention of Human Rights (Keene, 2020). In New Zealand the government's decision to impose a month-long stringency to stem the spread of coronavirus has also been challenged in court.

[7] See 'German police cracks down on anti-lockdown protesters', FT, May 17, 2020 (J. Miller).

[8] See 'US anti-lockdown protests: 'If you are paranoid about getting sick, just don't go out'', FT, April 22, 2020 (D. Crow).

In the following sections, we implement a diff-in-diff research design by focusing on the consequences of the stringency measures on the mobility level of the population. In particular, we use daily observations from February 15, 2020 to December 25, 2020 (315 days), across 35 countries[9] for which these data are available[10].

We exploit the staggered implementation of stringency measures adopted by countries along time, while controlling for country and daily fixed effects. Our results indicate that stricter stringency measures are significantly associated with lower mobility, and that this effect is larger the more people get information about the spread of COVID-19. These results survive a set of robustness tests, including the traditional event-study test à la Autor (2003).

1.2 Empirical strategy

Our baseline empirical model builds on the large and expanding literature that makes use of the staggered diff-in-diff method (as described in the Introduction) to investigate the effect of a policy on pre-determined outcomes. As noted above, while countries eventually adopted stringency measures in the year 2020 due to the COVID-19 outbreak, they differ in the timing of the adoption. This allows us to compare the change in the mobility in the treatment group before and after the adoption of the containment measures.

The estimated diff-in-diff model is the following:

$$mobility_{cd} = \alpha + \gamma stringency_{cd} + \beta X_{cd} + f_c + f_d + u_{cd} \tag{2}$$

where $mobility_{cd}$ is the Google mobility for country c in day d;$stringency_{cd}$ is the Stringency Index in country c and day d, ranging from 0—when stringency measures have not been adopted yet—to 100, with 100 denoting the maximum level of stringency; X_{cd} are daily variables at country level, such as temperatures, weekly moving average of the pandemic confirmed cases per capita and the intensity of searches on Google of the term "Covid" for each country a week before; f_c are country fixed effects that control for unobserved cross-country heterogeneity[11]; f_d are daily fixed effects

[9] The list of the 35 countries included is reported in Table A1 in Appendix A.

[10] The sample includes only the 35 countries which after February 15, 2020 (first day of available data on mobility), experienced at least one day without any COVID-related restrictions. This allows us to test for the parallel trend assumption via the Autor test. We explore the robustness of our baseline results by replicating them on the full set of available countries (109 countries), see Table A3 in Appendix A.

[11] In turn, this heterogeneity might be due to different levels of technology that affect both mobility

that capture time-specific shocks common to every country, such as Covid-related information that becomes available worldwide in a given day; and u_{cd} is the error term, clustered at country level. In some specifications, we also control for country specific trends. Within this specification, λ is the diff-in-diff estimate of the (average) effect of the stringency on mobility.

To investigate whether there has been a heterogeneous response to containment measures as a function of the knowledge about COVID-19 on a given day in each country, we interact Covid searches with the stringency measures. Covid searches ranges in each country from 0 - when there is no search in Google of the term "Covid" - to 100, with 100 denoting the maximum level of Covid searches.

The estimated model is a generalised version of Equation (1), taking the following form:

$$\begin{aligned} mobility_{cd} = \alpha + \gamma stringency_{cd} + \lambda covid\ searches_{cw} \\ + \vartheta stringency_{cd} \times covid\ searches_{cw} + \beta X_{cd} + f_c + f_d + u_{cd} \end{aligned} \tag{3}$$

where our variable of interest is the coefficient ϑ.

1.3 Data

In this section we describe the main variables of our analysis, namely movement of individuals, containment measures measured with the stringency index and the COVID-19 knowledge which is proxied with the google searches for the term "COVID". The summary statistics for all the variables used in the analysis are reported in Table A2 in Appendix A.

1.3.1 Movement of individuals

To measure the daily movement of people during the spread of COVID-19, we use the COVID-19 Community Mobility Reports provided by Google[12]. The mobility indicators measure the relative value of mobility for each day of the week which is compared to a baseline value for the same of the week. The baseline value is calculated as the median value recorded during the 5-week period from January 3

and Google hits, national differences in the contagion level, health-care systems (such as availability of testing and Intensive Care Unit capacity), as well as population density and the age profile of the population.

[12]For details see: https://www.google.com/covid19/mobility/

to February 6, 2020, i.e., before the start of the pandemic. So, the indicator takes on a value of 100 if mobility in given day during the pandemic, say on a Monday, is equal to the Monday pre-pandemic median. The Community Mobility Reports provide six different place categories: grocery & pharmacy, parks, transit stations, retail & recreation, residential and workplaces. In the main regression, we use as dependent variable the daily average of the above categories from which we exclude the 'residential' category as it has different units of measurement (i.e., change in duration vs change in total visitors)[13]. Following Helsingen et al. (2020), we use observed data on mobility because they are more reliable than individual surveys due to the potential confounding role of individual biases in the way respondents self-report their behavior.

1.3.2 Stringency Index

In order to deal with the COVID-19 outbreak, governments around the world adopted many and very different containment measures. We take into account the heterogeneity of governments' responses by making use of the Government Response Stringency Index (Stringency Index) developed by Hale et al. (2020). The Stringency Index is calculated using the mean of nine metrics: school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements, and international travel controls. Each of these variables is rescaled by its maximum value to create an overall score between 0 and 100. A higher score indicates a stricter response (i.e., 100 is equal to the strictest response). The index just indicates the strictness of government policies, and it does not give a measure of the effectiveness of a country's response.

1.3.3 COVID-19 knowledge

We use data from the Google Trends tool to measure the citizens' active pursuit of information related to the pandemic. As in previous work using Google Trends to predict disease outbreaks (Carneiro and Mylonakis, 2009), trading behavior in financial markets (Preis et al., 2013), and concern of public opinion about pension

[13]As robustness checks, we use as dependent variable the mobility indicator excluding one of each component at time (Table 4). We also use as dependent variable the mobility with its individual components (Table A6 in Appendix A).

systems (Fornero, Oggero and Puglisi, 2019), we assume that Google search indicators provide reliable information about citizens' (search for) knowledge. The tool provides an index for online search intensity of a specific term (and its components) over the time period under consideration within a specific area. The index is a weekly measure of intensity, which is computed as the number of weekly searches for the term divided by the maximum number of its weekly searches over the whole time period, in a given country. The result is scaled from 0 to 100, where 100 is the peak popularity and 0 means that there was not enough search volume for that specific term during that week. For our purposes, we collect the searches related to the term “Covid” for the period covering the 315 days of our analysis, from February 2020 to December 2020. In order to conduct a falsification test, we also collect the searches related to the terms that were most searched worldwide on Google from February 2020 to December 2020, i.e. “translate”, “porn”, and “maps”.

However, the active pursuit of information related to the "COVID" term is not independent from the media coverage. For example, a paper from Sousa-Pinto et al. (2020) shows that the Google Trends for COVID-19 symptoms such as cough, anosmia and ageusia are more strongly related to media coverage than to the underlying pandemic pace trends[14]. The authors find that the peaks for the Google searches on the various symptoms occurred simultaneously, irrespective of the country's pandemic stage.

1.4 Results

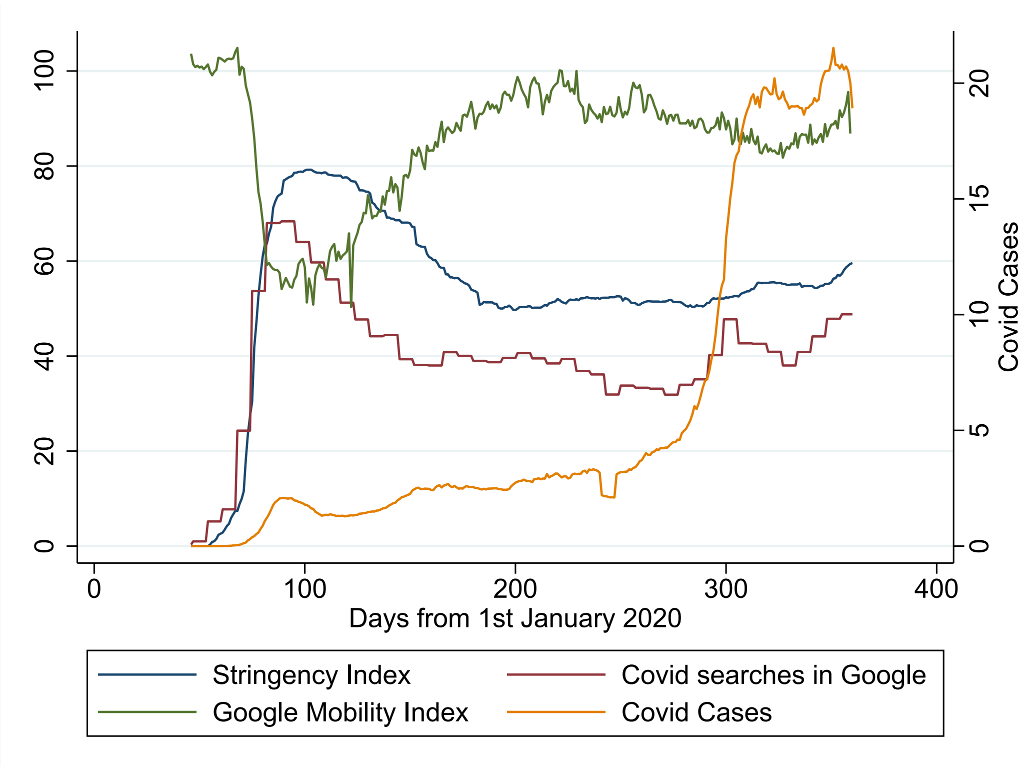
Figure 1 plots the relationship between our measure of people's movement, throughout the chapter called "Mobility", the level of the containment measures, called "Stringency Index", an indicator of the spread of the COVID-19, called "COVID-19 cases per capita", and a measure of the knowledge of the pandemic, called "Covid searches in Google". From the figure, we see that for the first 80 days of 2020 there is a clear inverse relationship between the stringency measures and COVID-19 online searches with population movement. After this initial period Covid searches decrease faster than stringency measures, while, at the same time, mobility starts increasing. This divergence between stringency measure and Covid searches and its relationship with mobility raises the issue of the role of citizens' knowledge about the seriousness of

[14] https://doi.org/10.1371/journal.pone.0152802; https://www.jmir.org/2020/8/e19611

COVID-19 pandemic.

Table 1 displays the findings from our regression analysis. The first three columns report the results based on the different specifications of Equation 2 for 35 "baseline" countries: these are the countries that – within our time frame- experienced an initial phase with no COVID-related restrictions, so that it is possible to test for the parallel trend assumption via the Autor test[15].

Figure 1: Mobility, stringency measures, Covid searches, and Covid cases per capita.



Note: This figure displays the Stringency Index, the Google Mobility, Covid searches on Google and weekly moving average of per capita Covid Cases over 2020, from February 15 (day 46) to December 30 (day 365). Observations for 35 countries are averaged by day. The Stringency Index and Covid searches vary from 0 to 100 inside each country. The Mobility indicator is equal to baseline value of 100 for each country, if on a given weekday it exactly equals its median value recorded during the first 5 weeks of 2020, i.e., before the start of the pandemic. Per capita Covid Cases correspond to new active cases and are calculated as the difference in per capita cumulated cases between day t and day t-1. For more details, see Section 1.3.

The baseline specification, which includes country and time fixed effects, and confirmed per capita cases[16] as control variable, is reported in Column 1. Column

[15]Table A3 in Appendix A extends the results shown in Table 1 to all countries available in our original dataset. The main results do not change significantly. From our main analysis, we drop the countries that, within our time frame, do not have a pre-treatment period with a stringency index equals to zero.

[16]In Table A4 in Appendix A, we replicate the regressions in Table 1 by replacing the 7-days moving average of per capita confirmed cases with the 14 days lag of per capita confirmed deaths. We do not observe any relevant difference in the coefficients of interest vis a vis the main specification.

2 adds to the previous specification the temperature variable, which would capture weather-related drivers of mobility[17]. In Column 3 we include as control variable the Covid searches. The last three columns show the results that are based on different versions of Equation 3. In Column 4 we include country and daily fixed effects, the temperature variable, and the interaction between Covid searches and the Stringency Index, while in Column 5 we include country-specific linear trends. In Column 6, in order to check whether the potentially heterogeneous reaction to the Stringency Index depends on real world events rather than on citizens' knowledge about those events, we add the interaction between confirmed per capita cases and the Stringency Index itself. Finally, in Column 7, we include a variable capturing the so called "Lockdown fatigue". Following Goldstein et al. (2021), starting from the first day of the Stringency Index greater or equal to 74.5 (which corresponds to the 75th percentile), we build the "Lockdown fatigue" variable by counting the number of days for which the Stringency Index was at least equal to 74.5. We include this variable in the regression in its linear and quadratic form.

In the first three specifications we find a negative and statistically significant relationship between mobility and stringency. The point estimates range from -0.489 to -0.398. This expected results suggests that, during the COVID-19 outbreak, the mobility in countries with stronger stringency measures decreases more than in those with weaker measures. The stringency variable is not a binary that switches on when the treatment starts. Conversely, it is a continuous variable that estimates the treatment intensity. In order to capture the impact of being treated at varying levels of the COVID searches, we compare the effect on mobility when the stringency and Covid searches are at extreme values of their joint distributions. For instance, following the point estimates of Column 3, the mobility is reduced by approximately 11.41 percentage points when considering a shift from Uruguay, whose level of both the stringency measure and Covid searches are the closest to the 25th percentile value, to Dominican Republic, whose level of both the stringency measure and Covid searches are the closest to the 75th percentile value[18]. In Column 4 the coefficient on the interaction term Stringency Index*Covid searches is negative and statistically significant at the 1% confidence level, with a point estimate of -0.003, while it is 5%

[17] Temperatures are retrieved from Global Historical Climate Network Daily (National Oceanic and Atmospheric Administration, 2020).

[18] This effect is computed as follows: $-11.42 = (-0.3979666 * (76.033 - 47.342))$, and it is statistically significant at the 1% level.

statistically significant in the last two specifications (Columns 5 and 6).

This result implies that the magnitude of the effect of the stringency measures on mobility is stronger for higher levels of COVID-19 knowledge, i.e., the effectiveness of stringency is amplified by the knowledge of the severity of the pandemic. Moreover, in our regressions we notice that the interaction of the stringency measure with the number of confirmed cases (Column 6) is not significant at ordinary confidence levels, while the interaction of stringency with Covid searches remains significant and with the same magnitude throughout all the specifications. This suggests that the role of Covid searches in determining the impact of stringency on mobility appears to be relevant and the real world events that are connected with the evolution of the pandemic by itself do not seem to matter. These coefficients confirm the results obtained by Sousa-Pinto et al. (2020) who find that the peaks for the Google searches on the different symptoms occurred simultaneously, independently of the country's pandemic stage.

Using the point estimates of Column 6, mobility is reduced by 35.90 percentage points[19] when the Stringency Index and the Covid searches are the closest to their 75th percentile values, i.e. 76.033 and 53.411 respectively; conversely, when the Stringency Index and the Covid searches are the closest to their 25th percentile value (47.342 and 29.069) the reduction in mobility is equal to 19.10 percentage points[20]. Therefore, the difference in mobility reduction is 16.80 percentage points, which is greater than what we obtained with the specification that does not include the interaction term between Covid searches and the stringency index (11.42 percentage points). Therefore, the Google search interaction term contributes to the mobility reduction by increasing it by 47% confirming that the impact of the stringency measures on mobility is not linear but it depends on the knowledge of the severity of the pandemic, proxied by the Google search for Covid.

In Column 7 the coefficient on the interaction term Stringency Index*Covid searches is negative and statistically significant at the 1% confidence level, with a point estimate of -0.002. We find a negative and significant coefficient on lockdown fatigue (-0.225, standard error 0.118) and a positive and significant coefficient on its squared term (0.001, standard error 0.0003): this suggests that for higher values of the vari-

[19] This effect is computed as follows: -35.90 = (-0.3215919*76.033-0.0028192 (76.033*53.411)), and it is statistically significant at the 1% level.

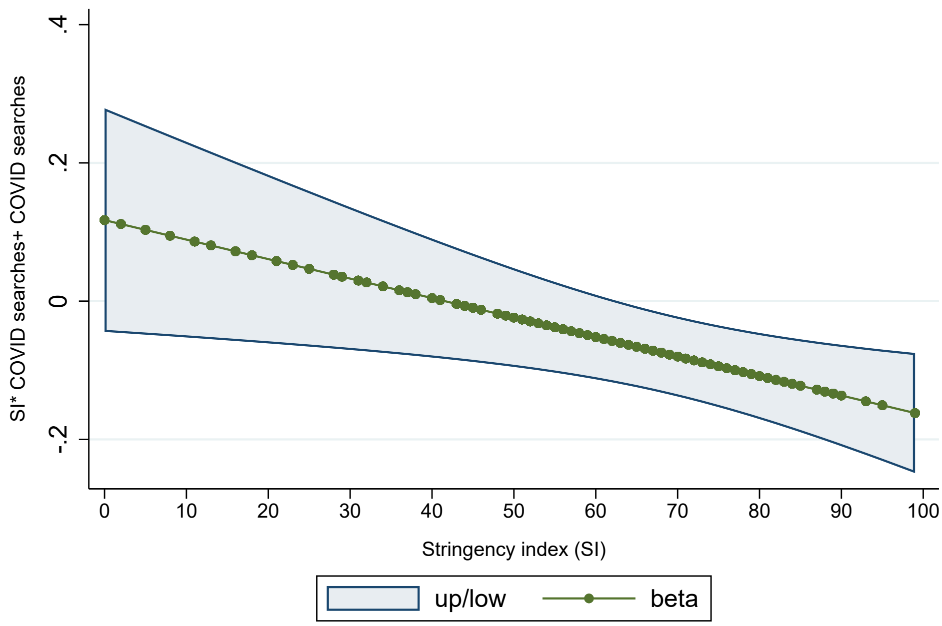
[20] This effect is computed as follows: -19.10 = (-0.3215919*47.342–0.0028192 (47.342*29.069)), and it is statistically significant at the 1% level.

Table 1: Diff-in-diff estimates, main specification.

DV: Mobility	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Stringency Index	-0.489***	-0.412***	-0.398***	-0.266***	-0.322***	-0.322***	-0.315***
	(0.059)	(0.055)	(0.058)	(0.064)	(0.071)	(0.071)	(0.065)
Conf. cases	-0.345***	-0.246***	-0.176**	-0.170**	-0.185*	-0.232	-0.218
per capita	(0.100)	(0.088)	(0.085)	(0.083)	(0.092)	(0.312)	(0.290)
Temperatures	-	0.118***	0.122***	0.119***	0.118***	0.118***	0.117***
		(0.017)	(0.018)	(0.017)	(0.020)	(0.020)	(0.019)
Covid searches	-	-	-0.079**	0.136	0.112	0.117	0.082
			(0.038)	(0.090)	(0.082)	(0.083)	(0.077)
Stringency Index*	-	-	-	-0.003***	-0.003**	-0.003**	-0.002*
Covid searches				(0.001)	(0.001)	(0.001)	(0.001)
Stringency Index*	-	-	-	-	-	0.001	0.001
Conf. cases pc						(0.004)	(0.004)
Fatigue	-	-	-	-	-	-	-0.225*
							(0.118)
Fatigue^2	-	-	-	-	-	-	0.001***
							(0.0003)
Observations	11,025	11,025	11,025	11,025	11,025	11,025	11,025
R-squared	0.755	0.790	0.793	0.797	0.8180	0.8181	0.826
Country FE	YES	YES	YES	YES	YES	YES	YES
Daily FE	YES	YES	YES	YES	YES	YES	YES
Country trend	NO	NO	NO	NO	YES	YES	YES

Note: This table shows the results of the diff-in-diff model described in equations 2 and 3. We regress country's mobility on different set of variables. In Column 1 we control for per capita Confirmed Cases (2); in Column 2 we additionally control for temperatures (2); in Column 3 we add Covid searches as a control (2); in Column 4 we also include the interaction between Stringency Index and Covid searches (3), while in Column 5 we add country-specific linear trends; in Column 6 we additionally include the interaction term between Stringency Index and per capita Confirmed Cases (3). Finally, in Column 7 we include both Lockdown fatigue and its squared term. Fatigue measures the lockdown fatigue is equal to the number of days in the past for which the Stringency Index was at least equal to 74.5, which corresponds to the 75th percentile. For all specifications we include country and daily fixed effects. In columns 5, 6 and 7 we also include country specific trend. The dataset is a country by day panel, for 35 countries and 315 days. Robust standard errors are clustered at country level (and shown in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 2: Plot of the linear combination of the interaction term and the Covid searches for different values of the Stringency Index.



Note: The figure shows the estimated coefficients on Covid searches for different values of the Stringency Index, together with their 95% confidence intervals. We plot the point estimates of the coefficients in correspondence of the mean values of each percentile of the Stringency Index.

able the impact on mobility turns out to be positive. More precisely, if we compute the first order condition with respect to lockdown fatigue, we find that the coefficient becomes positive after about 94 days of stringency above 74.5[21]. Finally, in Figure 2, we plot the values of the coefficient of interest, the interaction term between Stringency Index and Covid searches, for different values of the Stringency Index. We do this exercise to show how the Covid searches impact on the estimated coefficients for varying levels of the stringency index. The point estimates of the partial correlation of Covid searches with mobility are positive for low values of the Stringency Index, and negative for high values thereof. However, the confidence intervals are such that we cannot reject the null hypothesis that the partial correlation of searches with mobility is zero for low values of the Stringency index, while we reject the null hypothesis of a zero correlation for a large interval of high values of the Stringency Index.

In our main analysis, we show that Covid searches affect the compliance with the stringency measures. How to explain the fact that the interest in the pandemic affect

[21]-0.225*2*(0.0012)*lockdown fatigue

the compliance with the containment measures? Our intuition is that people comply with compulsory regulations when they perceive them as salient. In fact, when the pandemic becomes more relevant to them, people likely feel more pressure to comply with stringency measures themselves. In the next Section, we analyse how this effect changes for different levels of education and government indicators.

1.5 Heterogeneity analysis

The knowledge of the pandemic, that we proxy with the Covid searches, might matter more in economic and political environments with low levels of governance quality, whereas citizens do not necessarily trust the appropriateness of government interventions and/or news about those interventions. To explore this face of the issue, we consider the a country's level of education and its quality of institutions. We measure the quality of institutions with three different indicators: rule of law which measures citizens' subordinations to well-defined and established laws[22]; voice and accountability which is citizens' ability to participate in selecting their government[23]; and media repression, the state interference in communication and expression[24]. As for the education level we take as education indicator the average total years of schooling for adult population, over 25 years old[25]. We expect that when citizens have a low level of education they might be more prone to be influenced by Covid searches.

We expect that transparency of institutions (Voice and Accountability), citizens' confidence in the rules of society (Rule of law), low level of media repression and high level of education should narrow down the effect of citizens' knowledge, as measured by the volume of Covid searches: when people are more likely to trust institutions, abide by the law, receive fair information and are properly informed, Covid searches should not amplify or diminish the effects of stringency measures. Consistent with this argument, Table 2 shows that the amplification through the level of Covid searches of the negative impact of stringency on mobility is in fact driven by countries with lower level of government indicators and education, namely low values of Voice and

[22]As in Kaufman et al. (2010): "Rule of Law: capturing perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence."

[23]As in Kaufman et al. (2010): "Voice and Accountability: capturing perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.

[24]Freedom House. Data refers to the last available year (2014).

[25]This data is retrieved from Human Development Report (2018). Data refers to the last available year in the dataset, 2017.

Accountability and Rule of Law, high values of Media Repression and high values of Low Education.

The interacted coefficient for observations with Voice and Accountability above the median (Stringency Index*Covid searches*Above the median Voice and Accountability + Stringency Index*Covid searches) is almost zero and not significant, however, the not interacted coefficient due to observations with Voice and Accountability below the median (Stringency Index * Covid searches) is equal to -0.004 and statistically significant at the 1% level as well (Column 1 of Table 2). For Rule of Law, the interacted coefficient is equal to 0.003 but it is not statistically significant, while the not-interacted coefficient is equal to -0.004 and it is statistically significant at the 1% level (Column 2), therefore the impact of the coefficient for observation with Rule of Law above the median is not statistically different from that of observations with Rule of Law below the median. The interacted coefficient for observations with Media Repression above the median (Stringency Index*Covid searches*Above the median of Media Repression + Stringency Index*Covid searches) is -0.004 and 1% significant, however, the not interacted coefficient due to observations with Media Repression below the median (Stringency Index * Covid searches) is not statistically significant (Column 3). In the case of the Low Education dummy, the interacted coefficient for observations with Low Education above the median (Stringency Index*Covid searches*Above the median of Low Education + Stringency Index*Covid searches) is -0.005 and 5% significant and, the not interacted coefficient due to observations with Law Education below the median (Stringency Index * Covid searches) is not statistically significant (Column 4).

1.6 Autor test

The key identifying assumption for diff-in-diff estimates is that the variation in mobility in countries belonging to the control group is an unbiased estimate of the counterfactual. While we cannot directly test this assumption, we can test whether the time trends in the control and treatment countries were the same in the pre-intervention periods. If the trends are the same in the pre-intervention periods, then it is likely that they would have been the same in the post-intervention period, had the treated countries not adopted any stringency measure. An event-study analysis can shed some light on the validity of the research design. Following Autor (2003),

Table 2: Diff-in-diff estimates with conditioning macro variables.

DV: mobility	(1)	(2)	(3)	(4)
Heterogeneity variables	Rule of Law	Voice and Accountability	Media repression	Education
Stringency Index (SI) * Covid searches (CS)	-0.004***	-0.004***	0.001	-0.001
	(0.001)	(0.001)	(0.002)	(0.002)
SI * CS * Above median macro	0.004**	0.003	-0.006**	-0.005**
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	11,025	11,025	11,025	11,025
R-squared	0.8384	0.8418	0.8355	0.8385
Country FE	YES	YES	YES	YES
Daily FE	YES	YES	YES	YES
Country specific trend	YES	YES	YES	YES

Note: This table shows the effect of Covid searches on mobility for countries with low and high governance index. We split the dataset in two subsets, i.e. 18 countries below the median and 17 countries above the median, both observed for 315 days. Panel Asearches on mobility for countries with low and high levels of Rule of Law, Voice and Accountability, Media Repression and Low Education. Rule of Law, Voice and Accountability and Media Repression are equal to one if a given country is above the median level in our sample, and zero otherwise. Low Education is equal to one if a given country is below the median level of education in our sample, and zero otherwise. We have data for 35 countries and 315 days. We regress mobility on Stringency Index, Covid searches, the interaction between Stringency Index and Covid searchesm, per capita Confirmed Cases, Temperatures and the interaction of all the terms with the four macro-level dummy variables. We include country and daily fixed effects and country specific trends. Robust standard errors are clustered at country level (and shown in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

we create a dummy variable which takes on the value of one on the first day of the stringency index greater than zero, and zero otherwise. We do not introduce this dummy variable directly in our specification but we interact it with the mean of the Stringency Index adopted by each country in order to account for the overall intensity of the government measures. Hence, starting from this variable, we create its leads (one for each day prior the day of the stringency measure) and lags variables (one for each day after the stringency measure was introduced)[26]. If the trends in the mobility measure in adopting versus non-adopting countries are the same, then the leads should not be statistically significant. An attractive feature of this test is that the lags are informative and can show whether the effect changes over time. We

[26]As the number of countries with more 282 lags sharply decreases after the 283rd day from the stringency adoption, we replace each individual lag for the remaining 13 days with a single dummy variable interacted with the mean stringency.

estimate the following specification:

$$
\begin{aligned}
mobility_{cd} =& \alpha + \sum_{\pi=-32}^{-2} \beta_{\pi} first\ day_{c(d+\pi)} * Mean\ stringency_c \\
&+ \sum_{\tau=0}^{283} \beta_{\tau} first\ day_{c(d+\tau)} * Mean\ stringency_c + \gamma\beta X_{cd} + f_c + f_d + u_{cd}
\end{aligned} \tag{4}
$$

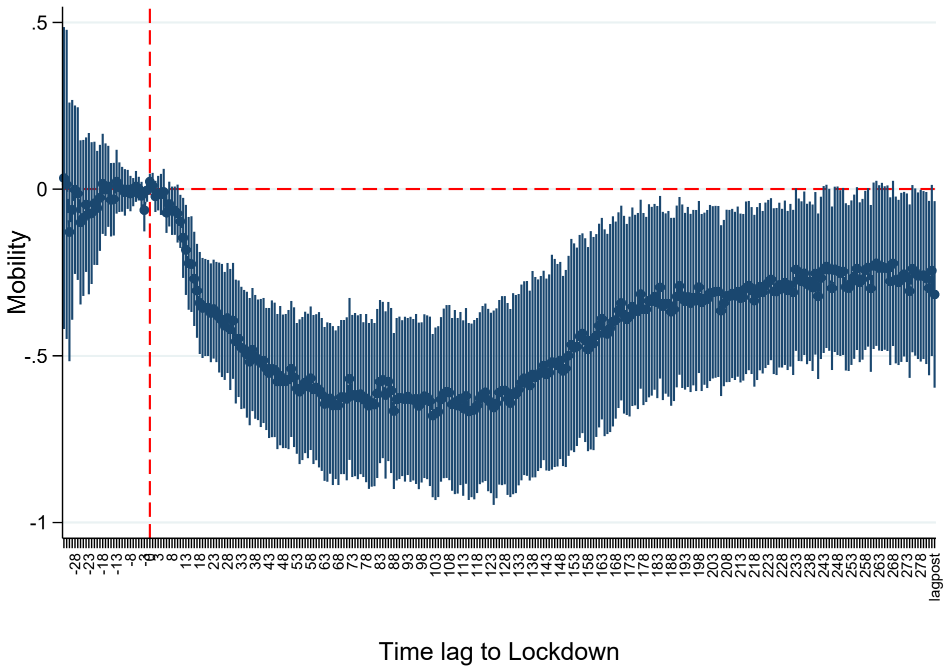
Where $first\ day_{cd}$ is a dummy equal to 1 only in the day the Stringency Index starts being greater than zero in country c and day d. Moreover, $first\ day_{c(d+\pi)}$ is a dummy variable equal to 1 in country c and day $d+\pi$, with π going from -32 to -2:, those dummies which stand for the leads of the variable $first\ day_{cd}$. We also include the lags of the $first\ day_{cd}$ by building the dummies $first\ day_{c(d+\tau)}$ equal to 1 in country c and day $d+\tau$, with τ going from 1 to 283. Finally, we have $Mean\ stringency_c$ which is the mean of the Stringency Index in country c. All the other variables and fixed effects are defined as in Equation 2. This specification allows for testing parallel trends in the pre-treatment period, namely, whether the coefficients associated with the lead (β_{π}, with π going from -32 to -2) are not statistically different from zero. This approach also helps understand whether the treatment effect fades, increases, or stays constant over time, depending on the estimated coefficients on the lags (β_{τ}, with τ going from 1 to 283).

The omitted day is the day before the stringency index becomes greater than zero, which -because of the staggered time of stringency measure adoption- differs by country. For example, in Sweden stringency measure started to be adopted on March 9 2020, therefore there are 13 leads and 270 lags, and omitted day is March 8 2020.

The estimates, together with their 90% confidence intervals, are plotted in Figure 3. According to the point estimates, in the pre-treatment period there is no difference in the movement until around the 10th day after the adoption of the stringency measure.

Turning now to the lag coefficients, we find that the stringency measures contribute to a reduction in mobility, but it takes some days for the effects to materialise. In fact, the coefficient associated with the lags turns out to be negative and statistically significant at the 5% after 11 days since the first day of the stringency measure.

Figure 3: Autor test.



Note: This figure plots estimates β from Equation 4, with their respective pointwise 90% confidence intervals. The plotted estimated coefficient is the interaction between the leads and lags and the mean over the all period of the stringency index adopted by each country. The dependent variable is the Mobility Google Index. The day before the first day of the adoption of the stringency measure is omitted, so the estimates are normalized to zero in that day. The model also includes country and daily fixed effects and temperatures and per capita confirmed cases as covariates. Errors are clustered at country level. The sample include 35 countries observed over 315 days.

From the 11th day after the introduction of the stringency measures, we get a steep decrease in mobility for the following two weeks, followed by a milder decrease up to the 120th day after the introduction of the stringency measure. Afterwards, the estimated coefficient starts increasing and reaches a plateau after the 160th day until the end.

1.7 Correlates of Covid searches

To explain and better understand the panel variation of Covid searches we implement a fixed effects regression analysis. First, in Column (1) of Table 3, we show that the difference in the number of confirmed cases between country i and the average of its four closest neighbouring countries is positively and significantly correlated with Covid searches in country i, with a coefficient of 0.43 (1% confidence level). The intuition is that the search of information by citizens of a given country appears to be driven by the excess of country's own cases vis a vis its neighbouring -and comparablecountries.

Second, in order to identify other observable factors that are significantly associated with Covid searches, we include additional variables in our regressions. In Column (2) we add the Stringency Index and find that it is positively and significantly correlated with Covid searches. Citizens are significantly more interested in searching about COVID-19 when governments implement stricter containment measures. In Column (3) we also include temperature, per capita confirmed cases and the interaction between the Stringency Index and the country-specific level of education.

We find that the coefficient on the interaction of the Stringency Index and the country-specific low education level is negative and statistically significant offsetting the significant positive coefficient not interacted. In other terms, in countries with low levels of education, the Stringency Index is not significantly correlated with the outcome variable, while in countries with high levels of education the Stringency Index has a positive and statistically significant correlation with Covid searches.

Table 3: Regression explaining Covid searches.

DV: Covid searches	(1)	(2)	(3)
Diff in number	0.426***	0.367***	0.257**
of conf. cases	(0.100)	(0.106)	(0.123)
Stringency Index (SI)	-	0.304***	0.351***
		(0.092)	(0.090)
SI * Low Education	-	-	-0.311*
			(0.104)
Temperatures	-	-	-0.025
			(0.029)
Constant	40.27 ***	23.93***	34.02***
	(0.06)	(4.94)	(7.62)
Observations	11,025	11,025	11,025
R-squared	0.68	0.70	0.71
Country FE	YES	YES	YES
Daily FE	YES	YES	YES
Country specific trend	YES	YES	YES

Note: The dependent variable is the amount of Covid searches in country i in day t. As explanatory variables, in Column 1 we include the difference in confirmed cases, i.e., the difference between the 7-days moving average of country i and the mean of the 7-days moving average of confirmed cases in the four closest neighbouring countries; in Column 2 we add the Stringency Index; in Column 3 we add temperatures, and the interaction between the Stringency Index and Low Education, a dummy variable which is equal to one when the country's level of education (as defined in Section 1.4.2) is below the median, and zero otherwise. Our dataset includes 35 countries and 315 days. We include country and daily fixed effects and country specific trends. Robust standard errors are clustered at country level (and shown in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

1.8 Robustness checks

We use a series of robustness checks to address possible issues related to the research design that could bias our baseline estimates. In Section 1.8.1 we replace the main dependent variable by excluding one by one each component of the Google mobility indicator; in Section 1.8.2 we perform a country-sensitive test to show that the estimated effects do not depend from a specific country; finally, in section 1.8.3 we perform a falsification test, replacing the Covid searches with other relevant terms searched on Google during the same time-span.

1.8.1 Alternative dependent variables

The dependent variable used in the main regression (Table 1) is a composite indicator which is calculated as the daily average of the mobility for visits to the following destinations: (i) retail & recreation, (ii) workplaces, (iii) grocery & pharmacy, (iv) transit stations, and (v) parks.

To check whether results are not driven by a specific individual component of the Google mobility composite indicator, in Table 4 we exclude one component at a time from the dependent variable[27]. The coefficient on the interaction term Stringency Index*Covid searches remains negative and statistically significant in all specifications, which is consistent with our results not essentially depending on one particular component of the index.

1.8.2 Country sensitivity analysis

We also test whether our main findings are sensitive to the exclusion of a single country. For this reason, we estimate Equation 3, by dropping one country at a time. The estimated coefficients of the interaction term Stringency Index*Covid searches and their 95% confidence interval (Figure 4) are very similar to those obtained in our baseline specification. Hence, it can be concluded that our main result is not driven by a particular country.

[27]As a robustness check, we also replace the composite Google mobility indicator with its individual components, namely: workplaces, parks, transit stations, and retail & recreation. In the case of mobility to workplaces, transit and retail & recreation we find very similar results to our baseline ones. On the other hand, the result on mobility to/from parks is slightly smaller in size and not statistically significant at ordinary confidence levels (see Table A6 in Appendix A).

Table 4: Diff-in-diff estimates, using alternative dependent variables.

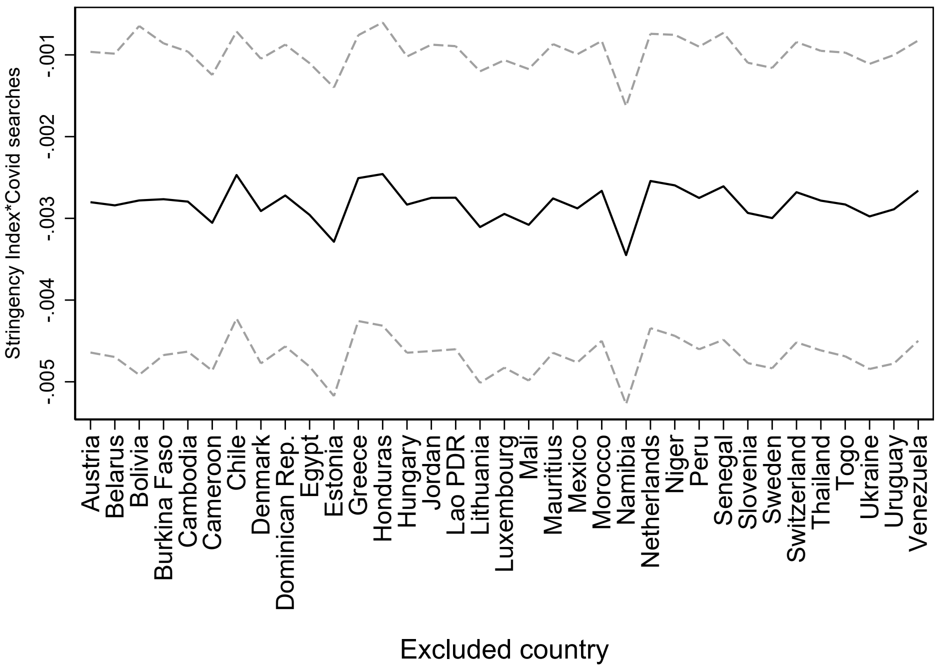
	(1) Mobility without retail and recreation	(2) Mobility without workplaces	(3) Mobility without grocery and pharmacy	(4) Mobility without transit stations	(5) Mobility without parks
Stringency Index	-0.318***	-0.344***	-0.342***	-0.313***	-0.304***
	(0.075)	(0.072)	(0.073)	(0.075)	(0.073)
Covid searches	0.109	0.089	0.118	0.125	0.131
	(0.085)	(0.085)	(0.086)	(0.090)	(0.080)
SI*Covid searches	-0.003**	-0.003**	-0.003**	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
SI*Conf. cases pc	0.002	0.001	0.000	0.001	-0.001
	(0.004)	(0.005)	(0.005)	(0.004)	(0.002)
Conf. cases pc	-0.331	-0.276	-0.235	-0.220	-0.078
	(0.351)	(0.362)	(0.377)	(0.340)	(0.150)
Temperatures	0.132***	0.149***	0.136***	0.134***	0.039***
	(0.022)	(0.023)	(0.023)	(0.021)	(0.011)
Observations	11,025	11,000	11,025	11,025	11,025
R-squared	0.78	0.80	0.80	0.79	0.79
Country FE	YES	YES	YES	YES	YES
Daily FE	YES	YES	YES	YES	YES
Country trend	YES	YES	YES	YES	YES

Note: Notes: This table shows the effect of Covid searches on different measures of mobility. In each Column from (1) to (5) we average the mobility indicator removing one component at a time, namely we exclude retail and recreation (1), workplaces (2), grocery and pharmacy (3), transit stations (4), and parks (5). We regress different dependent variables on Stringency Index (SI), Covid searches, the interaction between Stringency Index and Covid searches, per capita Confirmed Cases and its interaction with Stringency Index and temperatures, as in Equation 3. We include country and daily fixed effects and country specific trends. The dataset is a country by day panel, for 35 countries and 315 days. Robust standard errors are clustered at country level (and shown in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

1.8.3 Falsification exercise on Google searches

Within our diff-in-diff analysis we conduct a placebo test to simulate how alternative Google searches that are unrelated to the pandemic might impact mobility. This test arises from the concern that Covid related searches could be endogenous to mobility, e.g., the week by week volume of Google searches can be correlated with the fact of staying at home, i.e., with lower mobility. If the relationship between Covid searches and mobility were spurious, namely due to the stay-at-home order which causes more searching activity on Google, using our placebo variables we would get similar results to the ones obtained in the baseline specification which makes use of "Covid" searches. Specifically, we replicate the main analysis in Equation 3 by replacing Covid searches with the main three terms searched in Google in the year 2020 (translate, porn, and maps). Notice, moreover, that these terms are most likely

Figure 4: Country sensitivity analysis.



Note: The figure shows the coefficient of interest, ϑ, of Equation 3 and the relative 95% confidence interval, for different set of countries. We exclude from the original set of 35 countries one country at a time (reported on the x-axis). We include country and daily fixed effects and country specific trends. The dataset is a country by day panel, for 34 countries and 315 days. Robust standard errors are clustered at country level.

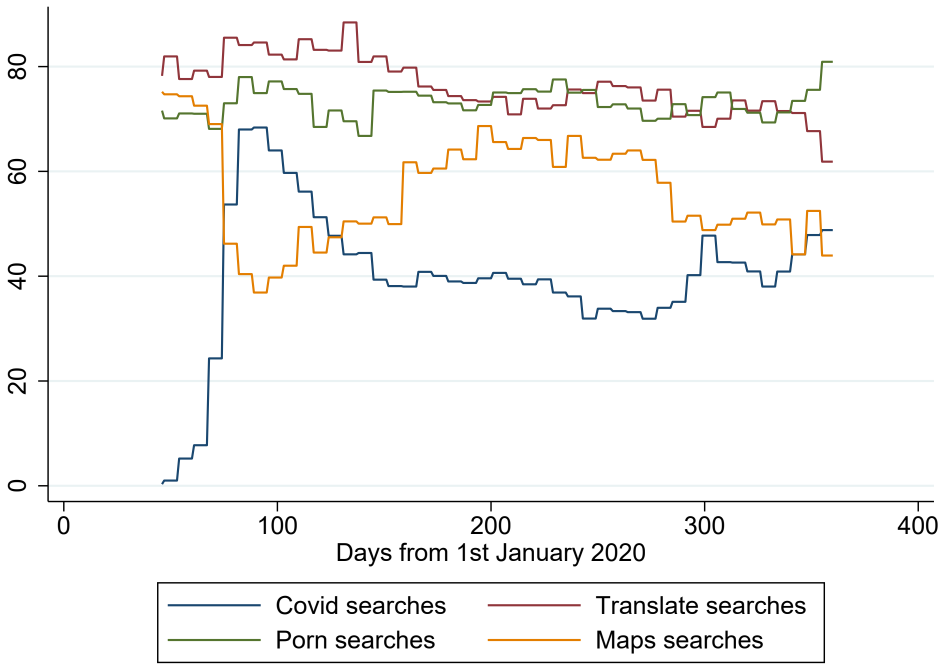
unrelated with the term Covid. The graphical analysis in Figure 5 shows that the searches for "porn", "translate" and "maps" are not correlated with the searches for Covid: the Pearson correlation index is respectively equal to 0.16, 0.17 and -0.09. As a preliminary analysis, from Figure 5, we observe that these three Google searches are not correlated with the Covid searches in the time-span of our analysis, but still relevant in terms of intensity.

In Table 5, we use as explanatory variable "translate" searches (Column 1), "porn" searches (Column 2) and "maps" searches (Column 3). In all specifications we find that the coefficients on the interaction terms are statistically indistinguishable from zero: thus, Google searches different from Covid-19 apparently do not affect the impact of stringency on mobility.

1.8.4 Scaled Covid searches

The variable "Google searches" we use in our main analysis (Table 1) is a weekly intensity, which is measured as the number of weekly searches for the term divided

Figure 5: Covid searches and fake Google searches.



Note: The figure describes Covid searches, Porn searches, Translate searches, and Maps searches over 2020, from February 15 (day 46) to December 30 (day 365). Observations for 35 countries are averaged by day. All the variables vary from 0 to 100.

by the maximum number of its weekly searches over the whole time period, within each country, and scaled to 100 for easier readability.

On the other hand, Google searches can be re-scaled at the aggregate level, i.e. jointly considering all sampled countries. In order to rescale the variable, we proceed as follows. First, we find the country with the maximum number of searches (i.e. Chile) in our sample of 35 countries. Then we collect the data from the other countries in groups of five from Google Trends, always including the leading country (Chile). Afterwards, we use the ratio between the leading country of two different groups to compare the observations from the remaining countries. Eventually, we come up with a dataset with variables from 0 to 100, where the maximum value of 100 is only reached by Chile on April 2020 (Brodeur et al., 2021).

We replicate our baseline specifications by replacing the original Google searches with the scaled version (Table A5): we find very similar results, and in particular a negative and significant coefficient on the interaction term Stringency Index*Scaled Covid searches.

Table 5: Diff-in-diff estimates, falsification test.

DV: Mobility Google searches:	(1) Translate	(2) Porn	(3) Maps
Stringency Index	-0.228	-0.501***	-0.495***
	(0.161)	(0.137)	(0.096)
Google searches	0.042	0.037	0.075
	(0.085)	(0.083)	(0.050)
Stringency Index*Google searches	-0.003	0.001	0.001
	(0.002)	(0.002)	(0.001)
Stringency Index*Confirmed cases	-0.001	-0.001	-0.001
	(0.004)	(0.004)	(0.003)
Confirmed cases per capita	-0.155	-0.149	-0.102
	(0.297)	(0.292)	(0.277)
Temperatures	0.114***	0.116***	0.106***
	(0.018)	(0.020)	(0.017)
Observations	10,696	10,696	10,696
R-squared	0.82	0.82	0.82
Country FE	YES	YES	YES
Daily FE	YES	YES	YES
Country Trend	YES	YES	YES

Note: Notes: This table shows the effect of different fake searches on mobility. In Column from (1) to (3) we report searches for "Translate" (1), "Porn" (2), and "Maps" (3). We regress Mobility on Stringency Index, fake search, the interaction between Stringency Index and the fake search, per capita Confirmed Cases and their interaction with the Stringency Index and temperatures, as in Equation 3. We include country and daily fixed effects and country specific trends. The dataset is a country by day panel, for 35 countries and 315 days. Robust standard errors are clustered at country level (and shown in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

1.9 Conclusions

In this chapter we used a staggered diff-in-diff to show that the containment measures are successful in reducing mobility and that the effect of the stringency index is sensitive to citizens' knowledge about the pandemic, which we proxy as Google Covid searches.

From estimating Equation 2 we find that mobility decreases by 11.42 percentage points when considering a shift from a country in the 25th percentile of the stringency measure and Covid searches to a country in the 75th percentile of the stringency measure and Covid searches.

Afterwards, we add the interaction term between the treatment variable and the Covid searches as shown in Equation 3. We show that the decrease in mobility due to the implementation of stringency measures is sensitive to citizens' knowledge about the severity of the pandemic itself. More precisely, mobility is reduced by 35.90 percentage points when the Stringency Index and the Covid searches are the closest to their 75th percentile values; conversely, when the Stringency Index and the Covid searches are the closest to their 25th percentile value, the reduction in mobility is equal to 19.10 percentage points. In conclusion, the Google search interaction term enhances the impact of the containment measures on the stringency index by about 47%. When we perform the heterogeneity analysis, we find that the effect of the Covid searches on the effectiveness of containment measures is driven by countries with low levels of education and low levels of quality of governance. Our intuition behind this result is that the in case of low trust in political institutions citizens' compliance with regulation is more likely guided by individual level of knowledge of the severity of the pandemic.

Our conclusion is that the knowledge about the pandemic is crucial in making the containment measures effective, and that this result is driven by countries with low levels of education and low institutional quality. This result suggests that coercive regulation must be supported by a proper communication plan. In particular, the suggested communication plan has to be considered by countries with low levels of governance institutions.

References

(1) [dataset] Google LLC (2020). "Google COVID-19 Community Mobility Reports", available: www.google.com/covid19/mobility/.

(2) [dataset] Hale, T., Petherick, A., Phillips, T., Webster, S. (2020). "Variation in Government Responses to COVID-19", Version 4.0, *Blavatnik School of Government Working Paper*, available: www.bsg.ox.ac.uk/covidtracker.

(3) [dataset] National Oceanic and Atmospheric Administration (2020). Global Historical Climate Network Daily https://www.ncdc.noaa.gov/ghcnd-data-access.

(4) Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., Yang, D. (2020). "Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic", *NBER Working Paper*, 26946.

(5) Ashofteh, A., Bravo, J. M. (2020). "A Study on the Quality of Novel Coronavirus (COVID-19) official datasets", *Statistical Journal of the IAOS*, 36(2), 291-301. https://doi.org/10.3233/SJI-200674.

(6) Autor, D.H. (2003). "Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing", *Journal of Labor Economics*, 21, 1–42, doi: 10.1086/344122.

(7) Bargain, O., Aminjonov, U. (2020). "Trust and compliance to public health policies in times of COVID-19", *IZA Discussion Paper* 13205.

(8) Briscese, G., Lacetera, N., Macis, M., Tonin, M. (2020). "Compliance with COVID-19 social-distancing measures in Italy: The role of expectations and duration", *NBER Working Paper*, 26916.

(9) Brodeur, A., Clark, A.E., Fleche, S., Powdthavee, N. (2021). "COVID-19, lockdowns and well-being: Evidence from Google Trends", *Journal of Public Economics*, Volume 193, 104346, ISSN 0047-2727, https://doi.org/10.1016/j.jpubeco.2020.104346.

(10) Brodeur, A, Grigoryeva, I., Kattan, L. (2020). "Stay-at-home orders, social distancing and trust", *IZA DP*, 13234.

(11) Carneiro, H.A., Mylonakis, E. (2009). "Google Trends: a web-based tool for real-time surveillance of disease outbreaks", *Clin. Infect. Dis.*, 49 (10), 1557–1564.

(12) Cucinotta, D., Vanelli, M. (2020). "WHO Declares COVID-19 a Pandemic", *Acta Biomed*, 19, 91(1), 157-160. doi: 10.23750/abm.v91i1.9397.

(13) Everett, J.A.C., Colombatto, C., Chituc, V., Brady, W.J., Crockett, M. (2020). "The effectiveness of moral messages on public health behavioral intentions during the COVID-19 pandemic", *PsyArXiv*, March 20. doi:10.31234/osf.io/9yqs8.

(14) Farzanegan, M.R., Hofmann, H.P. (2021). "A Matter of Trust? Political Trust and the Covid-19 Pandemic", *CESifo Working Paper*, 9121.

(15) Ferraresi, M., Kotsogiannis, C., Rizzo, L. and Secomandi, R. (2020). "The 'Great Lockdown' and its determinants", *Economics letters*, 197, p. 109628.

(16) Fornero, E., Oggero, N., Puglisi, R. (2019). "Information and Financial Literacy for Socially Sustainable NDC Pension Schemes", *World Bank discussion paper*, Social protection & jobs, n. 1924, doi: 10.1596/978-1-4648-1455-6_ch25.

(17) Frisch, R., Waugh, F.V. (1933). "Partial Time Regressions as Compared with Individual Trends", *Econometrica: Journal of the Econometric Society*, 387–401.

(18) Goldstein, P., Yeyati, E.L., Sartorio, L. (2021). "Lockdown Fatigue: The Diminishing Effects of Quarantines on the Spread of COVID-19", *CEPR*, 67.

(19) Helsingen, L. M., Refsum, E., Gjøstein, D. K., Løberg, M., Bretthauer, M., Mette Kalager, M., Emilsson, L. (2020). "The COVID-19 Pandemic in Norway and Sweden – Threats, Trust, and Impact on Daily Life: A Comparative Survey", *BMC Public Health*, 20 (1).

(20) Hyland-Wood, B., Gardner, J., Leask, J., Ecker, U. (2021). "Toward effective government communication strategies in the era of COVID-19", *Humanities and Social Sciences Communications*, 8, 30, doi: 10.1057/s41599-020-00701-w

(21) Keene, D.R. (2020). "Leviathan challenged – is the lockdown EHCR compliant?", *Working paper*, Crown Office Row.

(22) Levy Yeyati, E., Filippini, F. (2021). "The social and economic impact of Covid-19", *VoxEU.org*, 12 May.

(23) Lovell, M. C. (1963). "Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis", *Journal of the American Statistical Association*, 58 (304), 993–1010.

(24) Martin, G. J. Yurukoglu, A. (2017). "Bias in Cable News: Persuasion and Polarization", *American Economic Review*, 107(9), 2565–99.

(25) Mastrorocco, N. Minale, L. (2018). "News Media and Crime Perceptions: Evidence from a Natural Experiment", *Journal of Public Economics*, 165, 230–255.

(26) Oosterhoff, B., Palmer, C. (2020). "Psychological Correlates of News Monitoring, Social Distancing, Disinfecting, and Hoarding Behaviors Among US Adolescents During the COVID-19 Pandemic", *PsyArXiv*.

(27) Painter, M., Qiu, T. (2020). "Political Beliefs affect compliance with COVID19 social distancing orders, *SSRN* Posted on April 14, 2020.

(28) Preis, T., Moat, H.S., Stanley, H.E. (2013). Quantifying trading behavior in financial markets using Google Trends", *Sci. Rep.*, 3, 1684.

(29) Shi, L. (2009). "The Limit of Oversight in Policing: Evidence from the 2001 Cincinnati Riot", *Journal of Public Economics*, 93(1-2), 99–113.

(30) Sousa-Pinto, B., Anto, A. Wienia Czarlewski, W., Anto, J.M., Fonseca, J.A., Bousquet, J. (2020). "Assessment of the Impact of Media Coverage on COVID-19–Related Google Trends Data: Infodemiology Study", *Journal of medical internet research*, 22(8), e19611, doi: 10.2196/19611

(31) Spenkuch, J. L. Toniatti, D. (2018). "Political Advertising and Election Results", *The Quarterly Journal of Economics*, 133(4), 1981–2036.

(32) Stokes, A.C., Lundberg, D.J., Elo, I.T., Hempstead, K., Bor, J. (2021). "COVID-19 and excess mortality in the United States: A county-level analysis", *PLOS Medicine*, 18(5), doi:10.1371/journal.pmed.1003571.

(33) Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., Tatlow, H. (2021).

"A global panel database of pandemic policies", (Oxford COVID-19 Government Response Tracker) *Nature Human Behaviour*.

(34) Velásquez, D., Medina, S., Yamada, G., Lavado, P., Nunez-del Prado, M., Alatrista-Salas, H. Morzán, J. (2020). "I Read the News Today, Oh Boy: The Effect of Crime News Coverage on Crime Perception", *World Development*, 136.

(35) Koh, W.C., Naing, L., Wong, J. (2020). "Estimating the impact of physical distancing measures in containing COVID-19: an empirical analysis", *International Journal of Infectious Diseases*, 100, 42-49.

A Appendix

Table A1: List of countries in the sample.

Austria	Mali
Belarus	Mauritius
Bolivia	Mexico
Burkina Faso	Morocco
Cambodia	Namibia
Cameroon	Netherlands
Chile	Niger
Denmark	Peru
Dominican Republic	Senegal
Egypt	Slovenia
Estonia	Sweden
Greece	Switzerland
Honduras	Thailand
Hungary	Togo
Jordan	Ukraine
Lao PDR	Uruguay
Lithuania	Venezuela
Luxembourg	

Table A2: Summary statistics.

	(1) N.	(2) Mean	(3) Std. Dev.	(4) Min.	(5) Max.
Covid searches	11,025	38.94	26.44	0	100
Confirmed cases per capita	11,025	5.779	13.91	-28.51	114.3
Diff. in number of conf. cases with neighb.	11,025	-0.606	10.07	-64.83	77.13
Maps searches	10,701	57.63	23.94	0	100
Mobility	11,025	84.79	24.18	6.800	175
Mobility without retail and recreation	11,025	86.83	24.84	7.500	194
Mobility without workplaces	11,000	86.35	27.31	6	207.8
Mobility without grocery and pharmacy	11,025	82.86	26.00	7.250	191.8
Mobility without transit stations	11,025	88.19	25.33	7.250	198
Mobility without parks	11,025	79.74	20.47	6	146.5
Porn searches	10,701	73.17	14.90	14	100
Stringency Index	11,025	53.60	25.90	0	100
Temperatures	11,025	200.5	90.99	-77.50	388
Translate searches	10,701	76.16	15.05	7	100

Note: This table provides summary statistics. For more details about the variables, see Section 1.3.

Table A3: Diff-in-diff estimates, main specification, all countries.

DV: Mobility	(1)	(2)	(3)	(4)	(5)	(6)
Stringency Index	-0.505***	-0.446***	-0.432***	-0.312***	-0.386***	-0.384***
	(0.042)	(0.036)	(0.036)	(0.049)	(0.051)	(0.051)
Conf. cases	-0.264***	-0.228***	-0.171***	-0.177***	-0.203***	-0.055
per capita	(0.058)	(0.057)	(0.059)	(0.057)	(0.066)	(0.261)
Temperatures	-	0.085***	0.086***	0.086***	0.082***	0.083***
		(0.010)	(0.010)	(0.009)	(0.010)	(0.010)
Covid searches	-	-	-0.075***	0.153**	0.100	0.080
			(0.026)	(0.066)	(0.064)	(0.066)
Stringency Index*	-	-	-	-0.003***	-0.003***	-0.003***
Covid searches				(0.001)	(0.001)	(0.001)
Stringency Index*	-	-	-	-	-	-0.002
Conf. cases pc						(0.004)
Observations	34,320	34,320	34,005	34,005	34,005	34,005
R-squared	0.74	0.77	0.77	0.77	0.99	0.99
Country FE	YES	YES	YES	YES	YES	YES
Daily FE	YES	YES	YES	YES	YES	YES
Country specific trend	NO	NO	NO	NO	YES	YES

Note: This table shows the effect of Covid searches on mobility. We regress country's mobility on different variables. In Column (1) Stringency Index and per capita Confirmed Cases (Eq. (1)); in column (2) we additionally control for temperatures (Eq. (1)); in Column (3) we add Covid searches as a control (Eq. (1)); in Column (4) we also include the interaction between Stringency Index and Covid searches (Eq. (2)); finally, in Column (6) we additionally include the interaction term between Stringency Index and per capita Confirmed Cases (Eq. (2)). For all specifications we include country and daily fixed effects. In Column 5 and 6 we also include country specific trend. The dataset is a country by day panel, for 109 countries and 315 days. Robust standard errors are clustered at country level (and shown in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A4: Diff-in-diff estimates, confirmed deaths in place of confirmed cases.

DV: Mobility	(1)	(2)	(3)	(4)	(5)	(6)
Stringency Index	-0.537***	-0.443***	-0.412***	-0.289***	-0.345***	-0.340***
	(0.056)	(0.053)	(0.056)	(0.061)	(0.069)	(0.069)
Conf. Deaths	-6.23***	-3.494***	-2.012	-1.739	-2.304	-14.520
per capita	(1.93)	(1.636)	(1.519)	(1.467)	(1.578)	(9.200)
Temperatures	-	0.126***	0.128***	0.125***	0.118***	0.124***
		(0.018)	(0.019)	(0.018)	(0.020)	(0.020)
Covid searches	-	-	-0.100**	0.100	0.126	0.108
			(0.039)	(0.086)	(0.020)	(0.082)
Stringency Index*	-	-	-	-0.003***	-0.003**	-0.003**
Covid searches				(0.001)	(0.001)	(0.001)
Stringency Index*	-	-	-	-	-	0.158
Conf. deaths pc						(0.113)
Observations	10,675	10,675	10,675	10,675	10,675	10,675
R-squared	0.753	0.793	0.798	0.801	0.821	0.822
Country FE	YES	YES	YES	YES	YES	YES
Daily FE	YES	YES	YES	YES	YES	YES
Country specific trend	NO	NO	NO	NO	YES	YES

Note: This table shows the effect of Covid searches on mobility. We regress country's mobility on different variables. In Column (1) Stringency Index and 14 days lag of per capita confirmed deaths (Eq. (1)); in column (2) we additionally control for temperatures (Eq. (1)); in Column (3) we add Covid searches as a control (Eq. (1)); in Column (4) we also include the interaction between Stringency Index and Covid searches (Eq. (2)); finally, in Column (6) we additionally include the interaction term between Stringency Index and 14 days lag of per capita confirmed deaths (Eq. (2)). For all specifications we include country and daily fixed effects. In Column 5 and 6 we also include country specific trend. The dataset is a country by day panel, for 35 countries and 315 days. Robust standard errors are clustered at country level (and shown in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A5: Diff-in-diff estimates, with Scaled Covid searches.

DV: Mobility	(1)	(2)	(3)	(4)	(5)	(6)
Stringency Index	-0.489***	-0.412***	-0.405***	-0.289***	-0.367***	-0.368***
	(0.059)	(0.055)	(0.055)	(0.064)	(0.063)	(0.062)
Conf. cases	-0.345***	-0.246***	-0.198**	-0.198**	-0.196*	-0.261
per capita	(0.100)	(0.088)	(0.088)	(0.088)	(0.098)	(0.306)
Temperatures	-	0.118***	0.120***	0.121***	0.119***	0.119***
		(0.017)	(0.017)	(0.016)	(0.020)	(0.020)
Covid searches scaled	-	-	-0.210***	0.458***	0.334**	0.349**
			(0.086)	(0.111)	(0.128)	(0.150)
Stringency Index*	-	-	-	-0.008***	-0.006**	-0.006**
Covid searches scaled				(0.002)	(0.002)	(0.002)
Stringency Index*	-	-	-	-	-	0.0001
Confirmed cases pc						(0.0038)
Observations	11,025	11,025	11,025	11,025	11,025	11,025
R-squared	0.755	0.789	0.791	0.797	0.817	0.817
Country FE	YES	YES	YES	YES	YES	YES
Daily FE	YES	YES	YES	YES	YES	YES
Country specific trend	NO	NO	NO	NO	YES	YES

Note: This table shows the effect of Covid searches scaled on mobility. We regress country's mobility on different set of variables. In Column (1) Stringency Index and per capita Confirmed Cases (Eq. 1); in Column (2) we additionally control for temperatures (Eq. 1); in Column (3) we add Scaled Covid searches as a control (Eq. 1); in Column (4) we also include the interaction between Stringency Index and Scaled Covid searches (Eq. 2), while in Column (5) we add country-specific linear trends; finally, in Column (6) we additionally include the interaction term between Stringency Index and per capita Confirmed Cases (Eq. 2). For all specifications we include country and daily fixed effects. In columns 5 and 6 we also include country specific trends. The dataset is a country by day panel, for 35 countries and 315 days. Robust standard errors are clustered at country level (and shown in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A6: Diff-in-diff estimates, separate component of the dependent variable.

DV:	(1) Workplaces	(2) Parks	(3) Transit stations	(4) Retail & recreation
Stringency Index	-0.270***	-0.395**	-0.383***	-0.352***
	(0.085)	(0.183)	(0.066)	(0.076)
Conf. cases per capita	-0.008	-1.306	-0.277	-0.194
	(0.155)	(1.021)	(0.229)	(0.179)
Temperatures	-0.011	0.433***	0.059***	0.060***
	(0.012)	(0.074)	(0.015)	(0.013)
Covid searches	0.183*	0.097	-0.063	0.134
	(0.094)	(0.198)	(0.065)	(0.083)
Stringency Index*Covid searches	-0.003**	-0.004	-0.002*	-0.003**
	(0.001)	(0.002)	(0.001)	(0.001)
Stringency Index*Conf. cases pc	-0.002	0.012	0.001	-0.006**
	(0.002)	(0.014)	(0.003)	(0.002)
Observations	11,022	10,839	10,900	10,972
R-squared	0.720	0.773	0.863	0.839
Country FE	YES	YES	YES	YES
Daily FE	YES	YES	YES	YES
Country trend	NO	NO	YES	YES

*Note:*This table shows the effect of Covid searches on different indicators of mobility. The dependent variables are: workplaces, parks, transit stations, and retail & recreation. In each column, we regress country's alternative mobility indicators on Stringency Index, per capita Confirmed Cases, Temperatures, Covid searches, the interaction between Stringency Index and Covid searches, the interaction term between Stringency Index and per capita Confirmed Cases, country-specific linear trends, country and daily fixed effects. The dataset is a country by day panel, for 35 countries and 315 days. Robust standard errors are clustered at country level (and shown in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Chapter two

2 Taxing Big Firms' Extra Profits: evidence from the "Robin Hood" Tax

2.1 Introduction

The containment measures related to the COVID-19 pandemic and the massive public expenditure to face the health emergency resulted in severe economic drawbacks. The idea of a tax hike targeting specific sectors, companies or people that are seen as having come out of this crisis relatively well-off is gaining political traction. For instance, the International Monetary Fund is evaluating the possibility of introducing a series of temporary taxes to help fund the pandemic recovery, in particular the Fund proposed a temporary tax levied on companies who gained unusually high profits due to the COVID-19 pandemic. Likewise, during the Great Recession, several states in the U.S. (e.g., Connecticut, North Carolina and Oregon) introduced temporary corporate income tax surcharges. In this chapter we estimate the effect of a surtax that Italy introduced after the great recession. This surtax, called the "Robin Hood" Tax (RHT), is levied on profits and it targets big firms operating in specific sectors that, during the great recession, benefited from abrupt extra profits [28].

A key challenge is that raising taxes can make things worse because firms are likely to react to the introduction of the additional surcharge and shift the additional cost either onto capital or onto labor.

In the first case, corporate tax hikes substantially raise firms' cost of capital (see Harberger, 1962, 1966; Feldstein, 1970; Poterba, Rotemberg, and Summers, 1985) and make firms less prone to invest (see Hall and Jorgenson, 1967; Cummins, Hassett, and Hubbard, 1994; Caballero, Engel, and Haltiwanger, 1995). In fact, the depressing effect of corporate tax hikes on short-term output is often used as an argument against increasing taxes on corporations (Ryan, 2011, 2012; Hubbard et al. 2012). Nonetheless, larger and high-profit firms seem to be less sensitive to tax increase (see Lediga et al, 2019; Devereux et al, 2014, for recent empirical evidence). In broad terms, tax policies can either target investments directly - such as depreciation changes[29]

[28]See Financial Post, "IMF Proposes Temporary Solidarity Tax on Pandemic Winners and the Wealthy", 7 April 2021.

[29]See Zwick and Mahon, 2017, who find that bonus depreciation has a substantial effect on

- or affect the cost of capital gradually over time - such as corporate or dividend tax changes (Feldstein 1982; Auerbach and Hassett 1992; Yagan 2015). However, in contrast with standard economic theory, Yagan (2015) has shown a null impact of increasing dividend's tax rate. The relevant outcome variables he analyses are investment, employee compensation and payout to shareholders. Yagan finds an immediate behaviour response on payout, but a null effect on investment and employee compensation.

Corporate tax hikes might as well affect wages and employment and be shifted onto labor. In fact Bradford (1978) and Kotlikoff and Summers (1987) showed that if capital is perfectly mobile between countries but labour is not, the introduction of a tax on corporate income in a home country tends to reduce the world rate of return to capital and to shift capital away from the country to the rest of the world. This shift in capital reduces the return to labour in the home country, decreasing wages. Moreover, corporate taxes can also be passed on to workers through a bargaining mechanism between firms and workers. In this case both the wage and the number of workers is decided by the bargaining between unions of workers and firms, while firms choose capital by maximizing their profits (Arulampalam et al. 2012). In this model for given level of inputs (meaning that the number of workers and capital are fixed), an increase in corporate tax determines a decrease in wages and profits which is related, respectively, to the bargaining power of workers and firms. Given a hike in the corporate tax, higher bargaining power of workers will lead to a higher decrease in wage and a lower decrease in profits: in fact the higher the workers' bargaining power the bigger the quota of the quasi rent they get and the greater is the the reduction of their wage due to the increase in corporate tax. In particular, Arulampalam et al. (2012) find that an increase in corporate tax is 50% shifted onto wages. This estimation corresponds to the direct effect related to wages which is defined as the reduction of the quasi rent over which there is bargaining for given level of inputs. Consistently with these findings Fuest et al. (2018) find that the extent to which corporate tax is shifted onto labor depends on the profitability of firm because the higher the rents to be shared between firms and workers, the higher the pass-through on wages. Moreover, they find that the decrease in wage due to the corporate tax is stronger in the case of firm level bargaining than sectoral level collective.

---

investment.

To the best of our knowledge, past empirical works have not fully explored the economic implications of imposing a temporary tax on big firms. To address this question empirically, we exploit the introduction of the RHT in Italy over the 2008-2015 period. This policy led to the imposition of a surtax on profits of firms reporting revenues above 25 million euros and operating in the energy sector.

We use financial data on Italian firms from the *Analisi Informatizzata delle Aziende Italiane* (AIDA) dataset, provided by the Bureau van Dijk. We retrieve firm-level information on annual revenues, total assets, capital, value added, profits, labor cost, and total cost, giving us the possibility to analyze multiple margins of responses to the surtax. We focus on firms reporting revenues between 5 and 45 million euros during the 2007-2011 period[30]. Firstly, we perform a difference-in-differences where the treatment variable is the introduction of the "Robin Hood" Tax and we find no significant change in all the relevant outcome variables. Secondly, we confirm the estimated null effect with a regression discontinuity design. These results are robust to a set of robustness checks.

Our results are split between firms' response with respect to investments and to labor. We show that firms do not react to the temporary tax hike by modifying neither their investment behavior nor their labor cost. These results are coherent with the findings from Yagan (2015) who shows that a tax increase resulted in a null impact on investments and employee compensation. Our findings are also consistent with Lediga et al. (2019) and Devereux et al. (2014) who show that larger firms are less sensitive to tax hikes. Nonetheless, our findings are also in contrast with Bradford (1978) and Kotlikoff and Summers (1987) who find that in case of mobile capital and fixed labor a tax hike reduces wages, and with a wide branch of the literature on corporate tax finding that tax hikes result in lower investments (Hall and Jorgenson, 1967; Cummins, Hassett, and Hubbard, 1994; Caballero, Engel, and Haltiwanger, 1995). In our quasi-experimental design we use the introduction of a temporary tax levied on big firms that gives us a unique perspective on the big firms' response to an additional tax. We estimate a null effect by using multiple econometric models and we confirm our findings with a set of robustness checks that validate our null results.

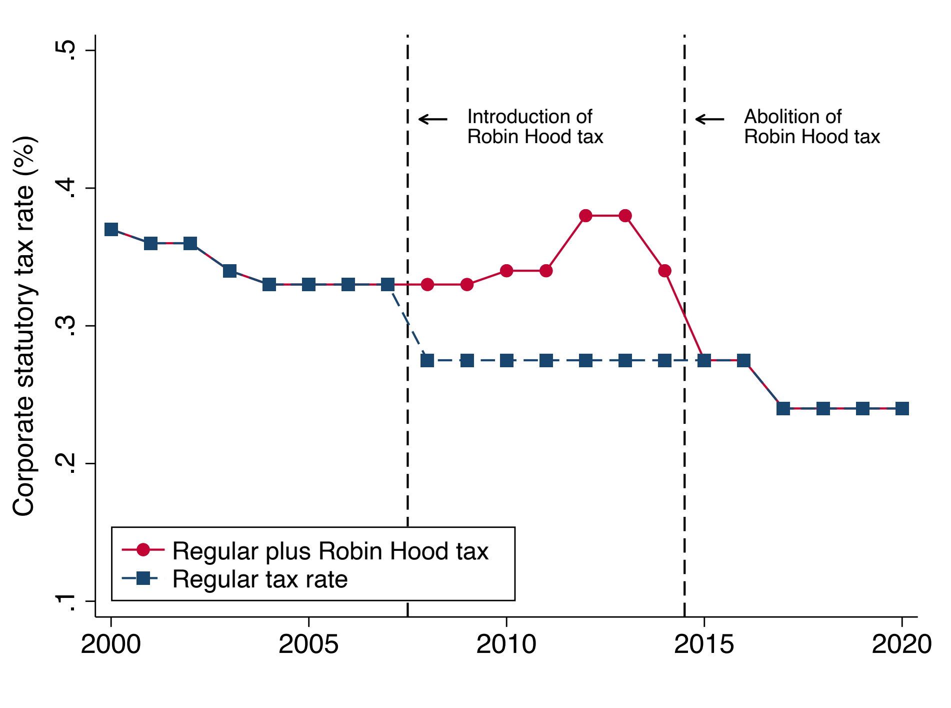
[30]We limit the analysis to this time range because across these years the revenue threshold remained constant at 25 million euros. For further details, see Section 2.2.2.

2.2 Institutional Background

2.2.1 Corporate Taxation in Italy

Italian companies are subject to a corporate income tax, known as *Imposta sul Reddito delle Società* (IRES). Similarly to most advanced economies, the corporate tax is charged on net profits (i.e., gross income minus allowable tax reliefs) of a company and the tax base is determined according to the worldwide taxation principle.[31] This corporate tax is relevant in terms of magnitude as, according to OECD Tax Database, corporate taxation in Italy accounted for around 2 percent of the GDP (or about 4.5 percent of taxation) in 2019. The IRES rate has registered a gradual reduction in the last twenty years: since 2016, it is set at 24%, while it decreased from 33% to 27.5% in 2008 (Figure 6).

Figure 6: Corporate Taxation in Italy.



2.2.2 The "Robin Hood" Tax

The introduction of the RHT had a redistributive objective by taxing, in a period of crisis, the windfall profits of the energy sector, namely those extra-profits resulting from an increase in speculation on commodity prices, such as crude oil and gas. At

[31] Non-resident companies are taxed only on Italian-source income.

the same time, RHT hit companies operating in sectors with high negative impact on the environment, the tax revenue thus obtained can be considered as compensation for these negative environmental externalities.

When firstly introduced, the RHT corresponded to a surtax equal to 5.5%, that, in addition to the IRES standard rate of 27.5%, effectively restored the 33% tax rate in force up to 2008 (Figure 6). The RHT was applied to certain companies operating on specific energy sectors[32], that exceeded revenues greater than 25 million euros in the previous fiscal year. The 25 million threshold includes all the revenues resulting from the balance sheet, regardless of the source, thus incorporating also the revenues generated from activities that are not related to the energy sector as long as the firm is eligible for the RHT. Producers of electricity from renewable sources (biomass, photo-voltaic and wind) were excluded from the RHT until 2010.

The RHT, since its first introduction, has been modified several times in terms of both tax rates and thresholds (Table 6). After the first two years, in 2010, the surtax was increased by one percentage point. In 2012, it increased by four additional percentage points, thus reaching 10.5%. Finally, in 2014, the surtax decreased to 6.5%. As for the thresholds, the fiscal years from 2008 to 2011, included companies whose revenues exceeded 25 million euros. Then, from 2012 to 2013, there were two simultaneous thresholds: 10 million euros of revenues, and 1 million euros of IRES tax base. Finally, in 2014 these thresholds further decreased: the revenues to 3 million euros and the IRES tax base to 300 thousand euros.

The RHT was applied until the fiscal year 2014, since in 2015 it was considered illegitimate by the Constitutional Court[33].

The RHT foresaw two rules that prevented tax avoidance. First of all, firms that were part of a corporate group had to pay the RHT based on their separate tax base and eligibility threshold. Secondly, in case of mergers (or divisions), RHT eligibility was based on the revenues of the year before the consolidation (or split). The RHT included a "user-friendly" rule, namely companies subject to the surtax

[32]More specifically, companies subject to the RHT are those whose main activity is the research and the production of liquid and gaseous hydrocarbons, the refining of oil, the production or the marketing of oil, lubricating oils and residues, liquefied gas and natural gas and the transport or distribution of natural gas, the production of electricity, its transmission and dispatching, its distribution or its marketing.

[33]The main aspects of unconstitutionality highlighted by the Constitutional Court with the judgment n.10/2015 were the violation of the principle of the ability to contribute and the lack of necessity and urgency.

Table 6: Surtax and thresholds of the RHT.

Law	Fiscal year	Threshold revenue (mil. euros)	Threshold IRES taxable income (mil. euros)	Surtax rate
Law n.81/2008	2008-2009	25	-	5.5%
Law n.99/2009	2010-2011	25	-	6.5%
Law n.138/2011	2012-2013	10	1	10.5%
Law n.69/2013	2014	3	0.3	6.5%

were prohibited from shifting the additional tax burden on consumer prices[34]. The task of monitoring compliance with this rule was entrusted to an independent agency: the Italian Regulatory Authority for Energy, Networks and Environment (ARERA). To convey an idea of the magnitude of this measure, we show the data of the tax bases and the revenues of the RHT during the period of its application (Table 7). In the first year of the RHT application, the surtax revenue amounted to 117 million of euros, while in the following two years, revenues reached about 500 million. With the increase in the surtax rate and the reduction of the thresholds, in 2012 and 2013, the tax base reached 29 billion and the revenue of the RHT increased to 2.7 billion. Finally, with the last reduction of the surtax rate, in 2014 the RHT revenue decreased to 700 million.

Table 7: Taxable income and revenues of the RHT (in thousand euros).

Fiscal year	Surtax taxable income	Surtax revenues
2008	2,137,905	117,585
2009	9,231,158	509,033
2010	8,171,764	531,452
2011	14,275,994	1,498,926
2012	13,213,471	1,387,414
2013	29,401,927	2,734,505
2014	10,717,848	697,952

2.3 Data

We use firm's level data from the AIDA dataset containing information on the balance sheets. Since the "user-friendly" rule did not allow firms to shift the burden

[34]The legislation also regulated that in the case of a merger the reference revenues must be those of the companies incorporated or merged in the period before the merger, in the case of a division the reference revenues refer to the company split before the division.

on consumer prices, we analyse whether firms shifted the tax burden on either capital or labor. We estimate the impact on capital by looking at total assets and one of its specific components, fixed tangible assets (capital). Then, we estimate the impact on labor by looking at labor cost. We also include value added, profit and total cost. Our main variables of interest are total assets and labor. We expect that total assets, and one of its components, capital, can decrease if the tax is beard by capital. On the other side, if the tax burden is shifted onto labor, we expect a decrease in labor cost. In case profits or/and labor cost decreases because of the tax, then the value added will drop as well. Finally, we investigate the total cost as it could be manipulated to decrease the tax base.

Our full dataset covers 15.627 firms operating in the energy sector over the 2007-2016 period. Our main estimations are based on the years from 2007 to 2011. This is due to two reasons. First, from 2012 on, the threshold is determined by both revenues and taxable income, which makes the identification fo the treated group difficult. Second, data on taxable income is not available. From the raw data, we keep firm-year observations satisfying the following criteria: i) firms observed over the whole 2007-2011 period that do not contain missing data; ii) firms that are active for the whole period, implying that we exclude companies that are dissolved, into liquidation or went bankrupt; iii) in the regression discontinuity analysis, extreme values of dependent variables at the top and bottom percentile are excluded from the dataset. Based on the local nature of the regression discontinuity (RD), we also restrict our sample to firms whose revenues are between 5 and 45 million euros, 20 million above and below the threshold of 25 million. We use the revenues in year *t-1* to determine whether a firm in year *t* is subject to the RHT. Our dataset also contains time-invariant information on the firm's economic activity classified with ATECO codes.

Empirical results

2.4 Difference in differences analysis

We run a diff-in-diff analysis from 2007 to 2011, our balanced sub-sample is given by 311 firms for a total of 1,555 observations. We use the firms in the 20 million revenues bandwidth from the threshold of 25 million euros and we exclude those firms that change their treatment status over the treatment period, i.e., firms that occasionally move above and below the threshold over the years from 2008 to 2011.

2.4.1 Identification

Diff-in-diff takes the before-after difference in treated firms' outcomes, which controls for factors that are constant over time in the treatment group. Then, to capture time-varying factors, it takes the before-after difference in the control group of firms, which was exposed to the same set of environmental conditions as the treatment group. Eventually, diff-in-diff subtracts the time-varying factors by taking the difference of these two differences. Our estimating equation is a diff-in-diff model that include an interaction term allowing for different time trends between the treated firms and the control ones:

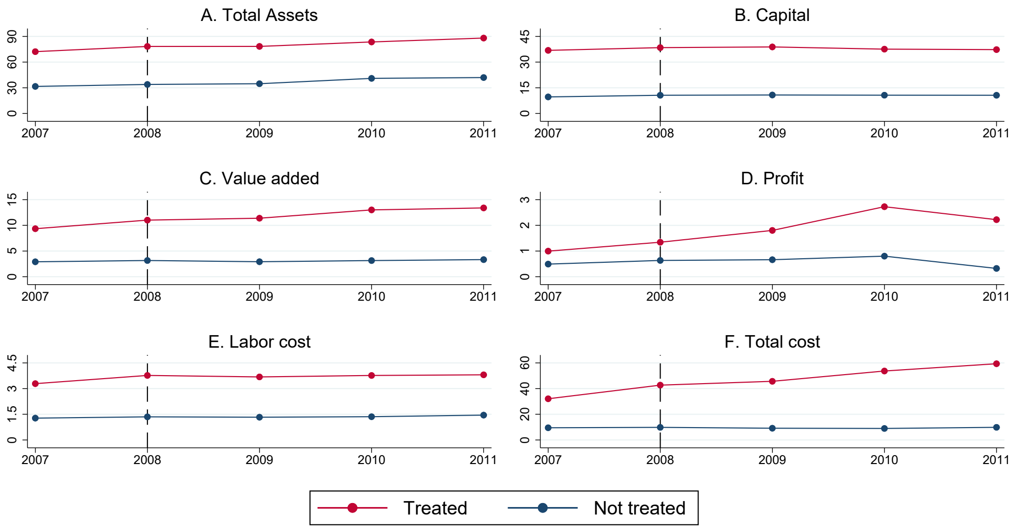
$$y_{i,t} = \alpha + \beta \; Treated_i \cdot Post_t + \gamma_i + \delta_t + \eta Trend_t \cdot Treated_i + u_{i,t}, \tag{5}$$

where $y_{i,t}$ is the outcome variable that can be total assets, capital, value added, profit, labor cost or total cost and varies for firm i and year t; $Treated_i \cdot Post_t$ is our variable of interest and indicates the effect of the "Robin Hood" Tax (RHT); γ_i and δ_t are respectively firms fixed effects and year fixed effects that capture time-specific shocks common to every firm, such as economics cycle, inflation and so on; $Trend_t \cdot Treated_i$ is the interaction between the treatment indicator and the trend; and $u_{i,t}$ is the error term, clustered at firm level. The identification assumption required for the diff-in-diff to be valid is that treatment and control groups follow a parallel trend in the pre-treatment period, i.e., there are no time-varying differences between the two groups. In this chapter, we can not directly test for parallel trend since our dataset only goes back to one year before the treatment, therefore we can not perform the typical parallel trend test. Nonetheless, we check the validity of the diff-in-diff approach by including the group-specific trends that allow for varying pre-treatment trends (de Jong et al., 2011; Green et al., 2014). As a side note, the parallel trend is very likely to hold since we only include firms in specific energy sectors and within the 20 million revenues threshold.

2.4.2 Effect of RHT on firms' outcomes

Figure 7 plots the graphical representation of the relevant outcomes, grouped by treatment. The red and the blue lines are the yearly average of each of the six outcomes of, respectively, the treated and the control groups of firms. The dotted line in 2008 indicates the introduction of the RHT. This initial graphical evidence suggests that the treated group is not affected by the RHT since most of the outcome variables seems to be following the same trend in both the treatment and the control group. It also shows a slight increase in profit and total cost from 2010 on.

Figure 7: Diff-in-diff - Graphical evidence.



Note: Yearly average of each of the six outcomes of the treated and the control groups. The dotted line in 2008 indicates the introduction of the RHT. Values are expressed in million euros.

The diff-in-diff estimation is based on 1,545 observations, that corresponds to 309 firms observed over 5 years (2007-2011). The treated group corresponds to 17% of the sample (54 firms out of 309). Our sample contains approximately 10% of the firms subject to the RHT. In fact, according to the Ministry of Economy and Finance, around 500 firms were treated by the RHT tax in 2008. We split the interpretation of our results in real tax responses (total assets and capital) and corporate tax incidence (labor cost, valude added, profit and total cost). Overall, as shown in Table 8, we do not find any statistically significant result for any of the outcome variables. More specifically we do not find any statistically variation in real tax responses, which we measure with total assets and capital. We also include labor cost, profit, value added and total cost to check whether the tax was shifted onto labor or profits.

Real Tax Responses: Investments If we look at the total asset margin there is no effect of the increase in tax given that the coefficient of the interaction term is not significant. Namely we get a coefficient of -3,056 with a standard error of 2,345. On the other hand, we get a positive but still not significant result if we look at part of the total asset like capital (fixed tangible assets). In the case of capital we get a coefficient of 753.7 with a standard error of 1,071.

Our results do not confirm what the most part of the existing literature finds. In fact, we do not find that firms are less prone to invest as opposed to Hall and Jorgenson (1967); Cummins, Hassett, and Hubbard (1994); Caballero, Engel, and Haltiwanger (1995). However, in line with Lediga et al. (2019) and Devereux et al. (2014) who find that larger firms are less sensitive to tax increases, we find that the absence of investment responses holds for big firms. Finally, our results are consistent with the findings from Yagan (2015) who finds a null impact of increasing the dividend's tax rate. His null result holds both for investments and for employee compensation.

Corporate Tax Incidence: Do Workers Bear the Tax? When we look at labor cost margin we do not find any significant change, namely we get a coefficient of 167.7 and a standard error equal to 777.3. Moreover if we examine the behavior of profits we do not find any evidence of distortion at threshold, in fact the coefficient is -55.10 and its standard error is 144.4. When we look at the value added there is a negative coefficient (-9.269) that is not statistically significant (standard error equal to 698.4). The coefficient of total cost is equal to -2,264 thousand euros with a standard error of 3,238. However, we do not find any statistically significant result and we can conclude that workers do not bear the RHT.

Our results on labor cost are consistent with the findings of Yagan (2015) on workers compensation. Moreover, they are partially consistent with the branch of the literature on the bargaining power. In fact, Arulampalam et al (2012) and Fuest et al (2018) find that the tax is shifted onto wages in case of high workers' bargaining power because of the reduction of the quasi-rent over which shareholders and workers bargain.

Table 8: Diff-in-diff estimation for different outcomes.

	(1) A. Total Assets	(2) B. Capital	(3) C. Value added
$Treated_i \cdot Post_t$	-3,056	753.7	-9.269
	(2,345)	(1,071)	(698.4)
Observations	1,545	1,545	1,545
R-squared	0.94	0.98	0.92
	(4) D. Profit	(5) E. Labor cost	(6) F. Total cost
$Treated_i \cdot Post_t$	167.7	-55.10	-2,264
	(777.3)	(144.4)	(3,238)
Observations	1,545	1,545	1,545
R-squared	0.37	0.96	0.74
Firm FE	YES	YES	YES
Year FE	YES	YES	YES
Treatment-specific trend	YES	YES	YES

Note: Coefficients from diff-in-diff estimates from years 2007 and 2011. Values are expressed in thousand euros. Capital refers to fixed tangible assets. Profit is the the net profit. We regress the outcome variables on the interaction between the treatment and the post variables, the group-specific trends and the country and time fixed effects. Standard errors are clustered at firm level and shown in brackets. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

2.5 Regression Discontinuity Design

After running the diff-in-diff, we replicate the analysis by using a regression discontinuity (RD) design. The reason behind our choice is based on the limited robustness checks that can be performed on our diff-in-diff models due to the short pre-treatment period. Since the treatment group is defined with a threshold, the RD design applies perfectly[35].

2.5.1 Identification

To re-estimate the response of firms to corporate taxes, we use a discontinuity in the corporate tax rate at the 25 million euros revenues threshold of firms operating in the energy sector, as described in Section 2.2. Following the recommendations of Imbens (2008) and Gelman (2019), our main specification uses local linear regressions within a given bandwidth of the treatment threshold, and controls for the running variable (one year lagged revenues) on either side of the threshold. We use a 20 million

[35]See the Introduction for further details.

revenues threshold. Formally, we estimate the following model:

$$
\begin{aligned}
y_{i,t} = \alpha + \beta \cdot 1(T_{i,t-1} \geq C) + \gamma \cdot (T_{i,t-1} - C) \\
+ \delta \cdot (T_{i,t-1} - C) \cdot 1(T_{i,t-1} \geq C) + u_{i,t},
\end{aligned}
\tag{6}
$$

where the outcome variable, $y_{i,t}$, is measured at year t for each firm i. $T_{i,t-1}$ is firm lagged revenues and C is the cutoff defining eligibility for the RHT, so that the dummy variable $1(T_{i,t-1} \geq C)$ defines treatment and control firms[36]. The RD estimator, β, calculates the local average treatment effect (LATE) of eligibility for the RHT on firm-level outcomes. In some specifications, we will add industry and year fixed effects. These controls are not necessary for identification, but they might improve the efficiency of the estimation by reducing the sampling variability. $u_{i,t}$ is the error term. Standard errors are clustered at firm-level. The sample used in this regression is balanced. To avoid that our results are driven by extreme values we drop the top and bottom percentile of each outcome variable. In the case of the regression discontinuity, we do not drop the firms that move across the threshold over the time period analysed. In fact, in the RD design we exploit the cross section variation.

2.5.2 Validity of RD design

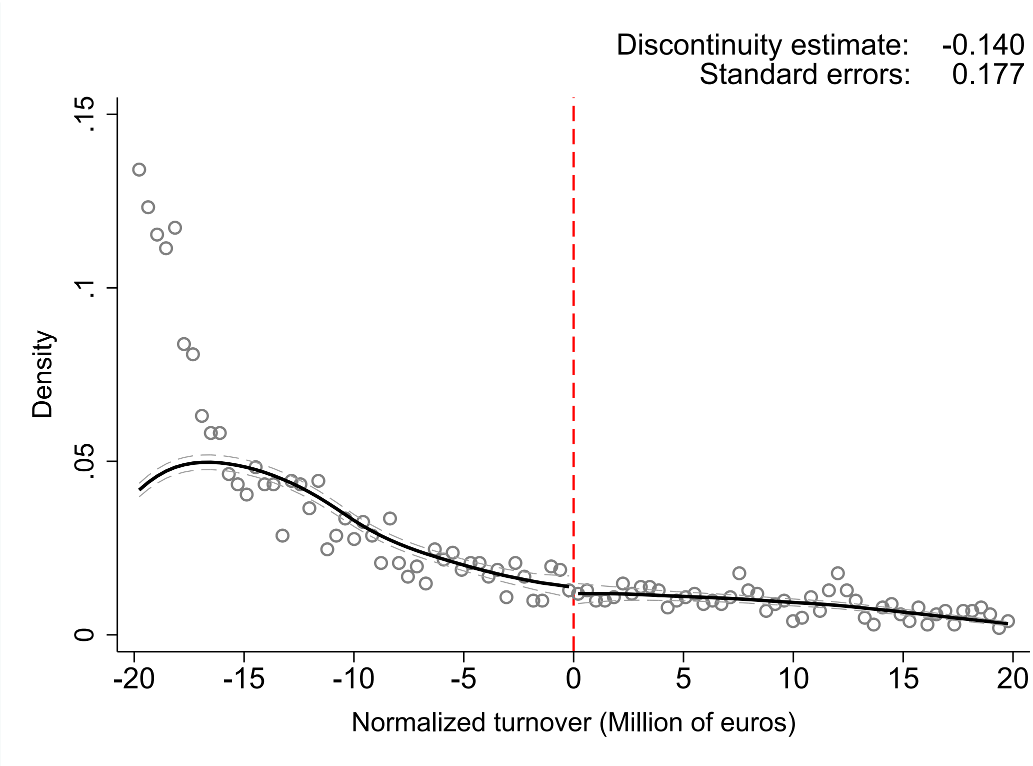
The key identifying assumption underlying our research design is that firms did not game the law by manipulating their revenues. We provide clear evidence that the revenue density is smoothly distributed around the 25 million threshold, as we would expect in a valid RD design. This is germane to the argument that it is harder to evade a broader base by concealing revenues (see, e.g., Best et al. 2015 JPE). We further show that key pre-determined firm-level outcomes look highly similar for firms below and above the threshold.

Strategic Manipulation of Revenues: McCrary test. The key assumption to identify β is that firms did not game the law by reporting revenues just below the threshold determining eligibility for the "Robin Hood" tax. In practice, firms have limited control over revenues that depends on market conditions which we assume not to be manipulated by firms subject to taxation. Moreover, the eligibility for the RHT

[36]We use the linear specification without additional controls as our baseline. This choice is consistent with the argument of Gelman and Imbens (2014) that using higher order polynomials in regression discontinuity settings can lead to high sensitivity.

is based on the revenues of the previous fiscal year (lagged revenues). We scrutiny the validity of this assumption in Figure 8, which depicts the density of firms around the revenues cutoff for RHT eligibility. As we would expect in a valid RD design (Lee 2008; McCrary 2008), the figure depicts a smooth distribution around the cutoff. Such result of no response over the revenues margin constitutes prima facie evidence that firms did not respond to the introduction of the "Robin Hood" tax. This evidence is supported by the fact that in case of division of big firms the revenue determining the RHT eligibility is based on the revenue of the year before the splitting. Therefore there is no incentive to split to manipulate the revenue in order to avoid the surtax.

Figure 8: Frequency of Firms and McCrary test.



Pre-Determined Characteristics: Balance Test. To check the validity of our RD design, we examine the patterns of a range of pre-determined characteristics around the threshold. In Table 9, we check for manipulative sorting by performing balance tests on the available invariant firms characteristics. If there is non-random sorting, we expect some of these characteristics to differ systematically between treated and untreated firms around 25 million revenues threshold. The available pretreatment characteristics are the geographical location (North-East, North-West, Center, South and Islands), the firm's age, and the legal form (limited company, public limited company and others). We run the balance test using one year, namely 2008, since the pre-determined characteristics are invariant over the years. No pre-treatment characteristics show a significant discontinuity at the threshold, with the exception of Center where the coefficient of interest is slightly statistically significant. This result might be guided by the city of Rome. Nonetheless, we are not concerned by the presence of this significant coefficient since the other geographical locations are perfectly balanced.

Table 9: Balance Test.

	(1) Age	(2) North-East	(3) North-West	(4) Center
$1(T_{i,t-1} \geq C)$	4.179	-0.167	0.008	0.138*
	(5.270)	(0.103)	(0.123)	(0.078)
Obs.	400	409	409	409
	(5) South	(6) Islands	(7) LTD	(8) PLC
$1(T_{i,t-1} \geq C)$	-0.001	0.022	-0.012	-0.020
	(0.052)	(0.020)	(0.071)	(0.070)
Obs.	409	409	409	409

Note: Values refer to 2007. Optimal bandwidth is fixed at 20 million euros of revenues.

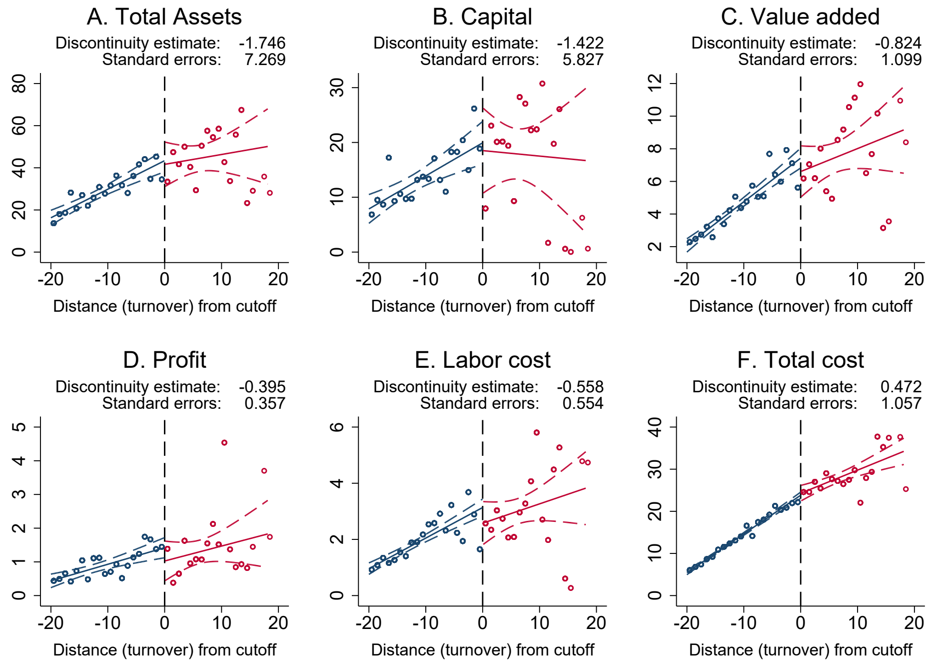
2.5.3 Effect of RHT on firms' outcomes

Similarly to the diff-in-diff estimation, we investigate total assets, capital, value added, profit, labor cost and total cost. First graphical evidence shows that there is no discontinuity in all the relevant outcomes. Our results from the diff-in-diff estimation are entirely confirmed by the graphical evidence of the regression discontinuity.

The regression discontinuity estimation of is based on 876 observations, that corresponds to 219 firms observed over 4 years (2008-2011). The treated group constitutes 17% of the sample (148 out of 876). Table 10 presents the RD estimates of the impact of the "Robin Hood" tax on our main outcome variables. For each variable, we present results using a bandwidth equal to 20 million euros. In Column 1, we show the estimated β coefficient obtained from equation (6) and robust standard errors clustered at firm level. In column 2 we add industry fixed effects and in column 3 we add both industry and year fixed effects.

Real Tax Responses: Investments If we look at the total asset margin there is no effect of the increase in tax given that the coefficient of the dummy variable equal

Figure 9: The impact of the "Robin Hood" tax on firm-level outcomes.



Note: Total assets, capital, value added, labor cost, profit and labor share are expressed in million euros. Bandwidth is fixed at 20 million euros and expressed in million euros. Capital refers to fixed tangible assets. Labor share is the ratio between labor cost and value added. Profit is the the net profit. For each outcome variable, values from the top and bottom percentile are dropped.

to one when the revenue is greater than 25 millions is not significant. Namely we get a coefficient of -1,745 with a standard error of 7,269. We also get a negative but not significant result if we look at part of the total asset like capital (fixed tangible assets). In the case of capital we get a coefficient of -1,421 with a standard error of 5,827. The finding of no significant increase in investment is robust to the addition of industry and year fixed effects, and alternative bandwidth choices (see Figure A1 in Appendix B).

Corporate Tax Incidence: Do Workers Bear the Tax? When we look at labor cost margin we do not find any significant drop at the threshold, namely we get a coefficient of -557.9 and a standard error equal to 554.2. Moreover if we examine the behavior of profits we do not find any evidence of distortion at threshold, in fact the coefficient is -395.2 and its standard error of 356.7. When we look at the value added there is a negative coefficient (-823.5) that is not statistically significant (standard error equal to 1,098). The coefficient of total cost is equal to 471.7 thousand euros with a standard error of 1,057. The finding of no significant increase in corporate tax incidence is robust to the addition of industry and year fixed effects, and alternative bandwidth choices (see Figure A1 in Appendix B) [37].

Table 10: Regression discontinuity estimation for different outcomes.

	(1)	(2)	(3)
		A. Total assets	
$1(T_{i,t-1} \geq C)$	-1,745	-5,549	-5,408
	(7,269)	(6,182)	(6,193)
		B. Capital	
$1(T_{i,t-1} \geq C)$	-1,421	-3,771	-3,702
	(5,827)	(5,141)	(5,150)
		C. Value added	
$1(T_{i,t-1} \geq C)$	-823.5	-1,026	-1,012
	(1,098)	(1,031)	(1,037
		D. Profit	
$1(T_{i,t-1} \geq C)$	-395.2	-467.7	-475.4
	(356.7)	(371.1)	(374.5)
		E. Labor cost	
$1(T_{i,t-1} \geq C)$	-557.9	-575.9	-569.2
	(554.2)	(426.5)	(426.9)
		F. Total Cost	
$1(T_{i,t-1} \geq C)$	471.7	447.6	590.0
	(1,057)	(998.5)	(971.7)
Observations	876	876	876
Industry FE	No	Yes	Yes
Year FE	No	No	Yes

Note: Coefficients from linear regression discontinuity estimates. Optimal bandwidth is fixed at 20 million euros of revenues. Values are expressed in million euros. Capital refers to fixed tangible assets. Labor share is the ratio between labor cost and value added. Profit is the the net profit. Industry fixed effects refers to ATECO codes. Errors clustered at firm level. Standard errors in brackets. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

2.6 Robustness checks

We test whether our results are sensitive to bandwidth choice. Figure A1 in Appendix B reports coefficient estimates and 95 percent confidence intervals obtained by estimating equation 6 on different bandwidths around the threshold. The figure shows that our coefficient estimate is qualitative similar when using a reasonable range of bandwidths. In particular all the coefficients are never statistically significant for bandwidth going from 3000 to 20000 and the other margins are never significant.

[37] With a negative and statistically significant coefficient for profit for bandwidth values between 10 and 15 million euros.

2.7 Conclusions

This chapter analyses the impact of an increase in corporate tax on big firms behavior. We focus on the Italian "Robin Hood" Tax: a surtax on firms of the energy sector with more than 25 million revenues. Combining different econometric models with firm-level balance sheets, our contribution finds a null effect on both investments and labor costs. Throughout the article, we focus on different outcome variables that can be grouped into two areas: real tax responses and corporate tax incidence. We do not find any effect for all the relevant outcomes. Our results are validated by the use of two different approaches: we first perform a diff-in-diff and then we validate our results with a regression discontinuity. Graphical representation of the regression discontinuity provides compelling evidence of no effect for any of the outcome variables. The estimated coefficients of the variables of interest confirm the graphical evidence. We add some robustness checks that confirm our findings.

Notice that our results are relative to big firms and for a temporary increase in tax. Our results are in line with Fuest et al. (2018) who show that while medium and small firms react to a tax hike by decreasing the wage, big firms response is less sensitive. However, the increase in tax either in Arulampalam et al. (2012) and Fuest et al. (2018) is not relative to a specific event but it exploits changes in tax rates over a given period of time. On the contrary the RHT is an unexpected shock introduced to face a period of economic crisis. Our contribution to the literature is twofold. Firstly, in line with the literature that finds that large firms are less sensitive to tax hikes than smaller firms, we find that very big firms, namely those whose revenues are above 25 million euros, are not sensitive to a tax increase. Secondly, to the best of our knowledge, the literature has never analysed firms' response to the introduction of a temporary tax hike. With our contribution, we analyse the very big firms' tax response to temporary tax hikes.

Our results are applicable to tax extra-profits in period of crises like that of COVID-19 pandemic and even if normative attention should be paid in avoiding the negative impact on consumers' prices, there is evidence that big firms do not shift the additional tax neither on investments or on labor.

References

(1) Auerbach, A. J., Hassett, K. (1992). "Tax policy and business fixed investment in the United States", *Journal of Public Economics*, Elsevier, vol. 47(2), pages 141-170, March.

(2) Barnett, S. A., Sakellaris, P. (1998). "Nonlinear response of firm investment to Q: Testing a model of convex and non-convex adjustment costs", *Journal of Monetary Economics*, Elsevier, vol. 42(2), pages 261-288, July.

(3) Boeri, T., (2015). "Perverse Effects of Two-Tier Wage Bargaining Structures", *IZA World of Labor*, ISSN 2054–9571, No. 101. http://dx.doi.org/10.15185/izawol.101.

(4) Bond, S. R., Cummins, J. G. (2000). "The Stock Market and Investment in the New Economy: Some Tangible Facts and Intangible Fictions", *Brookings Papers on Economic Activity*, 31, issue 1, p. 61-124.

(5) Caballero, R., Engel, E., Haltiwanger J. (1995). "Plant-Level Adjustment and Aggregate Investment Dynamics", *Brookings Papers on Economic Activity*, 26, issue 2, p. 1-54.

(6) Cummins, J. G., Hassett, K. A., Hubbard, R. G. (1994). "A Reconsideration of Investment Behavior Using Tax Reforms as Natural Experiments", *Brookings Papers on Economic Activity*, Economic Studies Program, The Brookings Institution, vol. 25(2), pages 1-74.

(7) Cummins, J. G., Hassett, K. A., Oliner, S. D. (2006). "Investment Behavior, Observable Expectations, and Internal Funds", *American Economic Review*, 96 (3): 796-810.

(8) de Jong, P., van der Klaauw, B., Lindeboom, M. (2011). "Screening disability insurance applications". *Journal of the European Economic Association*, 9(1), 106–129. http://www.jstor.org/stable/25836059.

(9) Desai, M. A., Goolsbee, D. A. (2004). "Investment, overhang, and tax policy", *Brookings Papers on Economic Activity*, no. 2 (2004): 285–338.

(10) Devereux, M. P., Li, L., Loretz S. (2014). "The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records", *American Economic Journal: Economic Policy*, 6 (2): 19-53.

(11) Erickson, T., Whited, T. M. (2000). "Measurement Error and the Relationship between Investment and Q", *Journal of Political Economy*, 108. 1027-1057.

(12) Feldstein, M. S. (1970). "Corporate Taxation and Dividend Behaviour", *Review of Economic Studies*, Oxford University Press, vol. 37(1), pages 57-72.

(13) Feldstein, M. S. (1982). "Government deficits and aggregate demand", *Journal of Monetary Economics*, Elsevier, vol. 9(1), pages 1-20.

(14) Fuest, C., Peichl, A., Siegloch A. (2018). "Do Higher Corporate Taxes Reduce Wages? Micro Evidence from Germany". *American Economic Review*, 108 (2): 393-418.

(15) Gelman, A., Imbens, G. (2019). "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs", *Journal of Business & Economic Statistics*, 37, issue 3, p. 447-456.

(16) Green, C.P., Heywood, J.S., Navarro M. (2014). "Did liberalising bar hours decrease traffic accidents?", *Journal of Health Economics*, Elsevier, vol. 35(C), pages 189-198.

(17) Hall, R., Jorgenson, D. (1967). "Tax Policy and Investment", *American Economic Review*, 57.

(18) Harberger, A. C. (1962). "The Incidence of the Corporation Income Tax", *Journal of Political Economy*, University of Chicago Press, vol. 70, pages 215-215.

(1) Harberger, Arnold C. (1966). "Efficiency Effects of Taxes on Income from Capital". In *Effects of Corporation Income Tax*, edited by Marian Krzyzaniak, 107–17. Detroit: Wayne State University Press.

(19) Hubbard, R. G., Mankiw, N. G., Taylor, J. B., Hassett, K. A. (2012). "The Romney Program for Economic Recovery, Growth, and Jobs". http://www.docstop.com/docs/125714335/Romney-Tax-Reform-White-Paper (accessed December 19, 2012).

(20) Kangur, A., (2018). "Competitiveness and Wage Bargaining Reform in Italy", *IMF*, Working Papers 18/61.

(21) Imbens, G., Lemieux, T. (2008). "Regression discontinuity designs: A guide to practice", *Journal of Econometrics*, volume 142, issue 2, pages 615-635.

(22) Lediga, C., Riedel, N., Strohmaier,K. (2019). "The elasticity of corporate taxable income—Evidence from South Africa", *Economics Letters*, Elsevier, vol. 175(C), pages 43-46.

(23) Lee D. S. (2008). "Randomized experiments from non-random selection in U.S. House elections", *Journal of Econometrics*, volume 142, issue 2, pages 675-697.

(24) McCrary, J. (2008). "Manipulation of the running variable in the regression discontinuity design: A density test", *Journal of Econometrics*, volume 142, issue 2, pages 698-714.

(25) Pedersini, R., (2019). "Italy: institutionalization and resilience in a changing economic and political environment", in *Collective bargaining in Europe: towards an endgame*, edited by Torsten Müller, T., Vandaele, K., and Waddington, J.

(26) Poterba, J. M., Rotemberg, J. J., Summers, L. H. (1985). "A Tax-Based Test for Nominal Rigidities", *Working papers* 376, Massachusetts Institute of Technology (MIT), Department of Economics.

(27) Ryan, P. (2012). "The Path to Prosperity", (accessed September 20, 2015).

(28) Yagan, D. (2015). "Capital Tax Reform and the Real Economy: The Effects of the 2003 Dividend Tax Cut: Dataset", *American Economic Review*, 105 (12): 3531-63.

(29) Zwick, E., Mahon, J. (2014) "Do Financial Frictions Amplify Fiscal Policy? Evidence from Business Investments Stimulus". *Working Papers* 1415, Oxford University Centre for Business Taxation.

B Appendix

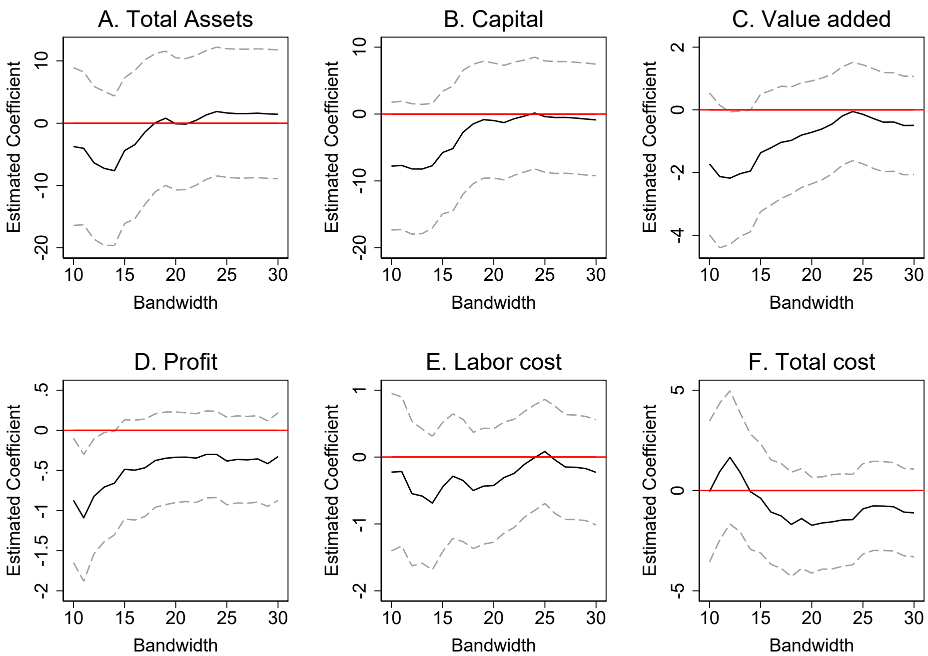
Table A7: Summary statistics: diff-in-diff.

Variable	Obs	Mean	Std. Deviation	Min	Max
A. Total assets	1,545	44,394	205,485	76.11	4,459,167
B. Capital	1,545	15,305	36,649	-10.48	367,032
C. Value added	1,545	4,597	6,546	-11,493	5,3061
D. Profit	1,545	808.2	4,869	-57,043	102,690
E. Labor cost	1,545	1,759	2,826	0	24,955
F. Total cost	1,545	15,966	23,556	2.138	497,527

Table A8: Summary statistics: regression discontinuity.

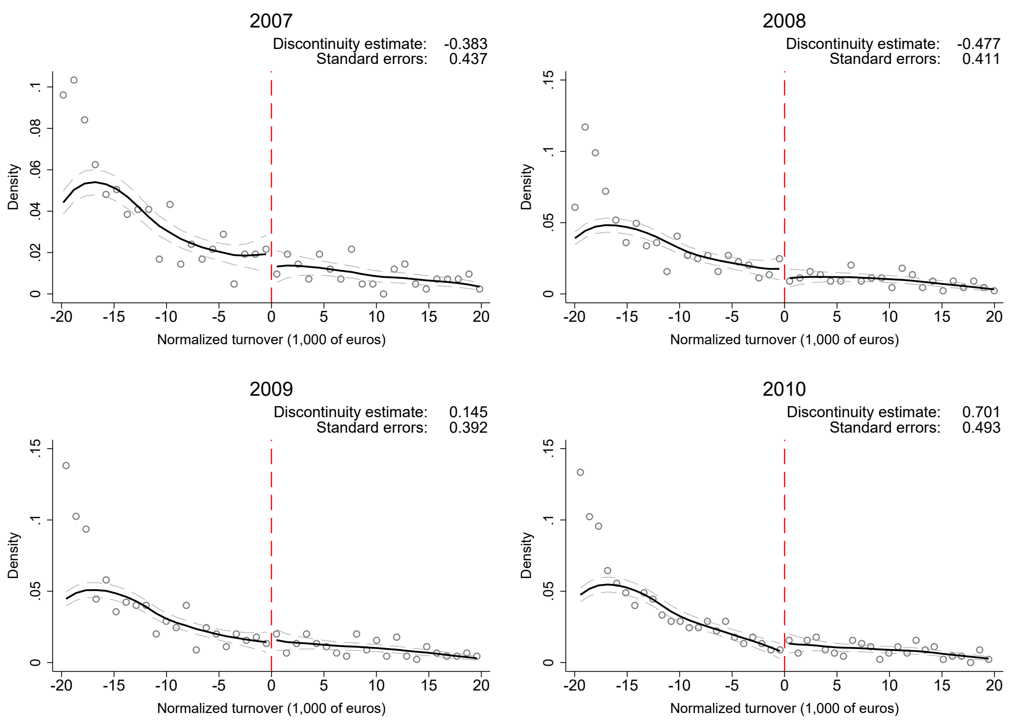
Variable	Obs	Mean	Std. Deviation	Min	Max
A. Total assets	876	29,426	32,132	2,608	178,835
B. Capital	876	13,293	22,934	0.132	125,005
C. Value added	876	4,641	4,259	-96.45	22,054
D. Profit	876	886.5	1,757	-5,497	10,596
E. Labor cost	876	1,976	2,024	0.020	13,619
F. Total cost	876	14,879	8,729	3,598	41,170

Figure A1: Sensitivity of estimates to varying bandwidth.



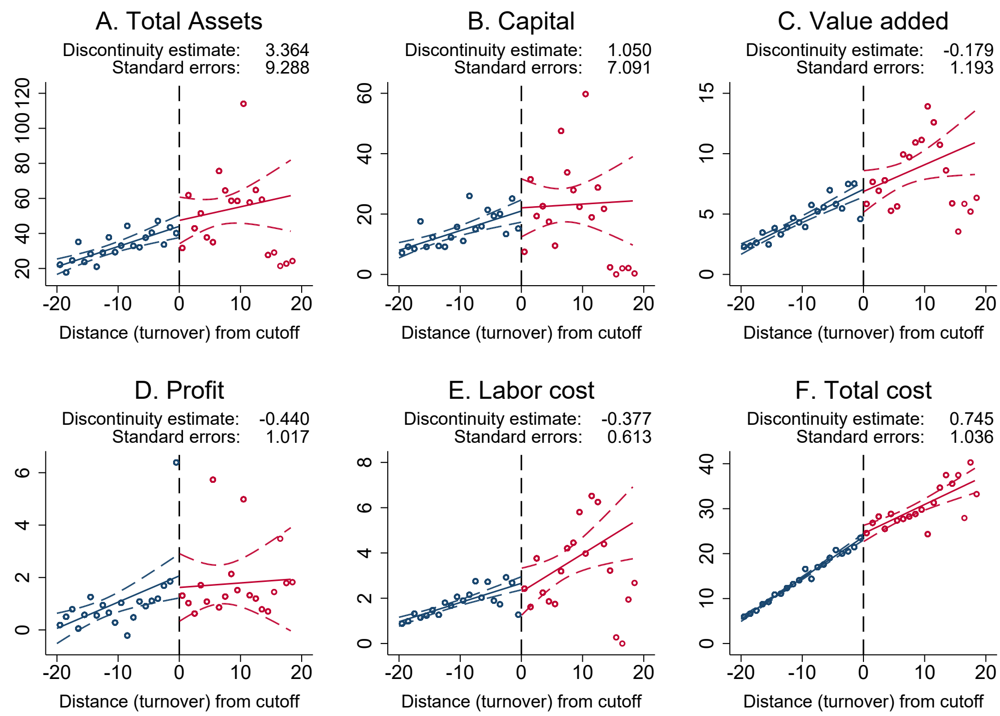
Note: This figure reports RD estimate and 90 percent confidence intervals obtained from estimating 6 on a various range of bandwidths.

Figure A2: Frequency of Firms and McCrary Test: Year by Year analysis.



Note: year by year McCrary test from 2007 to 2010.

Figure A3: The impact of the "Robin Hood" tax on firm-level outcomes. Top and bottom percentile observations are kept.



Note: Total assets, capital, value added, labor cost, profit and labor share are expressed in million euros. Bandwidth is estimated with inference procedures developed in Calonico, Cattaneo and Farrell (2020) using labor cost as dependent variable. Bandwidth is fixed at 20 million euros and expressed in million euros. Capital refers to fixed tangible assets. Labor share is the ratio between labor cost and value added. Profit is the the net profit. For each outcome variable, values from the top and bottom percentile are kept.

Chapter three

3 Assessing Bidding Zone Configuration: evidence from the Nord Pool market

3.1 Introduction

Given the ongoing spikes in gas and electricity prices, in this third chapter, we investigate one of the determinants of the energy market prices: the market configuration. Configurations of bidding zones in electricity markets are fundamental to give accurate signals while maintaining security of supply, especially given the increasing share of renewable energy sources (RES) in the market. Different configurations do not modify the actual network transmission capacity, but they affect price formation mechanisms and import/export flows.

The debate on the optimal bidding zone configuration is polarized; on one hand, the nodal market and smaller bidding zones are claimed to facilitate congestion management thanks to prices reflecting scarcity, on the other hand, larger zones should favour liquidity and avoid price spreads. Nevertheless, the ongoing debate on the configuration of the optimal bidding zones is mainly based on theoretical models. The goal of this chapter is to empirically assess the effects of the different bidding zones[38].

In particular, we focus on the market re-configuration in Sweden, where the Transmission System Operator (TSO) was accused to abuse its dominant position (Art. 102 Treaty on the Functioning of the European Union, from now on TFEU) by curtailing Available Transmission Capacity[39] (ATC) with the neighbouring countries. On April 2010, the European Commission announced that from November 2011, the single Swedish bidding zone would have been split into four smaller zones.

In fact, because of the Swedish network bottlenecks due to the localisation of inelastic demand in the South and cheap hydroelectric supply in the North, the Swedish TSO, was accused to curtail Available Transmission Capacity (ATC) with neighbouring countries in order to manage internal congestion. Although Art. 18 and 35 of TFEU expressly prohibit discrimination based on nationality and quantitative re-

[38]We provide the definition of a bidding zone and we explain the difference between the nodal and the zonal market in Section C.1 in Appendix C

[39]The Available Transmission Capacity is the maximum amount of electricity that can be traded in the day-ahead market through a specific inter-connector between two areas. National TSO determines ATCs hourly for each direction and each inter-connector.

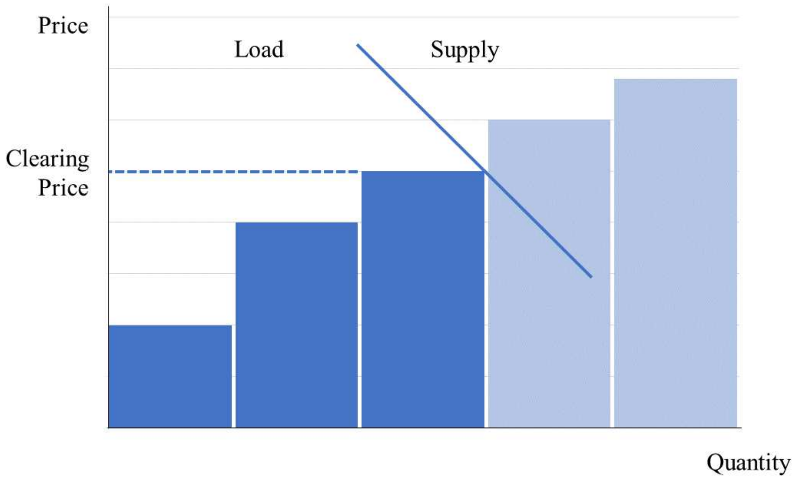
strictions on exports. As a consequence, during congestion hours, the average prices in Sweden were claimed to be significantly low and the TSO accused to curtail the ATCs. The market split was introduced to have price signals to reflect actual internal bottlenecks. Using empirical data, we consider that the re-configuration of the Swedish market should translate into changes in the average price in Sweden and to differences in cross-zonal and cross-country flows thanks to a more accurate use of the inter-connectors and congested lines.

In this chapter we perform both a difference-in-differences to estimate the impact of the policy intervention on prices and a regression discontinuity in time to estimate the impact on the policy on cross-zonal and cross-country flows.

The prices we analyse are spot prices, that are set in the electricity wholesale markets. In these markets, buyers and sellers simultaneously send their bids with the amount of energy they are willing to buy and sell and the relative price. Once that bids are set, they are aggregated by the power market using the merit-order criterion where bids are ranked in ascending order of price[40]. The market clearing price is then calculated at the point of intersection of buyers' and sellers' curves as in Figure 10. The difference between the price buyers were willing to pay and the clearing price is the consumer surplus, vice versa for the producer surplus. This mechanism is called "system marginal price" and it is the mechanism behind the formation of the electricity prices we use in this chapter. This mechanism takes place in each bidding zone and it is sensitive to the bidding zones configuration. In fact, as theoretically shown by Bjørndal and Jörnsten (2001) in case of a single nationwide zone, average prices in the day-ahead market will be lower compared to multiple smaller zones. In case of internal congestion, the multiple zonal prices will incorporate the congestions by taking them into account in the price formation. Conversely, the zonal pricing ignores internal congestions that are solved by the TSO either with costly remedial actions or by curtailing Available Transmission Capacities. We repeat this exercise by empirically showing how the average price in Sweden changed after the re-configuration. Our empirical analysis can be considered as a comparison between the nodal (four bidding zones) and the zonal market configuration (one nationwide bidding zone). The existing literature has widely covered the comparison between the nodal and the zonal systems, however, it is mainly based on theoretical models

[40]Since RES are considered to have zero marginal cost, they enter at the bottom of the merit order. In fact, RES are characterized by very low operating costs but high capital expenditures.

Figure 10: Uniform auction under perfect competition.



Note: Authors' elaboration based on Cretì and Fontini (2019).

that are sometimes combined with anecdotal evidence. Our contribution confirms the theoretical results from Sarfati, Hesamzadeh, & Holmberg (2019) that prove that the zonal market leads to distorted signals of transmission constraints. Indeed, we analyse to what extent the cross-zonal and cross-country flows changed thanks to non-distorted price signals that were better reflecting scarcity.

Throughout our empirical analysis we model electricity prices taking into account the regional seasonality. In particular, we use monthly and day of the week dummies, as in Hadsell, Shawky, and Marathe (2004) and Huisman & Kiliç (2013). Finally, as we use the temperature as an exogenous instrument for prices, we want to highlight that the strong relationship between prices and temperatures has already been observed by Huisman (2008). He uses average prices in peak hours (from 8 am to 8 pm) from the Dutch APX market, from January 2003 throughout February 2008, to show that temperatures are a key predictor for prices.

The different designs upon which our econometric analysis rely on show that the market efficiency improved after the market re-configuration. Our measure of market efficiency relies on accurate price signals and the exploitation of cross-country available transmission capacities. As theoretically shown by Sarfati, Hesamzadeh, & Holmberg (2019), we confirm that the zonal market gives distorted price signals on transmission constraints. Firstly, with our diff-in-diff analysis we find that prices in

Sweden increased, the estimated magnitude of the effect varies from 0.1% to 4% of the average price. Secondly, with our regression discontinuity in time we find that cross-zonal flows from the Northern to the Southern part of Sweden decreased by approximately 11%. The price differences resulting from the splitting pushed producers in the South to enter the market and therefore the South could decrease the import from the Northern area. Thirdly, we find that cross-country opportunities increased, namely, the net flow from Sweden to Finland increased from 5 to 14 percent of the average flow.

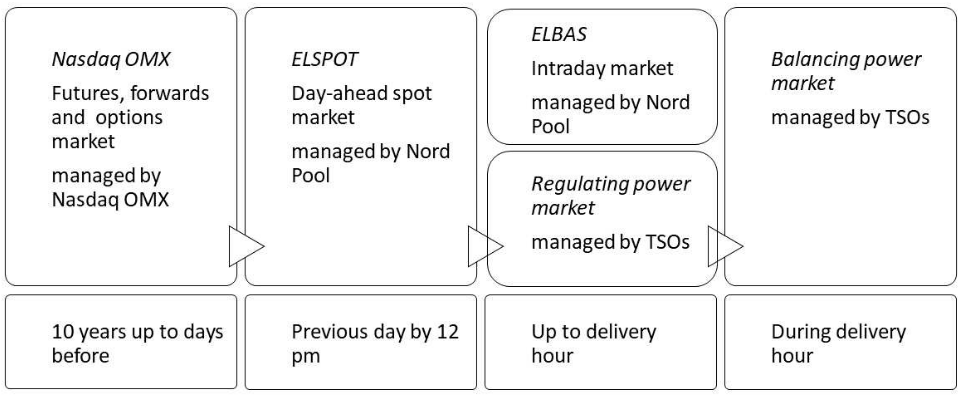
3.2 Description of the Swedish market

The Swedish market is an integrated part of the Nordic electricity market. The Nordic market is divided into bidding zones, and in each zone demand and supply curves meet to form the zonal price. When transmission capacities between bidding areas are not sufficient, congestion will lead to price divergence between areas; while when transmission capacities between bidding areas are within the capacity limits set by the national TSOs, congestion will lead to price convergence between areas. The zones configuration should be created to handle congestions in the grid.

Both Norway and Sweden underwent bidding zones re-configurations over the years 2008-2011. Norway's adjustment was driven by physical bottlenecks and the change was gradually implemented by the Norwegian TSO, Statnett, without a decision of the European Commission. The Swedish case, however, resulted from the Commission assessment that the national TSO, Svenska Kraftnät, might have abused of its dominant position (Art. 102 TFEU) by discriminating between domestic and foreign (namely, Danish) network users. The splitting in multiple bidding zones was announced on April 2010 and implemented on November 2011, in between the TSO had to avoid further curtailments on the Danish border. The market re-configuration is our treatment variable and it can be considered as an exogenous variation as it was imposed by the European Commission.

The Nordic electricity market consists of four markets (Figure 11). Firstly, the OMX Nordic Exchange is an equities and derivatives market where Nordic power products such as futures and options are traded from 10 years up to days before deliver. Secondly, the ELSPOT market (which is our main focus) is a day-ahead hourly two-sided auction market managed by the common power exchange Nord Pool where the wholesale electricity trading takes place. Participants can place offers for each hour of the next day until gate closure, 12 hours before the market opens (12pm), meaning that hourly bids for the following day generate in parallel. Thirdly, the ELBAS continuous market operates after the day-ahead market with products that can be traded up to the delivery hour. The ELBAS market, also managed by Nord Pool, works simultaneously with the Regulating Power Market, it counteracts imbalances related to the planned day-ahead market and is managed by the Transmission System Operators (TSOs). Lastly, the balancing market, also managed by the TSOs, takes place during the delivery hour. Either the generators or the buyers can increase/decreases their output/consumption, if needed for system balance.

Figure 11: Market Design in the Nordic Area.

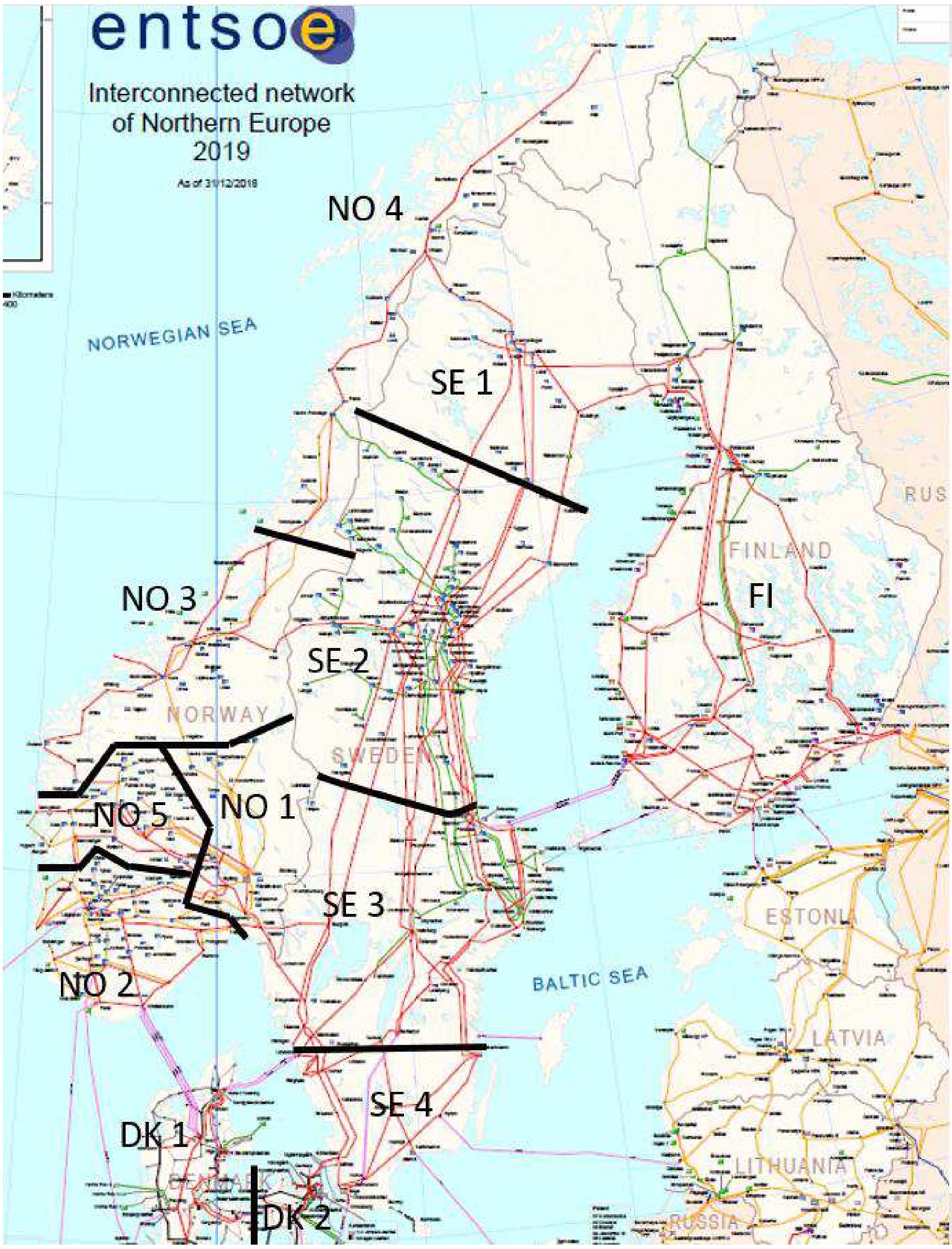


Note: Authors' elaboration based on TemaNord 2016:540, Nordic Council of Ministers.

The Nord Pool power market calculates one aggregate price based on four countries (Denmark, Finland, Norway, and Sweden) and twelve corresponding bidding zones[41] (Figure 12). This price, called the System price assumes away transmission constraints (capacities are set to infinity) and is determined as the unconstrained market clearing price of the above mentioned countries. Due to its calculation, the prices arising in each bidding zone differ from the system price because of congestions that are either country- or bidding zone- specific.

[41] On November 2011, after the re-configuration, the twelve zones are divided as follows: two zones for Denmark, one for Finland, five for Norway and four for Sweden. In these twelve zones, hydro-power contributes to about 50% of the total power generation, although countries have different generation mixes. Denmark takes its capacity evenly from thermal generation and wind power; Finland is based on conventional sources such as thermal and nuclear generation; Sweden is fuelled by hydro-power in the North and nuclear and thermal production in the South, even if it has been going through a process of nuclear decommissioning in the last decade; finally, Norway is almost entirely based on hydro-power, allowing for flexible and cheap energy.

Figure 12: Bidding zone configuration as of Nov. 1, 2011.



Source: Nord Pool AS

3.3 Data

We mainly collect data from the Nord Pool's Power Market Data and from the Svenska kraftnät[42] system data. We use data on electricity prices, consumption, production, import, export, temperatures and carbon dioxide prices. We focus on Sweden and its bidding zones. Our dataset includes hourly observations for 14 years, from 2005 to 2019. Data on prices, production and consumption is collected from the ELSPOT day-ahead market: it includes hourly prices and traded buy and sell volumes [43]. We integrate the ELSPOT dataset with data from Svenska kraftnät where consumption and production per bidding zone is recorded also before the market splitting[44]. We collect data on import and export with neighbouring countries from a dedicated section in the Nord Pool's Power Market Data. By combining production and consumption with import and export, we can calculate data on the flows between Swedish bidding zones. We build the daily price variable by averaging 24-hour prices and by summing up the hourly volumes per day. The country-level price in the post-treatment period is obtained as an average weighted by buy volume per bidding zone, as in (as in Equation A1 in Appendix C.2[45]).

In our analysis, we use two empirical applications that we apply at different levels: firstly, we use a diff-in-diff approach to compare the Swedish price to the system price; secondly, we use a regression discontinuity in time (RDiT) to measure the effects on the flows within the Swedish bidding zones and between each zone and its neighbouring countries. Hence, we work on two different dimensions: national level prices (diff-in-diff approach) and bidding zones level production and consumption volumes (RDiT approach).

3.3.1 Price

In Table 11 we report summary statistics before and after the intervention for the relevant outcome variables of the diff-in-diff analysis. Over our period of interest, prices both from Sweden and from the System show a decreasing trend: in the post-intervention period prices were, on average, 20% lower. The standard deviation is

[42]The Swedish Transmission System Operator
[43]Buy volumes are different from sell volumes because of trade between zones.
[44]The ELSPOT dataset only includes data on existing bidding zones.
[45]We use buy volumes as we take a consumer's perspective.

also similar in terms of magnitude with slightly smaller values for the System price[46]. Figure 13 shows the distribution of daily price in Sweden compared to the System

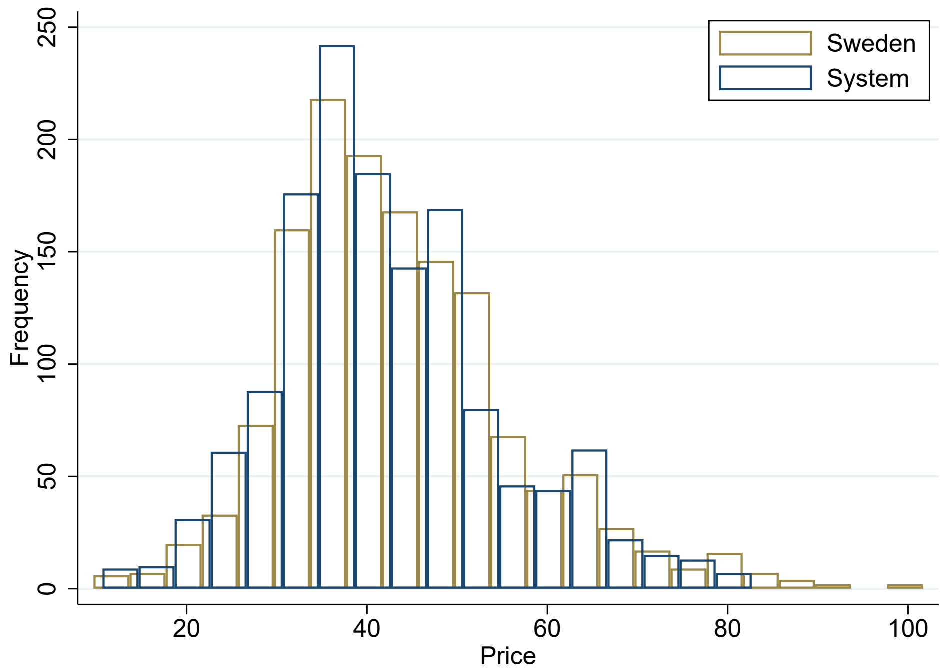
Table 11: Summary statistics of price by 12 months windows.

		Sweden		System	
Time window		Mean	Std. Dev.	Mean	Std. Dev.
2 windows before	357	47.91	11.43	46.99	8.48
1 window before	349	53.01	15.30	51.89	15.25
1 window after	353	33.21	10.89	31.58	9.99
2 windows after	365	40.08	7.12	38.78	6.19

Note: Summary statistics of average daily price by 12 months windows for System and Sweden prices. The windows start on November 1st and end on October 31st, e.g., "1 window after" the treatment is the average price from the first day after the intervention (November 1st, 2011) to 366 days afterwards (October 31st, 2012). Values in the top and in the bottom percentile are dropped.

values. The two sets of prices have a similar distribution, with the system price being slightly more right skewed.

Figure 13: Distribution of Swedish and System price.



Note: This figure shows the distribution of price in Sweden compared to the System. Price values in the top and in the bottom percentile are dropped.

[46] Summary statistics is replicated by including outliers, in Appendix C.3. Swedish higher values of the standard deviation are due to maintenance of nuclear power plants in 2010 that caused very high spikes in prices.

Electricity prices and volumes have periodic fluctuations since they typically have hourly, weekly and yearly seasonality. Hourly seasonality is not considered here because we use average daily prices. In the RDiT we account for day of the week, monthly and yearly seasonality with seasonal dummies. We show the day of the week and monthly seasonality by including dummy variables in an OLS regression with robust standard errors. We estimate the following two equations:

$$Price_t = \alpha + \beta_1 Monday_t + \beta_2 Tuesday_t + ... + \beta_6 Saturday_t + u_t \tag{7}$$

$$Price_t = \alpha + \beta_1 February_t + \beta_2 March_t + ... + \beta_{11} December_t + u_t \tag{8}$$

By estimating Equation 7 we find out that the day of the week seasonality has a clear and consistent pattern with higher prices during the week days that start decreasing on Fridays and reach their lowest point on Sundays. Table 12 shows the estimated coefficients: Sundays and Saturdays are not statistically different from each other and are, on average, much lower compared to the week days. Prices during the weekdays are at least 5€/MWh higher than the weekends' ones. By estimating Equation 8 we find that monthly seasonality shows a clear inverse relationship between cold temperatures and prices: the lower the average temperatures, the higher the prices [47]. Table 13 shows that estimated coefficients for each month: the highest prices are reached from December to March, while the summer months are characterized by lower prices with the lowest peak in July.

Table 12: Coefficients for day of the week dummies.

Monday	5.20***	Tuesday	5.79***
	(1.35)		(1.33)
Wednesday	5.22***	Thursday	6.13***
	(1.28)		(1.38)
Friday	5.17***	Saturday	1.57
	(1.43)		(1.34)

Note: OLS with Swedish price as dependent variable. Number of observations:1,424. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

[47] See Tables A5 in Appendix C.2

Table 13: Coefficients for monthly dummies.

February	4.17** (2.10)	March	-1.98 (1.93)	April	-7.53 (1.63)
May	-10.64*** (1.67)	June	-12.65*** (1.61)	July	-14.60 (1.69)
August	-12.77*** (1.64)	September	-11.11 (1.68)	October	-9396 (1.59)
November	-8.19*** (1.67)	December	-3.03 (2.26)		

Note: OLS with Swedish price as dependent variable. Number of observations:1,424. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

3.3.2 Swedish bidding zones

Table 14 describes the prices in the four bidding zones during the time period after the re-configuration. Zones 1 and 2 are almost identical for all the observed statistics, while they differ from Zones 3 and 4. In the observed time period, from November 2011 to November 2013, the prices among all the four zones converged for 466 days out of 731 (64%)[48] Zone characteristics help explain the prices and their relative difference.

Table 14: Prices in the four bidding zones.

Zone	Mean	Std. Dev.	Min.	Max.
SE1	35.67	10.24	7.38	99.61
SE2	35.70	10.23	7.38	99.61
SE3	36.15	10.79	7.38	101.26
SE4	37.49	10.98	7.38	101.26

Note: Prices in the four different zones in Sweden. Average of daily price over the post-treatment period from November 2011 to November 2013.

In the Northern parts of the country (SE1 and SE2) the power balance is significantly stronger than in the Southern zones, where 82% of the consumption takes place (Table 15). Moreover, while most of the consumption takes place in the Southern part (SE3 and SE4), the 81% of the electricity from hydroelectric is produced in the Northern part (SE1 and SE2)[49].

A proper bidding zone configuration allows for a more efficient usage of the transmission grid and improved trading opportunities[50]. An inefficient use of the network that requires many remedial actions favours intra-zonal trading at the expense of the

[48]SE1 and SE2 converged for 712 days (97%), SE2 and SE3 converged for 626 days (86%), SE3 and SE4 converged for 526 days (72%).

[49]For yearly summary statistics see A10 in Appendix C.2.

[50]Stakeholders argue that smaller bidding zones imply lower market liquidity and lower competition levels. This debate, although, is beyond the scope of this analysis

Table 15: Consumption and production by bidding zone.

Zone	Consumption	Production	Prod. Hydro
SE1	8.2	19.9	19.2 (96%)
SE2	15.2	39.5	37.4 (95%)
SE3	84.8	78.8	11.4 (14%)
SE4	23.7	6.3	1.7 (27%)
Total	131.9	144.5	69.7 (48%)

Note: Average yearly production and consumption from November 2009 to November 2013. Values are expressed in TWh.

cross-zonal flows[51]. Therefore, the Nord Pool re-configuration is expected to increase flows with the neighbouring countries due to lower prices in the North and to diminish the internal flows towards the South because of the entrance of additional suppliers in zones 3 and 4, driven by higher prices. The flows between zones and with the neighbouring countries are respectively reported in Figures 14 and 15. In Figure 14 we include flows from zone 2 to 3 and from zone 3 to 4[52]. Graphical evidence (Figure 14) suggests that both zones 4 and 3 are net importers and that cross-zonal flows seem to increase over time. As for trading with neighbouring countries, Figure 15 shows no evidence of an increase for Denmark and Norway, while it seems that there is a net increase in export to Finland[53].

3.4 Methodology and Results

3.4.1 Difference-in-differences

We apply the diff-in-diff approach to estimate the impact of the re-configuration on national level prices, as described in Section 3.3.1 and we use the System price as comparison group. We estimate the following model:

$$Price_{it} = \alpha + \beta_0 Post_t + \beta_1 Treated_i + \beta_2 Treated_i * Post_t + f_i + f_t + u_{it} \qquad (9)$$

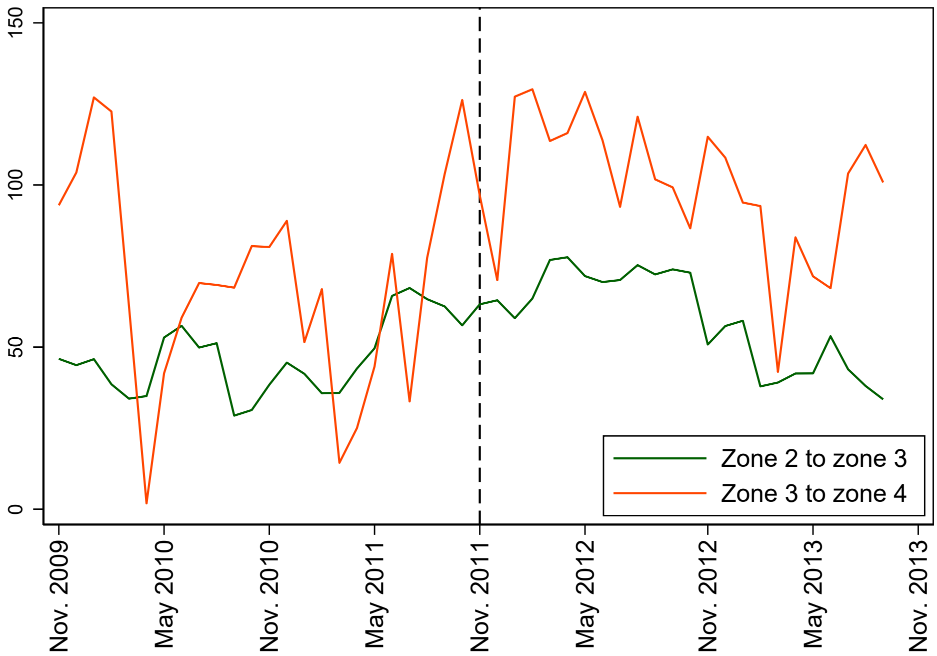
where i indicates either Sweden or the System and t is the day indicator. The key indicator variable in this specification is the interaction term between $Post_t$,

[51] Regulation (EU) 2019/943: *"the configuration of bidding zones in the Union shall be designed in such a way as to maximise economic efficiency and to maximise cross-zonal trading opportunities in accordance with Article 16, while maintaining security of supply."*

[52] For further details on the capacities between zones and countries, see Figure A7 in Appendix C.2

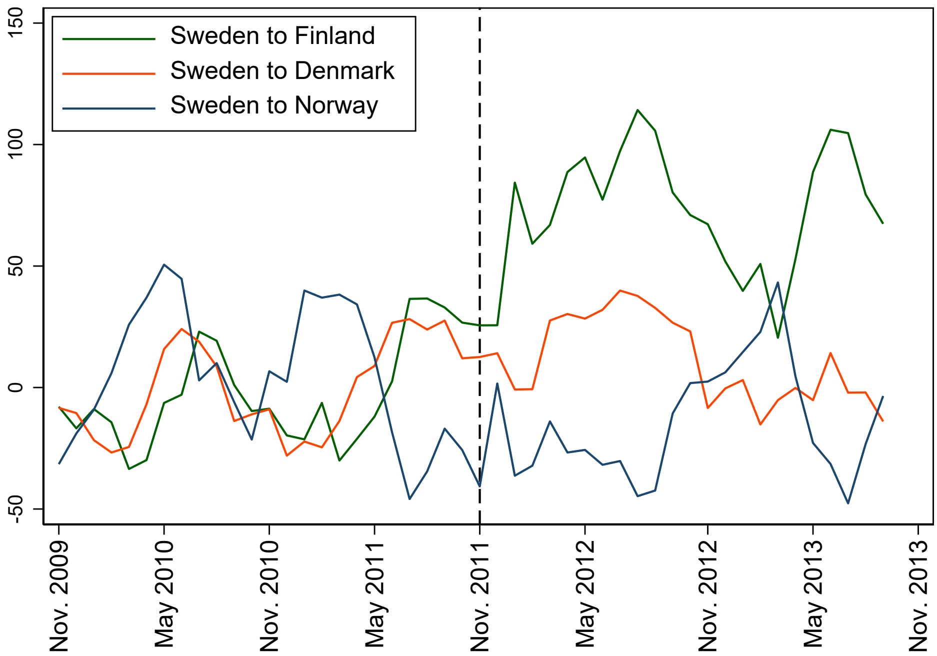
[53] The observed increase in flows between Finland and Sweden does correspond to additional available transmission capacities as shown in Table A12 in Appendix C.2.

Figure 14: Flows within zones.



Note: This figure shows the distribution of monthly net flows between zones in Sweden. Value are expressed in thousand MWh.

Figure 15: Flows with neighbouring countries.



Note: This figure shows the distribution of monthly net flows between Sweden and its neighbouring countries. Value are expressed in thousand MWh.

a dummy variable equals to one after the policy intervention, $Treated_i$, a dummy variable equals to one for the treated group (Sweden). The coefficient of the interacted term $Treated_i * Post_t$ estimates the effect of the policy intervention for Sweden in the post treatment period. We add daily and area fixed effects. Errors are clustered at the area level to allow for dependence of observations within these groups. The key identifying assumption is that the System price is not affected by the policy intervention. This assumption is reasonable since the System price, by definition, is calculated for the Nordic area by setting internal transmission capacities between the bidding zones to infinity. The System price is a common benchmark for the Nordic Market that allows for more liquidity because, by assuming no bottlenecks, it limits the effect of temporary transmission constraints and reduces the risk of manipulation. We compare the untreated group, the system price, to the treated group, the Swedish price, and we assume that if no treatment had occurred the distance between the two groups would have stayed the same. Unfortunately, we can not add additional control as we do not have country specific characteristics for the system price (e.g. GDP, temperatures, population). For the parallel trends assumption to hold, we need to restrict our sample to 500 days before and after the policy intervention (roughly 3 semesters before and afterwards November 2011). Before running the full regression we do a back-of-the-envelope calculation, whose results are shown in Table 16. In Column 1 we report the means for price for Sweden, before and after the policy intervention, in Column 2 the price average values for the System, before and after, and in Column 3 the difference between Sweden and System average values, before and after. Price mean values are higher for Sweden both in the pre-treatment and in the post-treatment period. Taking the difference of the differences we get an increase in price equal to 0.2 euros.

Table 16: Back-of-the-envelope diff-in-diff.

	(1)	(2)	(3)
	Sweden	System	Difference
Before	51.43	50.42	1.01
After	35.12	33.91	1.21*
Diff-in-diff	16.31***	16.51***	0.20

Note: The number of observations before the market re-configuration is 968; the number of observation after the re-configuration is 976. Values in the top and bottom percentile are dropped. The significance is estimated with a t test.

Table 17 presents estimates from a linear regressions with time and group fixed effects, as in Equation 9. Each Column reports the estimates of the coefficient of interest, the interaction between $Treated_i$ and $Post_t$. The dependent variable is always the price level, for two price areas, Sweden and the System. In column 1a, we only include day and area fixed effects and we find that Swedish price increases by 0.21 euros significant at the 1 percent level (approximately 0.5 percent of the average price) in the post-intervention period. In column 1b, we add to the model area specific trends, allowing for non-parallel trends. When the trend is included, the estimates remain positive and significant but gets smaller (coefficient equal to 0.06 euros significant at the 1 percent level, 0.1 percent of the average price). As a robustness check, we run the same regression on a reduced sample: we drop the 200 days in the window around the cutoff to check whether the estimated effect of the policy is due to an immediate reaction that does not last over time. In both the specifications we find that the result is stronger if we remove the window around the threshold, this implies that the result is stronger away from the threshold. In column 2a, with day and area fixed effects, we find that Swedish price increases by 0.50 euros significant at the 1 percent level (approximately 1 percent of the average price) in the post-intervention period. In column 2b, we add to the model area specific trends, we get a coefficient equal to 1.77 euros significant at the 1 percent level (approximately 4 percent of the average price)[54]. These results are not surprising since the re-configuration is expected to increase prices in the North and to decrease prices in the South, giving stronger price signals in the areas where supply is scarce. As theoretically proved by (Bjørndal & Jörnsten, 2001), we show that the average Swedish price increased after the re-configuration. This result also confirms that, as stated by the European Commission, prices in Sweden were low because the internal congestion was not reflected in the price formation mechanism. Before the market re-configuration, the internal congestion was more managed internally while keeping the day-ahead prices low. After the policy intervention, the internal congestion emerged because it was reflected into higher prices. Moreover, by curtailing ATCs, the Swedish TSO was also excluding the demand from the neighbouring countries by prioritizing

[54]We repeat the main regressions keeping values in the top and bottom percentile. Results are shown in Table A13 in Appendix C.3. In December 2010 daily prices in Sweden were influenced by both an increase in demand for electric heating due to extremely cold weather conditions and reduced availability of the Ringhals-4 nuclear reactor. This capacity reduction resulted in a positive deviation of Swedish price compared to the System price (Directorate-General for Energy, 2011). Due to the presence of these exceptional high spikes, the estimated coefficients have opposite sign.

internal flows.

Table 17: Diff-in-diff estimation.

	Full sample		Reduced sample	
	(1a)	(1b)	(2a)	(2b)
$Treated_i \cdot Post_t$	0.21***	0.06***	0.50***	1.77***
	(0.00)	(0.00)	(0.00)	(0.00)
Mean Dep. Var.	42.69	42.69	43.86	43.86
Observations	1,944	1,944	1,556	1,556
R-squared	0.99	0.99	0.99	0.99
Trend	No	Yes	No	Yes
FE	area, day	area, day	area, day	area, day

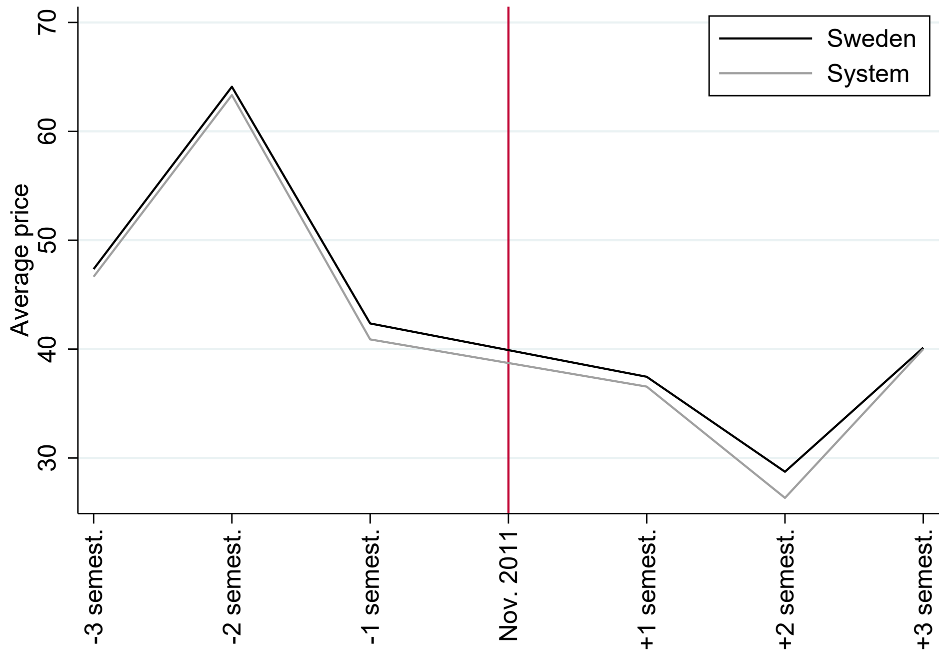
Note: The table reports the coefficients and standard errors (in brackets) associated with $Treated_i * Post_t$ from the estimation of Equation 9. In the full sample, Columns 1a and 1b, variables are observed 500 days before and after the policy intervention. In the reduced sample, Columns 2a and 2b, variables are observed 400 days before and after a window of 100 days from the policy intervention (100 days before and 100 days afterwards). Values are expressed in euros. Values in the top and bottom percentile are dropped. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Figure 16 shows that the treated and the untreated groups were heading in the same direction in the pre-treatment period. The distance between the Swedish and the System price is constant in the lead-up to treatment. Prices are averaged by six months intervals for simplicity. In the pre-intervention period trends are parallel, with a slight divergence in the last semester. In the post-intervention period prices keep following a parallel trend for the first semester, however, as captured by the diff-in-diff, the Swedish price decrease less than the system price and their trend starts diverging. We further verify that the parallel trends assumption holds, by performing a falsification test to see if the trends are different. We run the following regression:

$$Price_{it} = \alpha + \beta_0 Trend_t + \beta_1 Trend_t * Treated_i + u_{it} \tag{10}$$

As a robustness check, we run Equation 10 using only data from before the treatment period where the $Trend_t * Treated_i$ coefficient allows for trends to be different for the two groups (results shown in Table A14 in Appendix C.3). The coefficient of interest is -0.002 with standard errors of 0.003, meaning that there is no evidence of a different trend between the two groups. As this test shows that parallel trends assumption is likely to hold, we are more prone to rely on the coefficients in Columns 1a and 2a from Table 17.

Figure 16: Parallel trends.



3.4.2 Regression discontinuity in time

We apply the Regression Discontinuity in Time (RDiT) to estimate the impact of the re-configuration on cross-zonal and cross-country flows. In the RDiT , the running variable of the regression discontinuity is time and it defines the treatment eligibility: the treatment begins after a threshold in time is reached. This empirical analysis presents several differences with a typical regression discontinuity design (as the one presented in Chapter 2), the most relevant ones are: there is no need for testing for bouncing around the threshold, because it is not possible; it does not rely on cross-sectional analysis, therefore the window defined by the running variable can not be properly shrunk; a time-series approach needs to be adopted[55]. These issues raise the need for additional controls to identify the model correctly. We include both the auto-regressive term of the dependent variable to account for the time series nature of the data, and additional control variables to adjust for unobservable factors correlated with time. As for the time window, we use four years (two years before and two after the treatment) needed to absorb the seasonality effect. The identifying

[55]Among the papers using RDiT: Anderson, (2014); Auffhammer & Ryan Kellogg (2011); Chen & Whalley (2012); Davis & Kahn (2010).

assumption behind our discontinuity is that absent the bidding zone reconfiguration flows would have changed smoothly at the cutoff date. Therefore we assume that the flows on the days before the policy intervention work as a counterfactual for the days after the the intervention. Any difference that smoothly changes close to the cutoff is captured by the running variable.

Our main specification uses local linear regressions within a given bandwidth of the treatment threshold of two years, and controls for the running variable (distance from the cut-off day) on both sides of the threshold. Formally, we estimate the following model:

$$\begin{aligned} Flow_t = \alpha + \beta_0 Flow_{t-1} + \gamma_0 \cdot 1(Time_t \geq C) + \gamma_1 (Time_t - C) + \\ \gamma_2 1 \cdot (Time_t \geq C) \cdot (Time_t - C) + \eta_0 X_t + \eta_1 P_t + \\ f_{dow} + f_{month} + f_{year} + u_t \end{aligned} \tag{11}$$

where C is the cutoff day, equal to November 1st, 2011. In Equation 11, the RDiT estimator, γ_0, calculates the local average treatment effect (LATE) of the market re-configuration. The outcome variable $Flow_t$ is either the cross-zonal or the cross-country flows and is measured at day t, and $Flow_{t-1}$ is its lagged value. $(Time_t - C)$ is the distance in days from the cutoff and C is the cutoff defining the treatment group, so that the dummy variable $1 \cdot (Time_t \geq C)$ defines treatment and control groups, that correspond to before/after the treatment: it takes the value of one after the policy intervention, and zero otherwise. In some specifications, we also add the squared terms of the running variable and its interaction with the post-treatment variable. We add control variables that improve the efficiency of the estimation. In the vector X_t we include monthly carbon prices[56], and, for the cross-country flows we also include instrumented electricity prices. The vector P_t includes a third-order polynomial time trend to control for time series variation. We include day of the week monthly and year fixed effects. u_i is the error term, since the observations are likely to be serial correlated we cluster the standard errors at monthly-level.

Changes in net flows within Sweden and between Sweden and its neighbouring countries are an indicator of the re-configuration's effectiveness. Within Sweden, the policy is expected to decrease the flows going from North to South because it encour-

[56] Carbon prices are European Union Allowance prices from the European Union Emission Trading System.

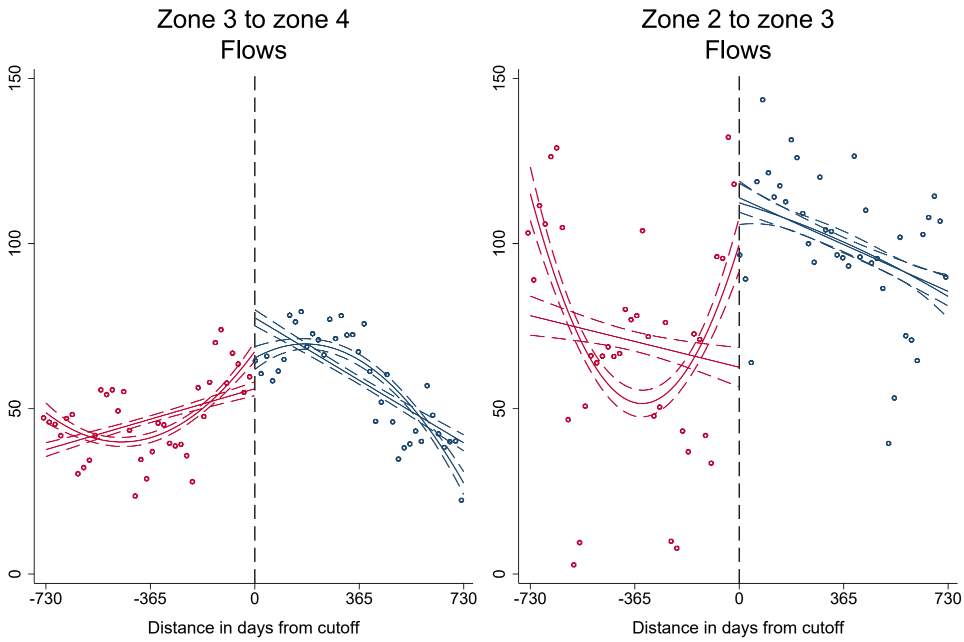
ages production in the South by setting different prices in case of congestion. Between Sweden and its neighbouring countries, flows are expected to change depending on the prices of the neighbouring countries. In fact, the national TSO was accused to curtail capacity with the neighbouring countries due to internal congestion. Absent internal congestion thanks to the four bidding zones, a decrease in net flows from North to South should be observed.

Cross-zonal flows. We estimate the cross-zonal flows with the regression described in Equation 11. We consider the flows between zones: from zone 3 to zone 4 and from zone 2 to zone 3. Both of the analysed net flows describe the flows going from North to South. We do expect a slight decrease in flows due to minor congestion. In fact, the four different prices resulting in the new configuration (shown in Table 14) might have modified internal flows decreasing the amount of electricity going from North to South. With higher prices, electricity producers in the South are encouraged to enter the market and decrease the imbalance between supply and demand in the Southern areas of Sweden.

Before running our regressions, we show graphical evidence of the RDiT around the cutoff (Figure 17). We plot the means of the daily flows grouped in bins, then we add the linear and the quadratic fit for OLS on both sides of the cut-off. Given the shape of the data, our preferred specification is the quadratic fit, although the seasonality of the data makes it difficult to visually inspect the discontinuity. For the flows from zone 3 to zone 4 we do not notice any significant decrease either in the plotted bins or in the quadratic form, although we get a positive change in the linear fit. As for the net flows from zone 2 to zone 3 we notice an increase in net flows at the cutoff in both the linear and the quadratic form. In addition, if we look at the trend in the post-treatment period, we see a decreasing trend that suggests that over time the South reduced its reliance on the production in the North of the country.

Our main estimations for cross-zonal flows are presented in Table 19. In Columns 1 and 3 we run the linear RDiT, while in Columns 2 and 4 we run the RDiT in its quadratic form. We use both the full sample and a reduced sample, the donut hole where we drop observations in the 40 days around the cutoff date. For the net flows from zone 3 to zone 4 we do not find neither statistically significant or consistent results across all the specification. For the net flows from zone 2 to zone 3, contrary to

Figure 17: Graphical evidence: cross-zonal flows.



Note: This figure shows the linear and quadratic fit for OLS on both sides of the cutoff. Values are expressed in thousand MWh.

graphical evidence, we notice a consistent negative sign that, however, is statistically significant only in the donut hole specification with the squared form. The coefficient of interest is equal to -9.63 with standard errors of 3.44 that correspond to decrease of 11% of the net flows.

To check how flows behave at varying bandwidths, we perform a bandwidth sensitivity test where we reduce the bandwidth from 730 (two years) to 330 days (less than one year) around the cutoff date. We run the RDiT as in Equation 11 using the quadratic form as in Column 2 of Table 18. We show results in Figure A8 in Appendix C.3, and we notice that there seems to be a slight decrease in net flows from zone 2 to zone 3. The negative effect found in the donut hole, is validated by this robustness test where we find a significant decrease from 380 to 630 days. As for the flows from zone 3 to zone 4, it is not notably sensitive to the bandwidth choice.

Finally, as a robustness check, we run a falsification test where we drop the observations after the policy intervention and we introduce a fake treatment on November 1st, 2010. We repeat our RDiT as in Equation 11 using two years (from November 2009 to November 2011) with a fake treatment introduced halfway. We run the test

Table 18: Regression discontinuity in Time estimation: cross-zonal flows.

	Full sample		Donut hole	
	(1)	(2)	(3)	(4)
	Panel A: Zone 3 to zone 4			
$1(Time_t \geq C)$	1.99	-5.41	3.17	-6.05
	(2.65)	(3.53)	(3.12)	(4.56)
Mean Dep. Var.	52.74	52.74	52.48	52.48
R-squared	0.75	0.76	0.76	0.76
	Panel B: Zone 2 to zone 3			
$1(Time_t \geq C)$	-2.38	-4.68	-4.59	-9.63**
	(2.67)	(2.81)	(4.08)	(3.44)
Mean Dep. Var.	85.05	85.05	84.48	84.48
R-squared	0.89	0.89	0.89	0.89
Observations	1,460	1,460	4,420	1,420
RD	linear	quadratic	linear	quadratic
Cubic Trend	Yes	Yes	Yes	Yes
FE	time	time	time	time

Note: The table reports the coefficients and standard errors (in brackets) associated with $1(Time_t \geq C)$ from the estimation of Equation 11. Variables are observed 2 years before and 2 years after the policy intervention. Values are expressed in thousand MWh. In all the regressions we include the lag of the dependent variable, daily temperatures and precipitations in Sweden, monthly carbon prices, a third-order polynomial trend and time fixed effects (day of the week, month and year). Columns 1 and 3 are based on linear RDiT, Column 2 and 4 on quadratic RDiT. In Column 1 and 2 we use the full sample. In Column 3 and 4 we use a donut hole and we drop observations in the 40 days around the threshold. Errors are clustered at monthly level. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

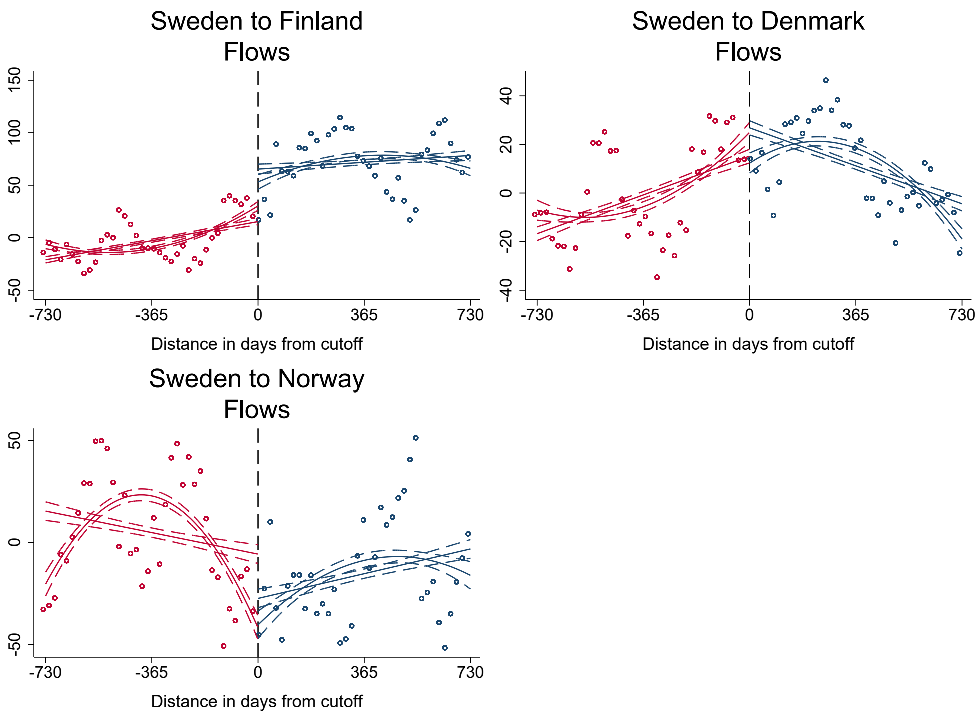
using the full sample and the quadratic form. Across all the specifications, we do not find any effect of the fake treatment.

Cross-country flows. We analyse cross-country flows because after the re-configuration the likelihood that the national TSO could manipulate the Available Transmission Capacities strongly decreased. For this reason we expect to find an increase in the flows with neighbouring countries. In the new configuration, the national TSO did not have to intervene by curtailing ATCs because the new market mechanism with different prices in each zone already took into account the internal bottlenecks.

We estimate the cross-country flows with the regression described in Equation 11. In this model specification we also include the Swedish price since it influences the net export to neighbouring countries. However, as the Swedish price is endogenous to the model, we use a two-stage least-squares (2SLS) including average temperatures in Sweden as an instrumental variable.

Before running our regressions, we show graphical evidence of the effect of the market re-configuration (Figure 18). As for cross-zonal flows, we plot the means of the daily flows grouped in bins and we add the linear and the quadratic fits for OLS on both sides of the cut-off. Also in this case, given the shape of the data, our preferred specification is the quadratic fit. We consider the flows with countries in the Nord Pool area: Finland, Denmark and Norway. For the net flows from Sweden to Finland we notice that the fitted lines present a break at the discontinuity in both the linear and the quadratic form and the observed data show an increase in the the days after the cutoff. We do not notice any discontinuity for net flows from Sweden to Denmark, however we get opposite although small variations for the linear and quadratic fit, suggesting that modelling the time series properly is of pivotal importance in order to have accurate estimates. For the net flows from Sweden to Norway we do not see any discontinuity either in the plotted data, or in the quadratic fit. However, the linear form shows a decrease in net flows with a decreasing trend on the left side and an increasing trend on the right side.

Figure 18: Graphical evidence: cross-country flows.



Note: This figure shows the linear and quadratic fit for OLS on both sides of the cutoff. Values are expressed in thousand MWh.

As we want to estimate the net flows after the policy intervention, we need to account for factors that influence their pace. One of the main determinants that we want to include is the price in Sweden, which however is endogenous to the model. Therefore, we use a 2SLS where we use average daily temperatures in Sweden as an exogenous source instrument[57]. We use this instrument to account for exogenous variation in the Swedish price. Temperatures contribute significantly to the price formation since they are strongly related with consumption and production. On one side, the demand for electricity in winter sharply increases due to the widespread electric heating in the whole Nordic region[58]. On the other side, the cold temperatures and the ice create operational constraints in the system.

We examine the validity of the instrumental variable by showing results from the first stage and from the weak identification tests (Table A16 in Appendix C.3). In the first-stage the endogenous variable, price in Sweden, is regressed on the exogenous variables and the excluded instrument, average temperatures in Sweden. The estimated coefficients show that when temperatures in Sweden decrease the electricity price goes up by almost 2 euros. We find that this result is significant for all specifications at 1% level. The weak identification tests we use include the traditional F-statistics and the Montiel-Plueger (M-P) test[59]. We include the M-P test as it reports the effective statistics at 5% level and critical values in presence of heteroskedasticity. The reported F-statistics always satisfy the rule of thumb of a value at least equal to 10 (Staiger et al., 1997). The M-P test also suggests that the instrument is strong because the statistics are (almost) above the critical value of 23.11 that corresponds to a maximum bias of 10% [60]. From these tests and from the first stage we can conclude that average temperatures are a relevant instrument for the price in Sweden.

Our main estimations for cross-country flows are presented in Table 19. In Columns 1 and 2 we run simple OLS regressions, respectively in their linear and in their quadratic form, that include the endogenous Swedish price. In Columns 3 and 4 we show the results from the 2SLS using the quadratic form, respectively using the full

[57]The instrumental variable, temperatures, is the beginning of a casual chain that influences the endogenous variable, price, that, in turn, affects the dependent variable, net flows (Angrist & Pischke, 2015).

[58]See the strong inverse relationship between temperatures and price in Figure A5 in Appendix C.2.

[59]In case of one instrument and one endogenous regressor the F test of excluded instrument and the M-P statistic coincide.

[60]The interpretation is that there is a 5% probability that the bias in the estimator is 10% of the worst possible case.

Table 19: Regression discontinuity in Time estimation: cross-country flows.

	OLS		2SLS	
	(1)	(2)	(3)	(4)
	Panel A: Sweden to Finland			
$1(Time_t \geq C)$	1.67*	2.95*	4.04**	4.95*
	(0.88)	(1.48)	(1.65)	(2.83)
F-statistic	-	-	23.19	22.24
Mean Dep. Var.	34.52	34.52	34.52	34.97
R-squared	0.95	0.95	0.95	0.95
	Panel B: Sweden to Denmark			
$1(Time_t \geq C)$	4.17	-6.53	1.60	1.97
	(3.96)	(5.63)	(2.89)	(4.22)
F-statistic	-	-	22.86	21.96
Mean Dep. Var.	5.21	5.21	5.21	4.96
R-squared	0.80	0.81	0.78	0.78
	Panel C: Sweden to Norway			
$1(Time_t \geq C)$	0.24	-8.39	-0.88	0.14
	(2.41)	(3.66)	(3.01)	(3.99)
F-statistic	-	-	28.39	26.91
Mean Dep. Var.	-5.25	-5.25	-5.25	-4.31
R-squared	0.86	0.86	0.86	0.86
Observations	1,460	1,460	1,460	1,420
RD	linear	quadratic	quadratic	quadratic
Cubic Trend	Yes	Yes	Yes	Yes
FE	time	time	time	time

Note: The table reports the coefficients and standard errors (in brackets) associated with $1(Time_t \geq C)$ from the estimation of Equation 11. In Columns from 1 to 3 variables are observed 2 years before and 2 years after the policy intervention. In Column 4 we use a donut hole and we drop observations in the 40 days around the threshold. In all the regressions we include the lag of the dependent variable, daily temperatures and precipitations in Sweden, monthly carbon prices, a third-order polynomial trend and time fixed effects (day of the week, month and year). In Column 1 we only include the linear terms of the RDiT, in Columns from 2 to 4 we also include the quadratic terms of the RDiT. Columns 3 and 4 report the estimated local average treatment effect from a separate two-stage least squares regression. The reported F-statistic is the F-statistic for excluded instrument. Values are expressed in thousand MWh. Errors are clustered at monthly level. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

sample and the donut hole[61]. Across the four specifications net exports to Finland are significantly positive, ranging from 1.67 (5% of the net export) to 4.04 (12%), suggesting that the re-configuration resulted in an increase in net export to Finland. From Table A12 in Appendix C.2 we get prima facie evidence that after the re-configuration there was an increase in Available Transmission Capacities that predicts the increase in net flows from Sweden to Finland that we find with our estimation.

In Column 1 the estimated increase in net flows is equal to 1.67 (standard errors

[61] In the donut hole we drop observations in the 40 days around the threshold.

of 0.88) while in Column 2 we get a coefficient of 2.95 (standard errors of 1.48) which corresponds to almost 9 % of the average flow. In Column 3the estimate coefficient is equal to 4.04 (standard errors of 1.65) which corresponds to an increase of 12% of the average flows. Finally, in the sample with the donut hole (Column 4) we find a slightly bigger effect of 4.95 (standard errors 2.83) which corresponds to an increase in net flows of 14%. This greater effect suggests that the effect is not guided by an immediate response. We do not find any significant effect for the other observed net flows (from Sweden to Denmark and from Sweden to Norway) with big standard errors of the estimated coefficients. These results are robust to both the linear and the quadratic form.

To check the robustness of these results, we perform a bandwidth sensitivity test where we reduce the bandwidth from 730 (two years) to 330 days (less than one year) around the cutoff date. We run the 2SLS as in Equation 11 using the quadratic form as in Column 3 of Table 19. Results, shown in Figure A9, are not notably sensitive to the bandwidth choice. The positive effect found for the net flows from Sweden to Finland holds for almost all the bandwidth choices, with smaller coefficients between 580 and 680 days. The coefficients of the net flows from Sweden to Denmark are approximately equal to zero and always statistically insignificant. Finally, the coefficient of the net flows from Sweden to Norway is statistically significant and negative between days 380 and 530, but firmly below zero and not statistically significant from 530 to 730 days.

As an additional robustness test, we run a falsification test as in Table A15 for cross-zonal flows. For the falsification test, we can not rely on the 2SLS as in Column 3 of Table 19 because in the shorter window adopted to run this test, the instrument chosen fails the weak identification tests. Therefore we use the OLS as in Column 2 of Table 19. The coefficients of the fake treatment, both in their linear and in their quadratic form show that there is no discontinuity at the fake cutoff date (see Table A17 in Appendix C.3).

3.5 Conclusions

Our measure of market efficiency is based on accurate price signals and the full use of cross-country transmission capacities. This short term signals are necessary to attract investments in generation, load and grid where they are needed the most to realise long-term solutions.

In this empirical work, we analyse the impact of the market re-configuration on price, cross-zonal and cross-country flows.

Our contribution to the literature is threefold. Firstly, we find evidence that in case of internal congestion, day-ahead prices slightly increase when smaller zones are adopted. The magnitude of the effect varies from 0.1 to 4 percent of the average price, depending on the preferred specification. As higher prices are set in zones where demand exceeds supply, this mechanism sends appropriate price signals for long-term investments, e.g., in additional power plants, in transmission lines. These results are robust to the control for trend that allows for diverging trends in the pre-treatment period and to the reduced sample where observations around the policy intervention are dropped. This result suggests that market efficiency improved as prices better reflected actual capacity constraints. Secondly, we find a decrease in the flows between the Northern and the Southern areas which reflects the more accurate price signals that emerged from the zone re-configuration. Higher prices in the South pushed producers in the South to enter the market and decrease the import from the Northern area. The size of the effect that we find is approximately equal to a decrease in average net flows of 11 percent. Thirdly, we find that the policy intervention increased the cross-country opportunities resulting in an increase in the net exports from Sweden to Finland. This increase was likely to be guided by the increase in Available Transmission Capacities between Sweden and its neighbouring countries. The size of the increase that we measure ranges from 5 to 14 percent. Both the RDiTs' results hold to varying bandwidths. We additionally run the falsification tests to exclude the hypothesis that the estimated effects are not driven by the policy intervention. These results suggest that inter-connectors between zones were fully used and this improved market efficiency.

We are aware that a full assessment of the bidding zones configuration also requires further analysis on liquidity and market structure. We hold space for future projects on these additional measures. Nevertheless, we want to stress out that our results prove compelling evidence that the market re-configuration was successful in giving more appropriate price signals and in eliminating congestion within the country.

References

(1) [dataset] EUA Emission Spot Primary Market Auction Report, available: eex.com/en/market-data/environmental-markets

(2) [dataset] National Oceanic and Atmosphere Administration, *Global Historical Climatology Network daily (GHCNd)*, available: https://www.ncei.noaa.gov/products/land-based-station

(3) Anderson, M. (2014). "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion", *American Economic Review*, 2014, 104 (9), 2763–2796.

(4) Angrist, J. D., Pischke , J.-S. (2015). "Mastering 'Metrics: The Path from Cause to Effect", *Economics Books*, Princeton University Press, edition 1, number 10363.

(5) Auffhammer, M., Kellogg, R. (2011). "Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality", *American Economic Review*, 2011, 101 (6), 2687–2722.

(6) Bask, M., Lundgren, J., Rudholm, N. (2009). "Market power in the expanding Nordic power market", *Applied Economics*, 43(9), 1035-1043. doi: 10.1080/00036840802600269

(7) Bjørndal, E., Bjørndal, M., Cai, H., Panos, E. (2018). "Hybrid pricing in a coupled European power market with more wind power", *European Journal of Operational Research*, 264(3), 919–931. doi: 10.1016/j.ejor.2017.06.048

(8) Bjørndal, M., Jörnsten, K. (2001). "Zonal Pricing in a Deregulated Electricity Market", *The Energy Journal*, 22, 51–74. doi: 10.2307/41322907

(9) Chen, Y., Whalley A. (2012). "Green Infrastructure: The Effects of Urban Rail Transit on Air Quality", *American Economic Journal: Economic Policy*, 2012, 4 (1).

(10) Davis, L.W., Kahn, M.E. (2010). "International Trade in Used Vehicles: The Environmental Consequences of NAFTA", *American Economic Journal: Economic Policy*, 2010, 2 (4), 58–82.

(11) Dijk, J., Willems, B. (2011). "The effect of counter-trading on competition in electricity markets", *Energy Policy*, 39(3), 1764–1773. doi: 10.1016/j.enpol.2011.01.008

(12) Directorate-General for Energy (2011). "Quarterly Report on European Electricity Markets".

(13) Egerer, J., Weibezahn, J., Hermann, H. (2015). "Two Price Zones for the German Electricity Market: Market Implications and Distributional Effects", *Energy Economics*, 59, 365–381. doi: 10.1016/j.eneco.2016.08.002

(14) European Union, "Regulation (EU) 2019/943 of the European Parliament and of the Council".

(15) Green, R. (2007). "Nodal pricing of electricity: how much does it cost to get it wrong?", *Journal of Regulatory Economics*, 31(2), 125–149. doi: 10.1007/s11149-006-9019-3

(16) Hadsell, L., Shawky, H. A., Marathe, A. (2004). "Estimating the Volatility of Wholesale Electricity Spot Prices in the US", *The Energy Journal*, Volume 25 (Number 4), 23–40. doi: 10.2307/4132335627

(17) Huisman, R. (2008). "The influence of temperature on spike probability in day-ahead power prices", *Energy Economics*, 30(5), 2697–2704. doi: 10.1016/j.eneco.2008.05.007

(18) Huisman, R., Huurman, C., Mahieu, R. J. (2007). "Hourly electricity prices in day-ahead markets", *Energy Economics*, 29(2), 240–248. doi: 10.2139/ssrn.918050

(19) Huisman, R., Kiliç, E. (2013). "A history of European electricity day-ahead prices", *Applied Economics*, 45(18), 2683–2693. doi: 10.1080/00036846.2012.665601

(20) Karakatsani, N. V., Bunn, D. W. (2008). "Forecasting electricity prices: The impact of fundamentals and time-varying coefficients", *International Journal of Forecasting*, 24(4), 764–785. doi: 10.1016/j.ijforecast.2008.09.008

(21) Kristiansen, T. (2012). "Forecasting nord pool day-ahead prices with an autoregressive model", *Energy Policy*, 49, 328 - 332. (Special Section: Fuel Poverty

Comes of Age: Commemorating 21 Years of Research and Policy) doi: 10.1016/j.enpol.2012.06.028

(22) Mazengia, D. H., Le Anh Tuan. (2008). "Forecasting spot electricity market prices using time series models", In *2008 IEEE international conference on sustainable energy technologies*(p. 1256-1261).

(23) Neuhoff, K., Twomey, P. (2010). "Wind power and market power in competitive markets", *Energy Policy*, 38(7), 3198–3210. doi: 10.1016/j.enpol.2009.07.031

(24) Sarfati, M., Hesamzadeh, M. R., Holmberg, P. (2019). "Production efficiency of nodal and zonal pricing in imperfectly competitive electricity markets", *IFN Working Paper*, No. 1264, 2019. doi: 10.1016/j.esr.2019.02.004

(25) Staiger, D., Stock, J. H. (1997). "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 65(3), 557–586. https://doi.org/10.2307/2171753

C Appendix

C.1 Bidding zone

In the electricity market, demand and supply must be balanced instantly, therefore when there is lack of capacity to transport the electricity from where it is produced to where it is demanded, the system incurs in congestion management issues. Congestion issues can be solved with market-based solutions that often take into consideration the market configuration, i.e. the bidding zones.

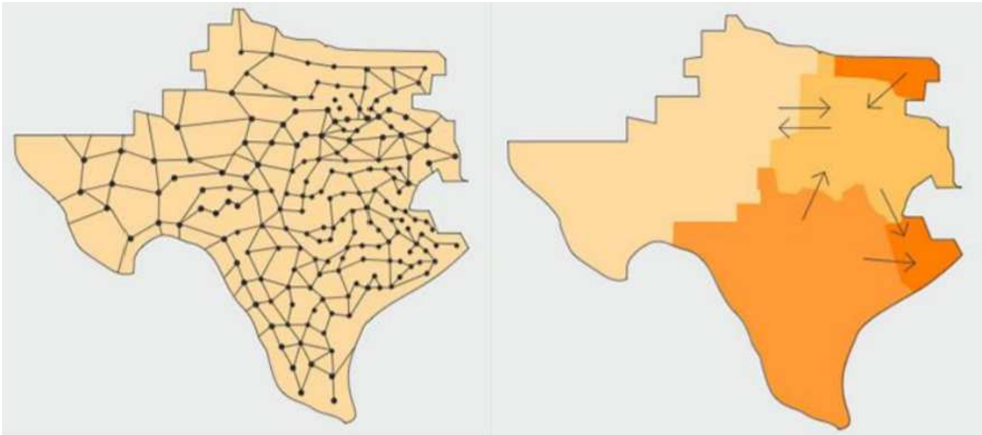
By definition, the bidding zone is the the largest geographical area within which market participants are able to exchange energy without capacity allocation (EU 543/2013).

In the zonal market, during the market clearing process that forms the price (as shown in Section 3.1) the transmission lines within a zone are set to infinity while the transmission lines between zones are taken into account. As the internal transmission constraints are ignored, the day-ahead dispatch may overload some transmission lines. For this reason, within a single bidding zone there should be no internal congestion. In fact, the internal congestion would affect both the internal market and the neighbouring zones. In case of internal congestion, as for the internal market, the Transmission System Operator (TSO) has to adopt costly remedial actions to remove congestion (e.g. re-dispatching, counter trading, Demand Side Response). As for the neighbouring zones, the TSO often reduce ex-ante the available transmission capacities because the operator wants to prioritize the internal flows. However, this response is not allowed as the Articles 18 and 35 of Treaty on the functioning of the European Union expressly prohibit discrimination based on nationality and quantitative restrictions on exports.

In a nodal market all the transmission constraints are considered in the day-ahead market and there are as many prices as nodes in the market. The market clearing mechanism itself eliminates congestion because the congestions are taken into account in the price formation. Zonal market can be interpreted as the aggregation of many nodal markets that converge in one single zone with uniform pricing.

In the European Union large bidding zones are widespread and in most cases they coincide with national borders. Nodal markets exist in the United States, Australia, New Zealand. In Figure A4 we give graphical representation of the difference between

Figure A4: Nodal market (on the left) vs zonal market (on the right).



the zonal and the nodal markets. On the left-hand side of the figure we show a nodal market where each dot corresponds to a node where the producer is paid with the local price. In this market configuration there are neither imports nor exports because the capacity constraints are already taken into account in the electricity dispatch. On the right-hand side of the figure we show a zonal market. The price is set within each of the colored areas and the arrows represent the inter-connectors that allow for flows between the zones. Throughout Chapter 3, we consider the re-configuration into four zones in Sweden as a proxy for the nodal market configuration because in extreme cases the small zones correspond to the nodes.

C.2 Data

$$Price\ Sweden = w_{1,t} * P_{1,t} + w_{2,t} * P_{2,t} + w_{3,t} * P_{3,t} + w_{4,t} * P_{4,t} \tag{A1}$$

Data on temperatures are retrieved from The Global History Climatology Network Daily, which is an integrated database of daily climate summaries from land surface stations in the world. It is a dataset from the National Oceanic and Atmosphere Administration from the United States of America. For each country, the database contains as many observations as the number of stations present in the country. Therefore, the country daily temperature is calculated by taking the averages of all the available stations.

Prices from the European Union Emission Trading Scheme are collected from the European Energy Exchange (EEX) where Emission Spot Primary Market takes place.

These prices are conventionally called EUA. Data is collected at monthly level. Flows between zone 3 and 4 are calculated as follows:

$$Flows_{3\to4} = Production_4 - Consumption_4 - Export_4 + Import_4 \quad \text{(A2)}$$

When calculating the flows from zone 2 to 3 we also subtract the net flows from zone 3 to 4 estimated with Equation A2. Therefore, we estimate:

$$Flows_{2\to3} = Production_3 - Consumption_3 - Export_3 + Import_3 + Flows_{3\to4} \quad \text{(A3)}$$

We can not estimate flows from zone 2 to 1 because we can not disentangle the exchange of zones 1 and 2 with the Northern part of Norway (NO4). We verify our results by comparing the flows we estimate with the Nord Pool data on the internal flows after the re-configuration: values are consistent with our estimation.

Table A9: Descriptive statistics from November 2009 to November 2013.

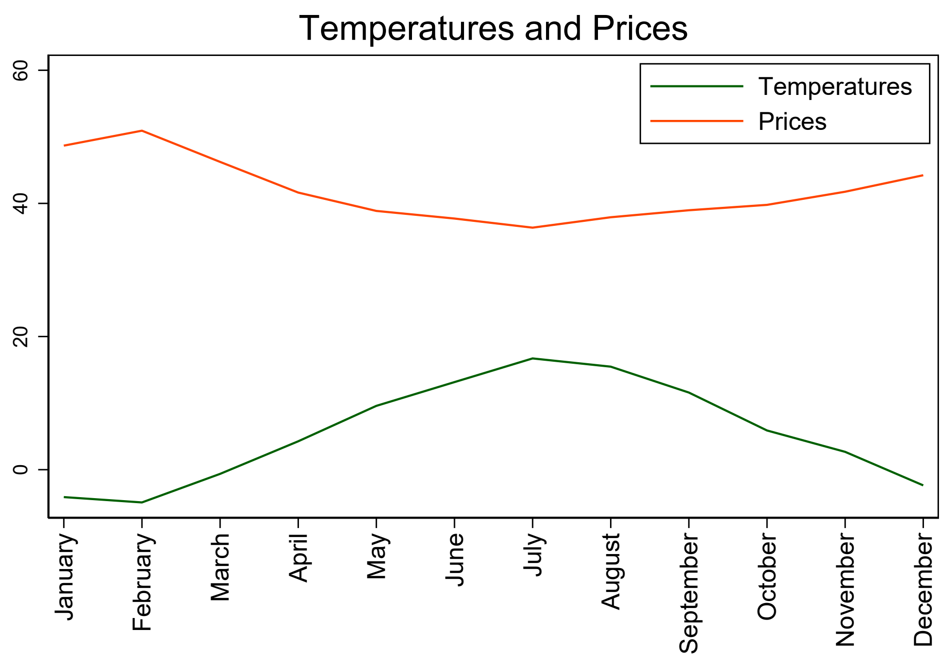
Variable	Mean	Std. Dev.	Min.	Max.
Price Sweden	44.495	21.883	7.377	505.681
Price System	42.490	14.623	5.790	134.804
CO_2 price	10.146	4.231	3.54	16.4
Flows from 2 to 3	85.053	39.265	-38.396	182.762
Flows from 3 to 4	52.743	187.827	-0.207	97.944
Flows from Sweden to Finland	34.525	46.939	-41.927	131.720
Flows from Sweden to Denmark	5.212	22.603	-44.986	47.52
Flows from Sweden to Norway	-5.250	33.606	-76.780	81.917
Temperatures Sweden	5.441	8.330	-16.725	21.25
N observations: 1,460				

Table A10: Consumption and production by bidding zone.

Zone	Consumption	Production	Prod. Hydro
	Nov 2009 - Nov 2010		
SE1	7.7	17.5	17.1 (97%)
SE2	15.2	38.2	36.7 (96%)
SE3	86.4	72.9	12.1 (17%)
SE4	23.9	6.8	1.7 (26%)
Total	133.2	135.4	67.6 (50%)
	Nov 2010 - Nov 2011		
SE1	7.8	17.5	16.8 (96%)
SE2	15.5	38.8	36.8 (95%)
SE3	85.8	79.2	10.8 (14%)
SE4	24.3	6.6	1.9 (29%)
Total	133.5	142.0	66.4 (47%)
	Nov 2011 - Nov 2012		
SE1	8.4	22.9	22.0 (96%)
SE2	15.1	45.0	42.9 (95%)
SE3	83.0	80.0	12.5 (16%)
SE4	23.2	5.8	1.8 (32%)
Total	129.7	153.7	79.2 (52%)
	Nov 2012 - Nov 2013		
SE1	8.7	21.8	20.7 (95%)
SE2	15.1	36.0	33.3 (93%)
SE3	83.9	83.1	10.0 (12%)
SE4	23.5	6.1	1.3 (21%)
Total	131.2	147.0	65.3 (44%))

Note: Average yearly production and consumption. Values are expressed in thousand MWh.

Figure A5: Temperatures and prices by month.



Note: Temperatures are expressed in degrees Celsius and prices in euros.

Table A11: Summary statistics of Swedish and System prices by 12 months windows.

Time window	Obs	Sweden		System.	
		Mean	Std. Dev.	Mean	Std. Dev.
2 windows before	365	51.50	32.66	47.98	11.07
1 window before	365	53.88	18.89	52.24	17.30
1 window after	366	32.55	12.07	30.99	11.18
2 windows after	365	40.08	7.12	38.78	6.19

Note: Summary statistics of average daily price by 12 months windows for System and Sweden prices. The windows start on November 1st and end on October 31st, e.g., "1 window after" the treatment is the average price from the first day after the intervention (November 1st, 2011) to 366 days afterwards (October 31st, 2012).

Figure A6: Cross-country flows between Sweden and Finland, import and export. Values expressed in thousand MWh and observed daily between November 2009 and November 2012.

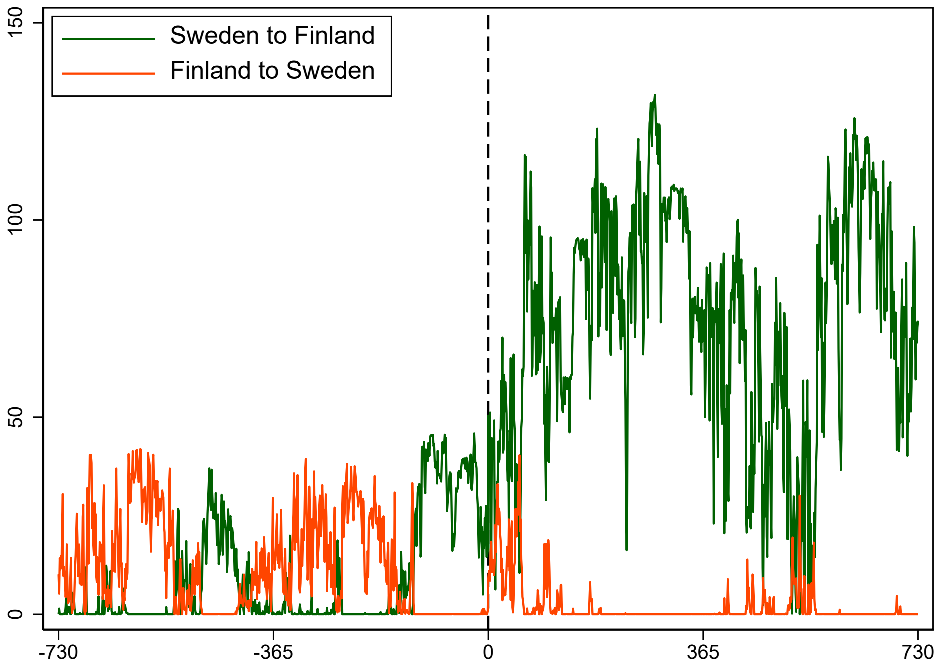
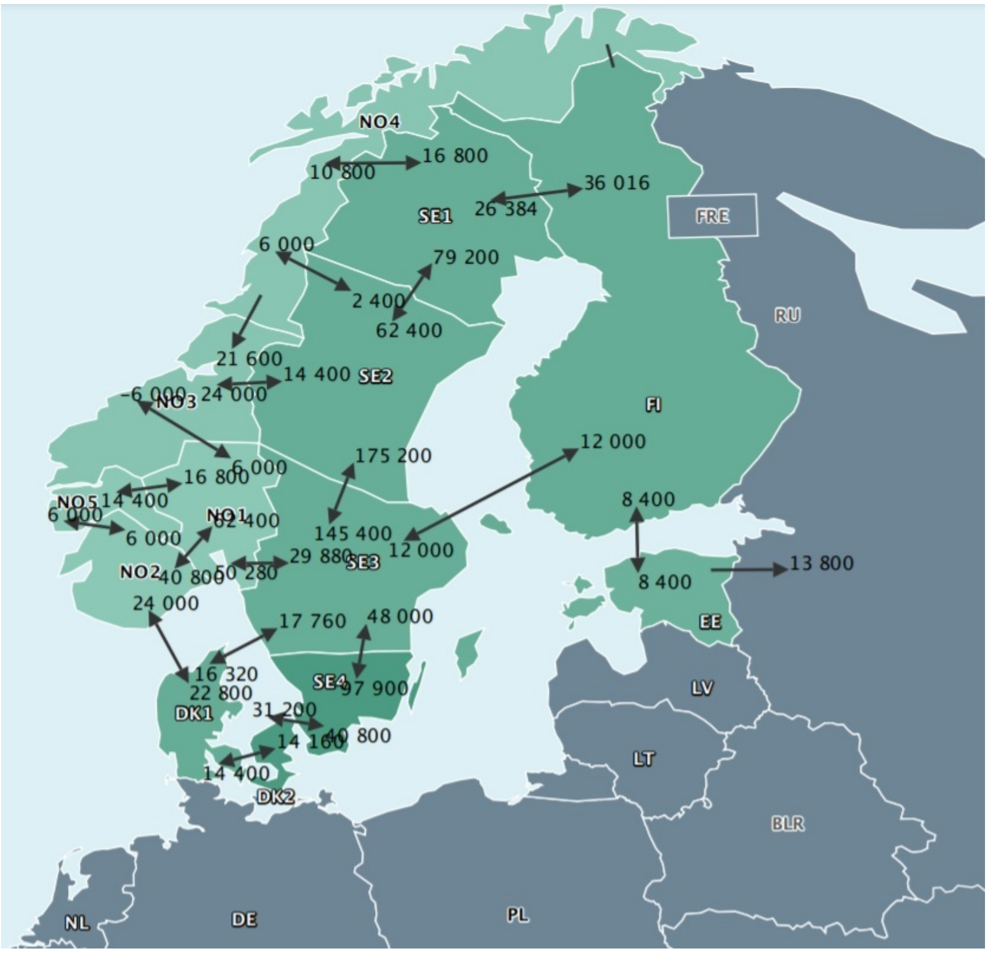


Table A12: Available Transmission Capacity (thousand MWh) between Sweden and Finland.

Date	SE → FI	FI → SE
2 windows before	43.492	38.848
1 window before	43.861	38.043
Date	SE → FI	FI → SE
1 window after	56.872	49.090
2 windows after	57.336	47.177

* *Note:* Summary statistics of average daily ATC. The windows start on November 1st and end on October 31st, e.g., "1 window after" the treatment is the average price from the first day after the intervention (November 1st, 2011) to 366 days afterwards (October 31st, 2012).

Figure A7: Capacities in the Nord Pool area as of November 1st, 2011.



C.3 Additional results

Table A13: Diff-in-diff estimation including outliers.

	Full sample		Reduced sample	
	(1a)	(1a)	(2a)	(2b)
$Treated_i \cdot Post_t$	-0.21***	-0.03***	-0.03***	1.45***
	(0.00)	(0.00)	(0.00)	(0.00)
Mean Dep. Var.	42.71	42.71	44.05	44.05
Observations	2,002	2,002	1,600	1,600
R-squared	0.99	0.99	0.99	0.99
Trend	No	Yes	No	Yes
FE	area, day	area, day	area, day	area, day

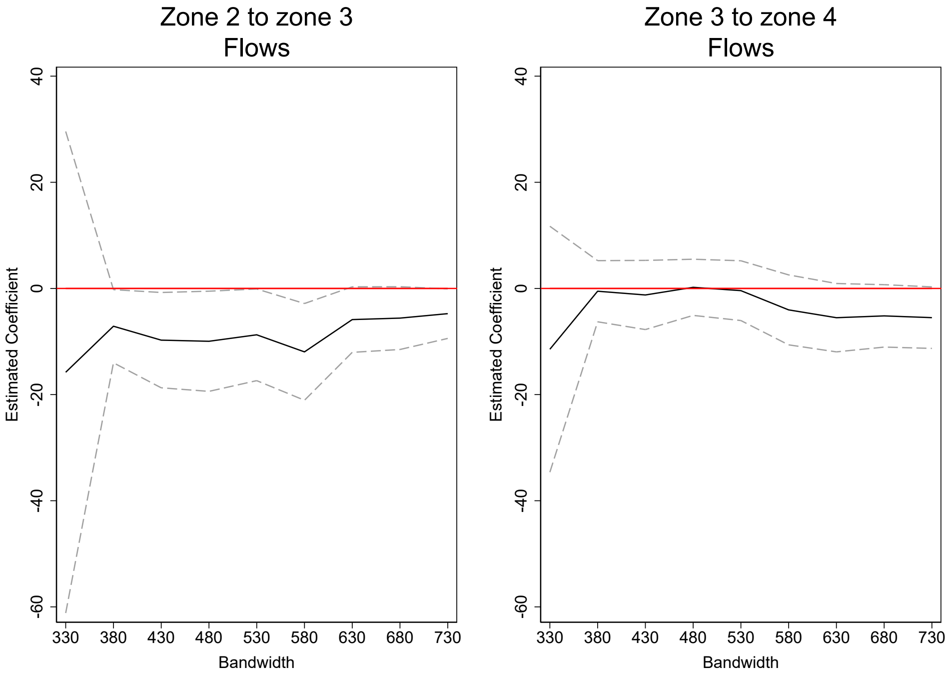
Note: The table reports the coefficients and standard errors (in brackets) associated with $Treated_i * Post_t$ from the estimation of Equation 9. In the full sample, Columns 1a and 1b, variables are observed 500 days before and after the policy intervention. In the reduced sample, Columns 2a and 2b, variables are observed 400 days before and after a window of 100 days from the policy intervention (100 days before and 100 days afterwards). Values are expressed in euros. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Table A14: Falsification test: price.

	Dep. Var.: Price
$Treated_i \cdot Trend_t$	-0.002
	(0.003)
$Trend_t$	0.005
	(0.004)
Mean Dep. Var.	50.93
Observations	968

Note: The table reports the coefficients and standard errors (in brackets) from the estimation of Equation 10. The sample is restricted to observations before the policy intervention. Values are expressed in euros. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Figure A8: Sensitivity of estimates to varying bandwidths.



Note: Coefficients and standard errors of the estimated coefficient $1(Time_t \geq C)$ from Equation 11 with the full sample and the quadratic RD as in Column 2 of Table 18. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Table A15: Falsification test: cross-zonal flows.

	Zone 3 to zone 4		Zone 2 to zone 3	
	(1)	(2)	(3)	(4)
$1(Time_t \geq F)$	34.65	-52.01	34.71	738.57
	(46.77)	(204.33)	(118.86)	(424.80)
Mean Dep. Var.	46.85	46.85	70.39	70.39
R-squared	0.78	0.78	0.92	0.92
Observations	729	729	729	729
RD	linear	quadratic	linear	quadratic
Cubic Trend	Yes	Yes	Yes	Yes
FE	time	time	time	time

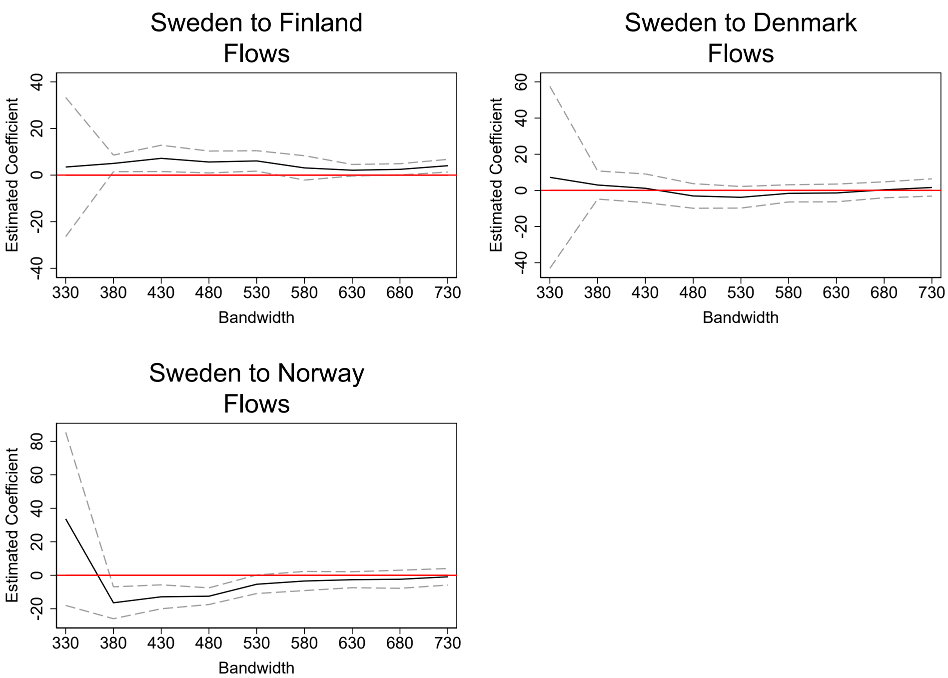
Note: The table reports the coefficients and standard errors (in brackets) associated with $1(Time_t \geq F)$ from the estimation of Equation 11. We replace the real cutoff date C with a fake treatment intervention F, one year before the actual treatment. F corresponds to November 1, 2010. Values after the real cutoff date are dropped, our sub-sample covers the period from November 2009 to November 2011. Values are expressed in thousand MWh. In all the regressions we include the lag of the dependent variable, daily temperatures and precipitations in Sweden, monthly carbon prices, a third-order polynomial trend and time fixed effects (day of the week, month and year). Columns 1 and 3 are based on linear RDiT, Column 2 and 4 on quadratic RDiT. Errors are clustered at monthly level. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Table A16: First stage regressions.

	Second stage DV: Price Sweden		
	(1)	(2)	(3)
From Sweden to	Finland	Denmark	Norway
Avg. Temper.	-1.88****	-1.85***	-1.78***
	(0.39)	(0.39)	(0.33)
F-statistic	23.19	22.86	28.39
M-P Critical Value	23.11	23.11	23.11

Note: The table reports first-stage coefficients from a separate two-stage least squares regression as in Column 3 of Table 19. The dependent variable is the daily price in Sweden and corresponds to the second stage dependent variable. The reported F-statistic is the F-statistic for excluded instrument. The M-P Critical Value refers to the Montiel-Pflueger robust weak instrument test for a maximum IV bias of 10% (we report this statistic because errors are clustered). Results differ because each regression has country specific controls. Variables are observed 2 years before and 2 years after the intervention. Errors are clustered at monthly level. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Figure A9: Sensitivity of estimates to varying bandwidths: cross-country flows.



Note: Coefficients and standard errors of the estimated coefficient $1(Time_t \geq C)$ from Equation 11 with the full sample and the quadratic RD as in Column 3 of Table 19. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Table A17: Falsification test: cross-country flows.

	Sweden to Finland		Sweden to Denmark		Sweden to Norway	
	(1)	(2)	(3)	(4)	(5)	(6)
$1(Time_t \geq F)$	-3.29	195.95	-33.24	200.81	-71.65**	12.90
	(23.41)	(125.62)	(22.77)	(210.51)	(28.42)	(159.74)
Mean Dep. Var.	-2.87	-2.87	-0.79	-0.79	4.83	4.83
R-squared	0.92	0.92	0.89	0.89	0.90	0.90
Observations	729	729	729	729	729	729
RD	linear	quadratic	linear	quadratic	linear	quadratic
Cubic Trend	Yes	Yes	Yes	Yes	Yes	Yes
FE	time	time	time	time	time	time

Note: The table reports the coefficients and standard errors (in brackets) associated with $1(Time_t \geq F)$ from the estimation of Equation 11. We replace the real cutoff date C with a fake treatment intervention F, one year before the actual treatment. F corresponds to November 1, 2010. Values after the real cutoff date are dropped, our sub-sample covers the period from November 2009 to November 2011. Values are expressed in thousand MWh. In all the regressions we include the lag of the dependent variable, daily temperatures and precipitations in Sweden, monthly carbon prices, a third-order polynomial trend and time fixed effects (day of the week, month and year). Columns 1 and 3 are based on linear RDiT, Column 2 and 4 on quadratic RDiT. Errors are clustered at monthly level. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Concluding remarks

Throughout this dissertation we estimated the effect of three different policies: the containment measures related to the COVID-19 pandemic, the "Robin Hood" tax and the electricity market bidding zone re-configuration.

In Chapter 1 we analyse the impact of the containment measures on the mobility indicator. As we aim at investigating the heterogeneous response to containment measures as a function of the knowledge about COVID-19, we interact Covid google searches with the stringency measure indicator. This quasi-natural experiment allowed us to show that the decrease in mobility due to the implementation of stringency measures is sensitive to citizens' knowledge about the severity of pandemic itself. We find that this effect is driven by countries with low quality of governance and low education. The intuition is that countries with low quality of institutions and low levels of education are less prone to comply with stringency index and more easily influenced by their individual level of knowledge. Our contribution suggests that coercive regulation such as pandemic related containment measures must be accompanied by an adequate communication effort.

In Chapter 2 we discuss how big firms in the energy sector reacted to a temporary surtax. This topic is of particular interest at the current moment as the idea of a temporary solidarity surcharge to pursuit economic recovery is being discussed. Moreover, the introduction of a temporary extra tax is also setting in because of the recent sharp increase in gas and electricity prices in Europe. Our main contribution is the investigation of the reaction of big firms to a temporary surtax, whereas the existing literature focuses on small and medium firms and on permanent tax increases. Our results show that the estimated effect of the surtax is null, and that the tax is neither shifted onto labor nor onto capital.

Finally, in Chapter 3 we discuss how the bidding zone configuration affects relevant market outcomes such as prices and cross-country and cross-zonal flows. We use these outcome variables as a measure of market efficiency because they reflect the appropriateness of price signals and the proper use of transmission capacities. Our contribution shows empirically that the nodal market design gives better market signals and improve the use of cross-country flows. We aim at further investigating the impact of the market re-configuration on liquidity and market structure.