Universidad Carlos III de Madrid

Master Degree in Economic Analysis

2022-2023

Master Thesis

"Measuring the Efficacy of Competition Policy: Identification using an RDD Approach"

Ali Musali

Raquel Carrasco Perea

Getafe, 2023

Abstract

The Hart-Scott-Rodino (HSR) Act plays a vital role in overseeing mergers in the United States, with its primary objective being to block anti-competitive deals that might hurt competition. In this study, we took a closer look at the existing notification threshold set by the HSR Act, specifically focusing on its impact on post-merger innovation around the $50 million mark. Diverging from previous studies that focused on the effects of amendments to the HSR Act, we precisely analyze the impact at the notification threshold itself, determining whether the existing threshold optimally fosters post-merger innovation. Contrary to the objective of antitrust policy, our research reveals that the current thresholds might actually be stifling innovation, rather than nurturing it. Using robust bias-corrected RDD estimation, we observe a significant decline of approximately 47% in post-merger patent intensity around the threshold, a sign that the premerger notification program might be inadvertently hindering innovation, particularly for entities that aren't heavy patent producers. This suggests that the notification program could be doing more harm than good, especially for smaller players in the market. The potential reasons behind this dip are discussed, pointing to the significant financial burdens and strategic uncertainties that come with the notification process.

Contents

Chapter 1

Introduction

Antitrust law in the U.S. started with Sherman Act in 1890 which was designed to prohibit anti-competitive acts and forbid monopolies. Today, competition policy in the U.S. is primarily regulated by the Federal Trade Commission (FTC) and the Department of Justice (DOJ). Starting from 1976, with the passing of the Hart-Scott-Rodino Act[1] (HSR Act from hereafter), all mergers and acquisitions that affect US commerce have to file a pre-merger notification form with the FTC and the DOJ and must wait for a certain period of time before the deal can be consummated. In 2001, the act was amended and deals with transaction size below $ 50 million were not required to notify authorities which led to a 70% reduction in notifications and resulted in an increase in anti-competitive mergers (Wollmann, 2019).

The amendment to the notification threshold was a result of an overwhelmingly high number and cost of processing notifications made to the authorities (Howell, 2002). However, how this amendment together with other regulatory changes would affect competition was not explicitly communicated by the FTC. As expected, the deals between competitors that had just become exempt from antitrust scrutiny increased in numbers compared to pre-amendment levels (Wollmann, 2019). In addition, Morzenti (2022) finds a reduction in post-merger innovation as a result of revised notification thresholds, again by comparing newly exempt deals with the ones before the amendment. These papers show that the modification of the thresholds led to a huge number of potentially anti-competitive deals to not face any pre-merger scrutiny and be consummated with little to no regulation.

The premerger notification itself is time-consuming and financially costly for the merging entities as they have to hire lawyers to prepare required documents, collaborate with economic consulting firms to prove to the authorities that their deal is not anti-competitive, pay the filing fees, and many

[1]Hart-Scott-Rodino Antitrust Improvements Act, Pub. L. No. 94-435, 90 Stat. 1383 (1976) (codified in 15 and 28 U.S.C. (2000)).

more. Therefore, the effect of a costly mandatory notification system on competition is ambiguous and has to be empirically measured. Despite the ambiguity, this analysis has not been done yet. This paper will analyze how premerger notification affects the post-merger innovation score of merging entities by comparing the deals that had to be notified as they were just above the notification thresholds with the deals that did not have to be notified as they were just below it. To put it differently, this paper will exploit the notification threshold of $ 50 million by using sharp Regression Discontinuity Design (RDD), to measure the average effect of notifying the authorities.

This paper will contribute to the not-yet vastly researched literature in various ways. First of all, this paper will employ sharp RDD, which is proven to be a very reliable method when the treatment rule is based on a cutoff, as an identification strategy to causally estimate the impact of premerger notification on post-merger innovation, where innovation measure is based on patenting activity. This, to the best of our knowledge, is the first paper that uses a sample of deals that took place only after the amendment till as recently as 2018, while previous papers have pooled the pre-amendment and post-amendment deals until 2005 together. We are also presenting an empirical analysis to the issue of notification thresholds being completely based on transaction sizes, company revenues, and net assets, and not on their innovative capacity, as the latter may not always be reflected in financial measures well.

Moreover, we are able to distinguish the impact of notification between intensive innovators with high records of patents and the companies that exhibit relatively lower patenting activity. By identifying the local average treatment effect of premerger notification, we are also effectively answering whether the notification threshold is optimal or not. Optimality in this context does not mean that an increased threshold enables anti-competitive deals that fall below the notification threshold to consummate without being scrutinized as this has already been shown by Wollmann (2019) and Morzenti (2022). It is rather related to the idea that the notification program should not have any effect, neither positive nor negative, on the post-merger innovation score of merging companies at the thresholds if the notification thresholds are indeed optimal.

We use SDC Platinum and PatentsView as data sources for M&A and patent activity, respectively. Using patent data, we generate 3 measures with different year windows around the deal date. Contrary to the objective of the antitrust policy and HSR Act, we find that for all the measures of innovation available to us, the ATE is negative for all year windows and is statistically significant in wider windows around the deal date. This suggests that antitrust scrutiny actually lowers the post-merger innovation score of reported deals at the notification threshold. To give an example, premerger notification actually reduces the long-term patent intensity, which is formally defined in Section 3.1, by 47% in the nonparametric robust estimation. Moreover, when we compare high patent producers with low patent producers, we find that this effect is larger in magnitude for low patent producers, suggesting that costly notification might make companies with relatively less patenting activity prioritize the R&D activity less than the ones with larger patent profile. These findings are in contrast with the objective of the program. They also indicate that although amendment to the HSR Act resulted in anti-competetive deals to consummate with no scrutiny as described in the aforementioned papers, the effect of costly notification at the newly established threshold is indeed negative and is more detrimental to deals with relatively less patenting activity.

Chapter 2

Data

There are 2 main databases used in the paper: Thomson Reuters SDC Platinum, the most used database in M&A analysis that also incorporates private deals or deals where the target is a private company, and PatentsView Database [1] for patenting activity, which also covers the patents of private companies together with public companies, government institutions, research laboratories, university departments, etc. Both of them are analyzed in detail below.

SDC Platinum has been the most important tool to analyze M&A activity as it covers private deals and has extensive coverage. As the amendment to notification thresholds was effective starting from the beginning of 2001, we retrieve the deals after that date till 2018. We retrieve the deals where both acquirer and target are U.S.-based companies or have significant operations in the U.S. For each deal, we can access the acquirer name, the target name, the name of the ultimate parent of the acquirer company, the main state of residence, the deal date, the value of the transaction, percent of shares being acquired in the transaction together with percent of shares owned after the transaction. We then apply a series of data cleaning based on the HSR Act reportability.

Firstly, the notification thresholds have been increased by the growth rate in Gross National Product starting from 2005. As the transaction value is reported at the time of the announcement of a deal, we deflate the deal value of those deals after 2005 by the growth rate in GNP, and we will refer to the notification threshold as \$ 50 million, but it should be noted that the transaction values are based on 2004 US dollars after they have been deflated. [2] Secondly, if an acquirer company previously earned a positive share of the target company and wants to acquire more of it, for the new deal to be HSR-reportable the size of the transaction is calculated according to the percent of shares held as a result of

[1] PatentsView was developed jointly by U.S. Patent & Trademark Office (USPTO) and other institutions as a platform to serve patenting data to students, researchers, and policy-makers.

[2] The official guide to HSR Act reportability, therefore, refers to this threshold as \$ 50 million (as adjusted).

the transaction, therefore aggregating all previously acquired shares. To accommodate this, we correct the transaction value of those deals by multiplying the deal value by the ratio of the percent of shares owned after the transaction to the percent of shares owned during the transaction, whenever they differ. Lastly, we exclude the deals that are exempt from the HSR Act irrespective of their deal size. These deals mostly involve oil and gas sector, real estate companies, investment entities, bankruptcy acquisitions, etc. We follow Wollmann (2019) to apply this correction.

Once the deflation adjustments and other corrections have been made, we need to choose a bandwidth around the threshold \$ 50 million (as adjusted). We choose \$ 15 million as the largest bandwidth to be used in the analysis, leaving us with deals with transaction value between \$ 35 million and \$ 65 million. This choice is arbitrary and smaller bandwidths are used later in the analysis to check the sensitivity of results to this bandwidth choice. After this, we are left with 4675 observations where each observation is a deal between an acquirer and a target company.

Afterward, we match the company names in our sample from SDC Platinum to assignee names in PatentsView. The simple string-matching algorithms alone do not deliver very accurate results as there are misspellings and different versions of the same company name in both samples. This problem is especially prominent in PatentsView as one company might have more than a few research departments registered in USPTO, making it hard to combine all the patents under one company name. To overcome the issue, we follow a similar pattern described in Mei (2019). We first match company names with simple and loose string-matching algorithm, allowing for various legal company name endings and common words, and we allow for multiple matches for each company in SDC Platinum to cover all possible profiles of a given company in PatentsView. Then, to confirm each match, we use Google Search API to search for the name being matched and alleged matches. If the matched name and alleged match have at least 2 exact search result in Google Search or share the same website domain excluding common financial, corporate or informative websites, only then it is considered a correct match.

Patents, just like academic papers, can be cited by other patents which we use to measure their effectiveness or innovative capacity. For each patent, we retrieve information about the date of application, the number of citations the patent received, and the technological field it has been assigned to. We choose the date of the application over the date of granting the patent as granting patents might take years, sometimes up to a decade. There are several technological field systems used by PatentsView. We pick CPC (Cooperative Patent Classification) as it is developed jointly by USPTO and European Patent Office (EPO), is hierarchical, has broad coverage, and offers a high level of detail[3]. For each patent, we pick the first assigned CPC (as they may have more than one assigned field), and we have 630 technological fields in total.

To be able to compare one patent with another, the use of citations may not be enough as some fields are more active than others. Moreover, it is generally observed that, the older the patent, the higher the citation, as one would expect. Therefore, to be able to compare patents across years and technological fields, we follow Lerner et al (2011) and define a citation intensity that is equal to the number of citations the patent received divided by the average number of citations in that technological in that year. We will refer to this measure as patent intensity from hereafter.

To assess the impact program on the innovation score of the merging entity, we choose the aggregate innovation as our variable of interest rather than differentiating between the research departments of the acquirer and target company. To this end, for each deal, we combine the patents of the acquirer and target company together with the patents of the ultimate parent of the acquirer company starting from 1995. This also allows for managerial and structural changes in R&D departments after the deal, since new patents could be assigned to any of the companies in the deal. Then we divide the patents into two parts: those before the deal date and those after it. After all of this, we are left with 1101 deals that produced patents in the 5-year window around the deal date, and 1782335 patents in total.

This paper measures innovation in three dimensions: patent intensity, patent citations and the number of patents. Patents are an imperfect measure of innovation due to high levels of heterogeneity in patent propensity and patent quality across different fields. Moreover, the citation dynamics of patents have changed in the last two decades as a large portion of patent citations have been produced by a few patents (Kuhn et al. 2020). This together with a direct relationship between the age of a patent and its citations makes the use of citations less suitable to measure innovation. The use of number of patents, or size from hereafter, is even less suitable for our analysis due to aforementioned heterogeneity in patenting activity. Patent intensity, on the other hand, corrects for the relationship between the age of patent and its citations and allows us to compare patents across different fields since it is a citation measure relative to the average patent in a given technological field and year. To see this more clearly we plot average citations and average intensity over the years. Figure 2.1 shows average citations and intensity starting from 1995. Average citation exhibits a steady decline while average intensity varies around 1. Nonetheless, all three measures are analyzed for the sake of completeness.

[3]This system has a higher granularity and is well depicted for almost all patents

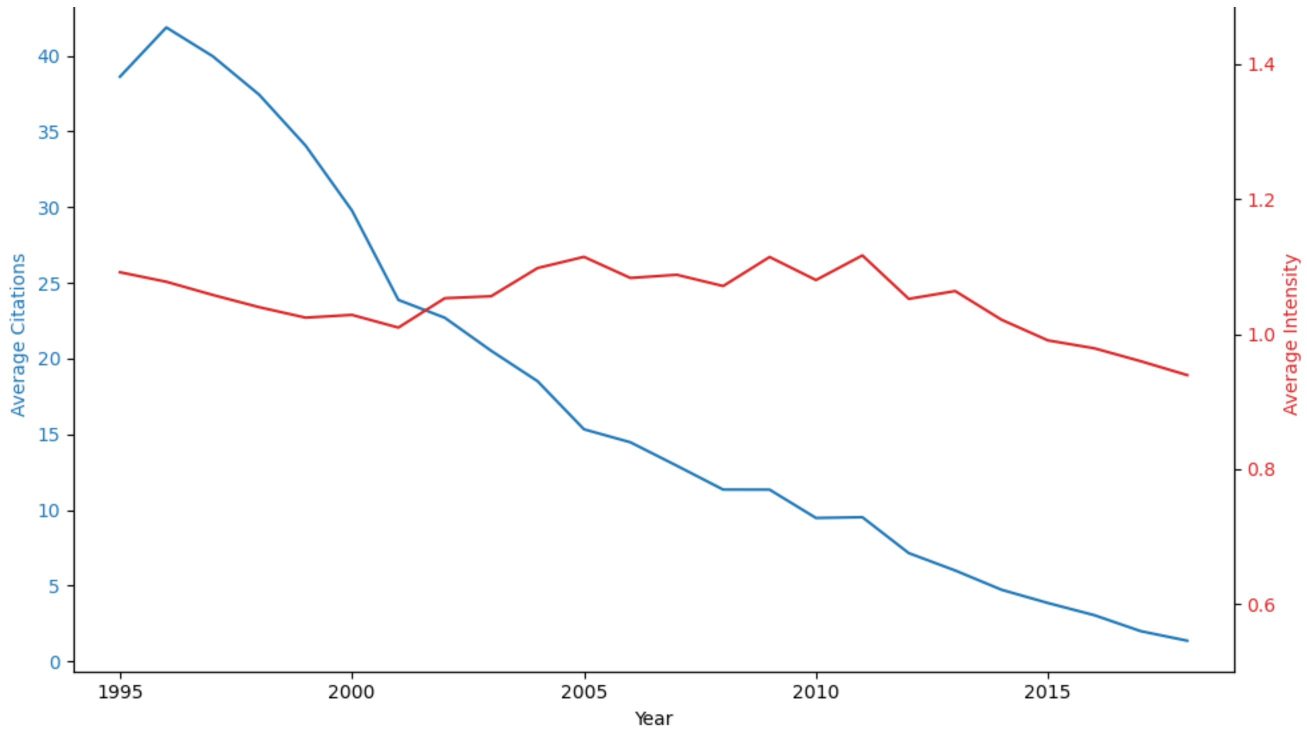

Figure 2.1: Average patent citations and average patent intensities from 1995 to 2018.

Chapter 3

Methodology & Identification

3.1 Innovation Measures

To measure the post-merger innovation score, we consider 5 years and 2 years as the maximum and minimum time frame around the deal date, respectively. We have 3 measures to be used and many possible year-window combinations. To be more robust and agnostic, we use all the possible year combinations. That is, for each variable we consider i years before and j years after the deal date, where $i, j \in \{2, 3, 4, 5\}$. This results in 16 year combinations for each of the three variables, amounting to 48 variables of interest in total. This also allows for measuring the short-term and long-term effects of a merger or acquisition and deeper analysis in general. Then, for each of these variables we take the difference between the natural logarithm of the average of the measure used after the deal date and the average of the same measure before the deal date. We define P_{dij}, C_{dij}, and S_{dij} for patent intensity, citation and size, respectively, for each deal d, for i years before and for j years after. More formally, we have the following definitions:

$$P_{dij} = \log\left(\frac{1}{N_{dj}} \sum_{\substack{k=1 \\ k \in After_{dj}}}^{N_{dj}} Intensity_k\right) - \log\left(\frac{1}{N_{di}} \sum_{\substack{k=1 \\ k \in Before_{di}}}^{N_{di}} Intensity_k\right)$$

where

$$After_{dj} = \{\text{Patents of deal } d \mid \text{Patent Date} \in [\text{Deal date, Deal date} + j \text{ years}]\},$$

$$Before_{di} = \{\text{Patents of deal } d \mid \text{Patent Date} \in [\text{Deal date} - i \text{ years, Deal date}]\},$$

and N_{dj} and N_{di} are the cardinality of $After_{dj}$ and $Before_{di}$, respectively. Similarly, we have for citations and size:

$$C_{dij} = \log\left(\frac{1}{N_{dj}} \sum_{\substack{k=1 \\ k \,\in\, After_{dj}}}^{N_{dj}} Citation_k\right) - \log\left(\frac{1}{N_{di}} \sum_{\substack{k=1 \\ k \,\in\, Before_{di}}}^{N_{di}} Citation_k\right),$$

$$S_{dij} = \log\left(\frac{N_{dj}}{N_{di}}\right).$$

P_{dij} (C_{dij}), as defined above, has an intuitive interpretation. It measures, approximately, the percentage change in the average patent intensity (citation) from i years before to j years after the deal date. Similarly, S_{dij} can simply be interpreted as the percentage change in the number of patents from i years before to j years after the deal date.

3.2 Framework & Assumptions

The identification strategy in this paper is based on exploiting the discontinuity in the probability of premerger notification at the notification threshold. As per the official guide for HSR Act reportability, the \$ 50 million threshold (as adjusted) is the first step to check whether a deal has to be reported in the Size of Transaction Test (SOT Test). If the deal is below this threshold, then it should not be reported, rendering the probability of notification zero for those deals below the threshold. If the deal is valued at more than \$ 50 million but less than \$ 200 million, however, a further test, called the Size of Person Test (SOP Test), is applied which requires filing if one of the parties has at least \$ 100 million (as adjusted) or more in net sales or total assets, and the other party has \$ 10 million or more. Unfortunately, SOP Test requires data on financial figures which are not available for private companies. As our sample contains both public and private companies, we assume that all the deals above the \$ 50 million threshold meet the requirements in the SOP Test, which makes the probability of notification 1 for the deals above the notification threshold. This consequently results in Sharp Regression Discontinuity Design as the main framework to be used in this analysis.

This assumption is also used in Morzenti (2022) to determine whether the companies had to file a notification report. One possible approach could be to use only the deals where both acquirer and target are public companies which would result in a significantly smaller dataset and selection bias. However, this assumption is not too restrictive. The amendment to the HSR Act in 2001 increased the size threshold from \$ 15 million to \$ 50 million, however, did not apply any change to the SOP Test thresholds that were proposed when the law was initially passed in 1976. These SOP Test thresholds of \$ 100 million and \$ 10 million have always sought to be too low and capture most of the deals if the size test is met (Howell, 2002). This can also be seen in the unexpectedly high number of reporting received in the first years of the program. In 1979, for example, 859 transactions were reported, and in 2000 this figure rose to 4926, which greatly exceeded Congress's target of capturing 150 biggest mergers.[1] Moreover, adjusting for inflation from 1976 to 2002, the original SOP Test threshold would have stood at \$ 311 million and \$ 31 million in 2002, but they remained in their original value in 1976 (at \$100 million and \$ 10 million), while size threshold (or SOT Test threshold after 2001) realized a 3-fold increase, compatible with the inflation (Howell, 2002). The adjustment to SOP Test thresholds only came around in 2005 when all the thresholds were to increase by the growth rate in GNP. These indicate that the SOP Test thresholds were too lax, low, and unrestrictive after the amendment in 2001. Therefore, assuming that the probability of notification is 1 for all the deals above the threshold is not too unrestrictive and should not be affecting the estimation significantly as most of the deals are probably reported.

One of the identifying assumptions of Sharp RDD is that the observations that are just below the threshold (as adjusted) are a suitable group of controls or they are comparable to the ones just above it. This is also referred to as the continuity of potential outcomes in the literature. Applied to this paper, we also add the comparability across the time dimension to this assumption. More formally, $\forall\, i, j \in \{2, 3, 4, 5\}$, we have the following:

$$\lim_{z \to 0^+} \mathbb{E}[Y_{dij} \mid Z_d = z] = \mathbb{E}[Y_{1dij} \mid Z_d = 0],$$
$$\lim_{z \to 0^-} \mathbb{E}[Y_{dij} \mid Z_d = z] = \mathbb{E}[Y_{0dij} \mid Z_d = 0],$$

where Y_{dij} is either one of the innovation measures, namely P_{dij}, C_{dij}, S_{dij}, and Y_{0dij} and Y_{1dij} are corresponding potential outcome notation, described in Rubin (1977). Z_d is the adjusted and centered Transaction Value for deal d, or simply the running variable. The emphasis is on "adjusted" as we deflate the transaction size of those deals after and including 2005 and adjust them to 2004 US dollars. As a result, two different deals may have the same running variable despite having different un-deflated transaction values. Despite having different un-deflated transaction values, the deals around the threshold are assumed to be comparable. The comparability across time means precisely this. This assumption is somewhat plausible as the size of the economy grows, so does the amount of innovation, lowering the chance of producing new innovation for relatively smaller firms as the boundaries of innovation expand with the size of the economy. This is similar to the rationale for adjusting the

[1] Workload Statistics Report FY 1978-1987 and Annual Report of FTC for FY 2000

reportability thresholds after 2005 to keep them more "relevant" (Howell, 2002). As this continuity assumption contains counterfactuals, it is not testable, as usually is the case in an application of RDD. Note that the same deflation has also been applied in Wollmann (2019).

Another identifying assumption in sharp RDD is the no-precise manipulation of the running variable, which is tested by checking the continuity of the probability density function of the running variable at the threshold. The ability to manipulate the running variable in this context would mean that when two parties agree on the deal value, they might agree on a value just below the threshold to avoid the costly notification while the true valuation methods indicate a higher number than the threshold. This, however, is illegal as per the wording in the "Deliberate Avoidance" section in the official Act[2]. It can be argued that if two parties do indeed agree to manipulate the running variable, then the FTC or DOJ would not be informed about the deal, hence they could not be caught. This argument, however, is fallacious as the authorities are not restricted to scrutinizing only reported deals and can examine any transaction, regardless of its size, that has been reported to be anti-competitive, possibly by their competitors, or that it just seems "suspicious". Not complying with this deliberate avoidance rule can be more costly to the parties than the notification itself. Nonetheless, for each regression or non-parametric estimation, we will be testing the continuity of the running variable using local-polynomial density estimators described in Cattaneo et al. (2017).

What parties can do legally, however, is agree on a number below the threshold while actual valuation methods show an even lower value. This, in the literature of M&A and corporate finance, is called acquisition premium and is mostly dependent on bargaining between the parties. Most deals usually have a premium, while some deals might get a discount. Then, for those deals where the valuation methods show a very close value to the threshold, but still smaller than the threshold, the parties may agree on a round number between the actual value and the threshold. One likely possibility is that they indeed agree on \$ 50 million (as adjusted) as the reporting is required for the deals with size strictly above the threshold. This also can be seen in a high number of observations at the threshold, while there is only a few right above and below it in our sample (26 observations at the threshold while only 3 in 0.1 bandwidth when we consider the intensity of 5 years before and 5 years after as a dependent variable, namely P_{55}). This is known as non-random heaping in the running variable in the literature and might induce a bias in the estimation. This can easily be solved by applying a "donut" estimator that entails dropping those observations at the threshold, or at heaps (Barreca et al., 2011). For our sample, these heaps happen at the multiples of 5, and dropping these observations together with the ones at the threshold gives us very similar results as when we only drop those at the threshold.

[2] Premerger Introductory Guide - Federal Trade Commission.

Moreover, due to the unavailability of financial data for private companies, we are not able to test the smoothness of relevant observables at the threshold, but it is assumed, together with the smoothness of unobservables at the threshold, for identification. It can be argued that even if we had data on private companies, the relevant observables that are potentially discontinuous at the threshold would most likely be used as a measure of innovation, rather than a threat to the identification strategy of this paper. One possible such variable would be post-merger R&D expenditure. If this variable is indeed not continuous at the threshold, then this would undermine the reliability RDD estimates presented in this paper. However, if this is indeed the case, then R&D expenditure would itself be a measure of innovation alongside the others, hence would not pose any threat to the results. More formally, we assume that if there is a relevant covariate that is discontinuous at the threshold, then it would necessarily be itself a measure of innovation, hence would not pose a threat to the identification.

Hence, under these assumptions, the Average Treatment Effect (ATE) for all $i, j \in \{2, 3, 4, 5\}$, defined as τ_{ij}, at the reporting threshold of a compulsory premerger notification for merging entities is identified as:

$$\tau_{ij} = \lim_{z \to 0^+} \mathbb{E}[Y_{dij} \mid Z_d = z] - \lim_{z \to 0^-} \mathbb{E}[Y_{dij} \mid Z_d = z]$$

where Y_{dij} and Z_d is a post-merger innovation measure and adjusted running variable, respectively, defined as before.

3.3 Estimation

Next, we turn to the estimation of ATE with various methods. To this end, we will use local linear and quadratic regression to estimate ATE at the threshold. In addition to this, we will also estimate the treatment effect with a standard non-parametric estimator that corrects for first-order bias in distributional assumption caused by bandwidth choice and gives bias-corrected estimation and improved confidence interval coverages. These are discussed in detail below. In all estimations, we use different kernels and bandwidths and add yearfixed effects as covariates. Adding year fixed effects does not change the identification strategy, but is aimed at improving efficiency, especially with citation as a measure as the change in average citation from before to after might be different in deals across years. This can be seen in Figure 2.1 where the "slope"[3] of average citation curve is different across years. Moreover, in every specification, we drop the observations with the running variable being exactly at the threshold, hence applying "Donut" RDD estimation. In addition to estimation with full data, we will also estimate ATE when we condition the maximum number of patents in the 5-year window

[3]Slope is an incorrect term here as the curve in Figure 1 is not a continuous function, but rather is drawn by interpolation between consecutive years

around the deal date on a shrinking set of values from 120 to 50, to be able to distinguish the impact of costly notification between high-patent producers and low-patent producers.

Parametric estimation in this application refers to assuming a shape for the conditional expectation function of the innovation on the adjusted transaction value for a given bandwidth. One of the most used methods in applied RDD literature is Local Linear Regression, hence assuming linearity between the dependent variable and running variable. This method is especially powerful and has a small bias when we take a small bandwidth around the threshold (Hahn et al., 2001). One can use high-order polynomials and a more flexible bandwidth to estimate ATE. Gelman and Imbens (2019), however, argue that polynomials of degree higher than 2 might be a bad approach as it results in noisy estimates, high sensitivity to the degree chosen, and bad coverage for CI's, instead, suggest using estimators based on local linear or quadratic polynomials. We will be assuming a quadratic relationship with high bandwidths and will be switching to local linear regression when smaller bandwidths are used. In all quadratic specifications, we will report estimation with a triangle kernel, one that gives higher weights to the observations closer to the threshold and weights decline linearly to 0 at the bandwidth, and in local linear regression, we will use Epanechnikov kernel that is smooth and gives more weight to the observations further away from the cutoff than a triangular kernel. We also interact the dummy for being above the threshold with linear and quadratic terms, allowing ATE to vary across the running variable. Hence, $\forall i, j \in \{2, 3, 4, 5\}$, our baseline model is the following:

$$Y_{dij} = \beta_{0ij} + \beta_{1ij} Z_d + \beta_{2ij} Z_d^2 + \beta_{3ij} T_d \times Z_d + \beta_{4ij} T_d \times Z_d^2 + \theta_{ij} T_d + \sum_{t\,=\,2001}^{2017} \gamma_{tij} \times Year_{td} + \epsilon_{dij} \tag{1}$$

where Y_{dij} either one of the three measures, namely P_{dij},C_{dij} or S_{dij}, Z_d is the adjusted and centered transaction value, $T_d = \mathbb{I}(Z_d > 0)$, and finally $Year_{td}$ is year dummy that takes the value 1 only if the deal d occurred in year t, 0 otherwise. The coefficient of interest, or ATE at the threshold, is θ_{ij}. Note that local linear regression is equivalent to setting $\beta_{2ij} = \beta_{4ij} = 0$ in the above equation.

In principle, the running variable in this application is a continuous variable. However, due to the predisposition of humans to round numbers, we end up having a high number of observations at integer values, and even more at the multiples of 5, than fractional values. Although its primary target is discrete running variables, Card & Lee (2008) corrections may help in this situation, which suggests using clustered standard errors for inference with clusters on the running variable. Kolesar and Rothe (2018) argue that confidence intervals with clustered SE's might have poor coverage, however, they also show that when the number of clusters grows with the sample size, which is the case in this application, then clustering on the running variable improves CI coverage over conventional heteroskedasticity-robust SE's. We will use clustered SE's for inference, however,it should be noted that the results are very similar when we use usual heteroskedasticity-robust SE's due to high number of clusters (around 60% of the sample size on average).

In addition to weighted least squares estimation, we will also use "rdrobust" command that fits a different local polynomial regression to each side of the threshold with different bandwidths and corrects for first-order bias caused by bandwidth choice (Calonico et al., 2017). The authors (2014) argue that conventional Mean Squared Error optimal bandwidth choice, such as the one described in Imbens and Kalyanamaran (2012), delivers too "large" bandwidth choice since it does not satisfy the bias condition $nh_n^5 \rightarrow 0$, where h_n is the bandwidth choice for a sample size of n. To present a more robust inference, we will choose the bandwidth selection method that is Coverage Error Rate optimal (CER optimal) for both sides separately and report bias-corrected estimation with robust standard errors, as this is the preferred method for robust inference (Calonico et al., 2014). The variance-covariance estimation is done by the heteroskedasticity-robust Nearest Neighbor method with a minimum number of 3 neighbors, described as in Abadie and Imbens (2006). For all specifications, linear regression is used for estimation and quadratic polynomial regression is used for bias correction, for each side separately. In all estimations, we use an Epanechnikov kernel and year dummies as covariates.

In addition to estimating ATE with full data, we will also restrict our sample by bounding above the number of patents in 5-year window around the deal date, we will estimate the effect with upper bounds decreasing from 120 to 50 with a decrement of 1. With this method, we are able to differentiate the impact of the notification between intensive patent producers and relatively less active patent producers. Note that the same objective can be achieved by setting a minimum limit for the number of patents in 5-year window around the deal date, and then by increasing this limit. We do not report this procedure as this yields very similar results.

Chapter 4

Results

We then turn to report results. First, we will describe some sample statistics, then will move to the results of regression estimation. For all measures, we report results of both Equation 1 and so-called non-parametric estimation via "rdrobust" [1] command that gives bias-corrected estimation and robust inference.

4.1 Sample Statistics

We begin to introduce some sample statistics. In Table 4.1, we present mean, median, standard deviation, 10^{th} and 90^{th} percentile together with the sample size of two measures, namely patent intensity and patent citation. We omit size as a measure for its redundancy for the discussion. We see that the means of all the variables are negative, which means a merger or acquisition, on average, leads to a decline in innovation. However, some deals may get lucky and significantly improve innovation. This can be seen from the 90^{th} percentile of every variable as they are all positive. Another important fact that stands out is that roughly all the means and medians go down as we consider a longer time window around the deal date. The percentiles in citation measure go down too, however, this is not the case for intensity. The reason why the distribution, so to speak, of the citation measures shifts to the left is also discussed in Chapter 2 and in Figure 2.1, as average citations go down across years. Since this trend does not exist in average intensity, the shift is only in means and modes for this measure.

[1] Interested reader is referred to Calonico et al.(2017) paper regarding the technicalities of the command and references therein.

	Mean	Median	Std Dev	10^{th} Percentile	90^{th} Percentile	Sample Size
Intensity						
Years Before - Years After						
2 - 2	-0.07	-0.04	0.93	-1.11	0.89	769
2 - 3	-0.09	-0.04	0.95	-1.13	0.88	805
2 - 4	-0.07	-0.05	0.97	-1.12	1.05	828
2 - 5	-0.07	-0.04	0.96	-1.15	1.01	836
3 - 2	-0.09	-0.05	0.93	-1.16	0.81	800
3 - 3	-0.11	-0.06	0.95	-1.15	0.84	841
3 - 4	-0.10	-0.07	0.96	-1.18	1.00	866
3 - 5	-0.10	-0.06	0.96	-1.16	0.96	875
4 - 2	-0.10	-0.06	0.92	-1.16	0.88	824
4 - 3	-0.11	-0.07	0.95	-1.22	0.88	867
4 - 4	-0.11	-0.07	0.96	-1.23	0.94	892
4 - 5	-0.10	-0.07	0.96	-1.19	0.94	901
5 - 2	-0.13	-0.09	0.92	-1.27	0.84	833
5 - 3	-0.15	-0.10	0.94	-1.29	0.83	876
5 - 4	-0.14	-0.09	0.94	-1.31	0.94	902
5 - 5	-0.14	-0.07	0.94	-1.25	0.92	911
Citation						
2 - 2	-0.41	-0.34	0.97	-1.60	0.59	803
2 - 3	-0.50	-0.45	0.98	-1.64	0.55	843
2 - 4	-0.56	-0.51	0.99	-1.78	0.50	863
2 - 5	-0.61	-0.56	0.97	-1.79	0.49	870
3 - 2	-0.51	-0.44	0.98	-1.73	0.48	831
3 - 3	-0.60	-0.53	0.98	-1.81	0.35	872
3 - 4	-0.66	-0.58	0.98	-1.88	0.38	893
3 - 5	-0.71	-0.64	0.96	-1.91	0.33	902
4 - 2	-0.59	-0.54	1.00	-1.80	0.47	854
4 - 3	-0.67	-0.61	1.00	-1.90	0.38	897
4 - 4	-0.74	-0.66	1.00	-1.99	0.35	918
4 - 5	-0.79	-0.71	0.98	-2.03	0.29	927
5 - 2	-0.69	-0.61	1.03	-1.96	0.40	864
5 - 3	-0.77	-0.69	1.03	-2.02	0.34	907
5 - 4	-0.84	-0.74	1.02	-2.14	0.28	929
5 - 5	-0.89	-0.79	1.01	-2.17	0.20	938

Table 4.1: Sample statistics of Patent Intensity and Citation Count (P_{ij} and C_{ij} where i and j is the number of years considered before and after the deal date, respectively).

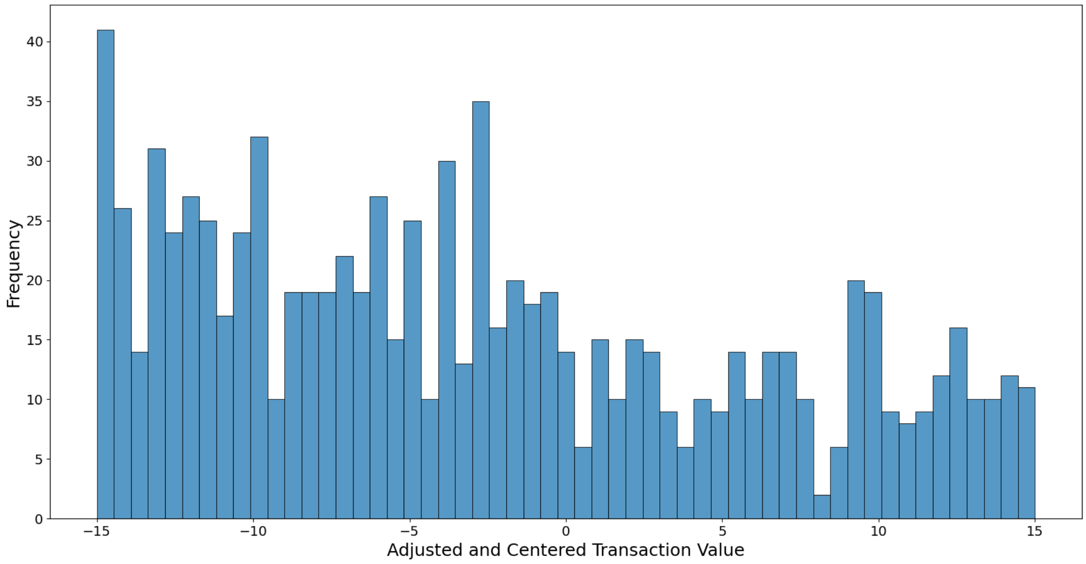

Figure 4.1: A frequency histogram of the adjusted and centered transaction value when the dependent variable is P_{55}.

Figure 4.1 is the frequency histogram of the adjusted and centered running variable when the considered dependent variable is P_{55}. The observations exactly at the threshold are dropped as per the "Donut" RD estimation. The heaps are at the multiples of 5 and are mostly prominent to the left of the cutoff. For the results reported here, we only consider dropping the observations at the threshold, but it should be noted that the results do not change much if the observations at the multiples of 5 dropped too. Moreover, the frequency decreases with the transaction value, in general, as expected. As one of the identifying assumptions, the continuity of density of the running variable at the threshold is tested, for every estimation, with local-polynomial density estimators described in Cattaneo et al. (2017), and for all estimations, we fail to reject the null hypothesis of continuity at 5% significance level.

4.2 Full Sample ATE

We then turn to estimate the ATE at the thresholds with a full sample. For each measure, we report 16 estimated discontinuities. To better visualize these estimates, we create a 4×4 matrix for each measure where the rows and columns are considered years before and after the deal, respectively. For example, the entry in row i and column j for the patent intensity matrix corresponds to the estimated ATE at the threshold where the dependent variable is P_{ij}. Each matrix is also supplied with a heatmap to give better readability and comparison across entries. We report estimates from Equation 1 first

Figure 4.2: Regression Discontinuity plot for post-merger average patent intensity, P_{55}.

Notes: The figure is plotted with "rdplot" command developed by Calonico et al. (2014). A global quadratic relationship is assumed. The covariates are year dummies and weights are given by a triangular kernel where the bandwidth is 15. The bins are constructed with mimicking variance evenly-spaced method using spacings estimators described as in the aforementioned paper.

with no restrictions and then with linearity assumption. Later, we report the estimates from the non-parametric RD-Robust method.

4.2.1 Quadratic Regression

As we have 48 variables in total, we are unable to report regression discontinuity plots for all of them, hence we only report it for patent intensity in the long term, namely P_{55}. Figure 4.2 depicts this. We can see that the discontinuity is negative, meaning that ATE at the threshold is negative. This is surprising as it means that the notification to authorities actually lowers post-merger innovation, contrary to the main objective of the antitrust policy. This suggests that the cost of notifying the authorities may indeed exceed the benefits of a blocked anti-competitive deal in the long term. Next, we turn to numerical results.

Figure 4.3 shows the estimated discontinuities from Equation 1 for patent intensity. There are several important results worth discussing. First of all, all the estimated coefficients are negative, meaning that ATE is negative at the threshold for all year combinations. Moreover, for each row, the coefficients decline if we consider longer duration, hence for a given row, roughly all the coefficients decline along the columns[2]. To put it differently, for a given year before the deal, ATE decreases when

[2]This can also be seen with the help of a color bar in Figure 4.3 as for a given row, the shaded colors get "colder" if we move from left to right, or if we consider a longer duration after the deal

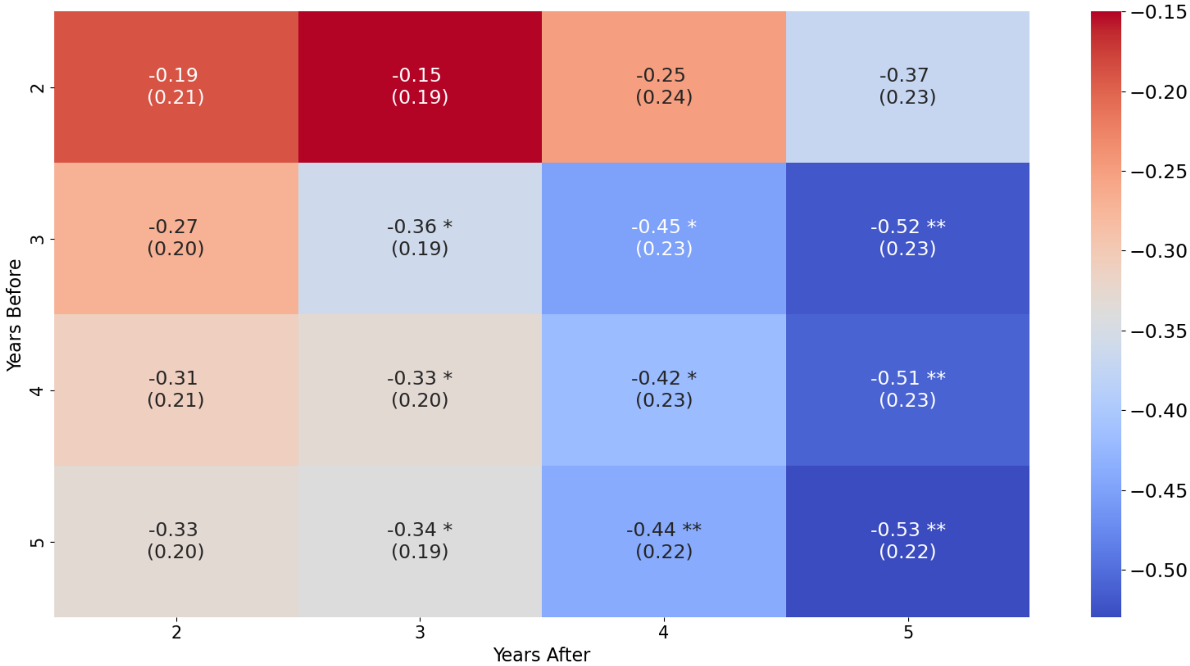

Figure 4.3: Estimated discontinuity in post-merger average patent intensity, P_{ij}.

Notes: The entry in row i and column j is the coefficient of θ_{ij} (estimated discontinuity) in Equation 1 where the dependent variable is P_{ij} and we use "Donut" RD estimation with a triangular kernel. The bandwidth is 15. The observations with the absolute value of the dependent variable higher than 3.5 are dropped. The standard errors are clustered at the running variable level and reported inside parentheses. (* ~ P-value < 10%, ** ~ P-value < 5%, *** ~ P-value < 1%)

we consider a higher number of years after the deal. More importantly, the coefficients become statistically significant if we consider a longer duration after the deal for a given year before the deal. As an example, the numerical estimate for the discontinuity for the RD-plot of P_{55} in Figure 4.2 is estimated to be -0.53 log-points, or 41% reduction in average patent intensity in the long-term[3], which is quite large when we consider that the unconditional mean of P_{55} is -0.14 log-points, or 13% reduction, and it is statistically significant at 5% level. These findings suggest that post-merger average intensity decreases for the notified deals that were just above the notification threshold, contrary to the expected, and this reduction is even more pronounced in the long term.

Next, we use citation as a measure of innovation. Figure 4.4 shows the estimated discontinuities from Equation 1. It is also noticeable that all the coefficients are negative and they become statistically significant in the long term. We do not see the trend as in Figure 4.3 and the most probable reason for this is that we are considering a given time window around a deal date for the deals in different technological fields. Even if we add year dummies, this does not control for the relationship between

[3]Long-term refers to, depending on context, either higher number of years after the deal date for a given year before it, or higher number of years for both before and after a deal date. In this case, the meaning is the latter as we are referring to the widest window around the deal date, P_{55}

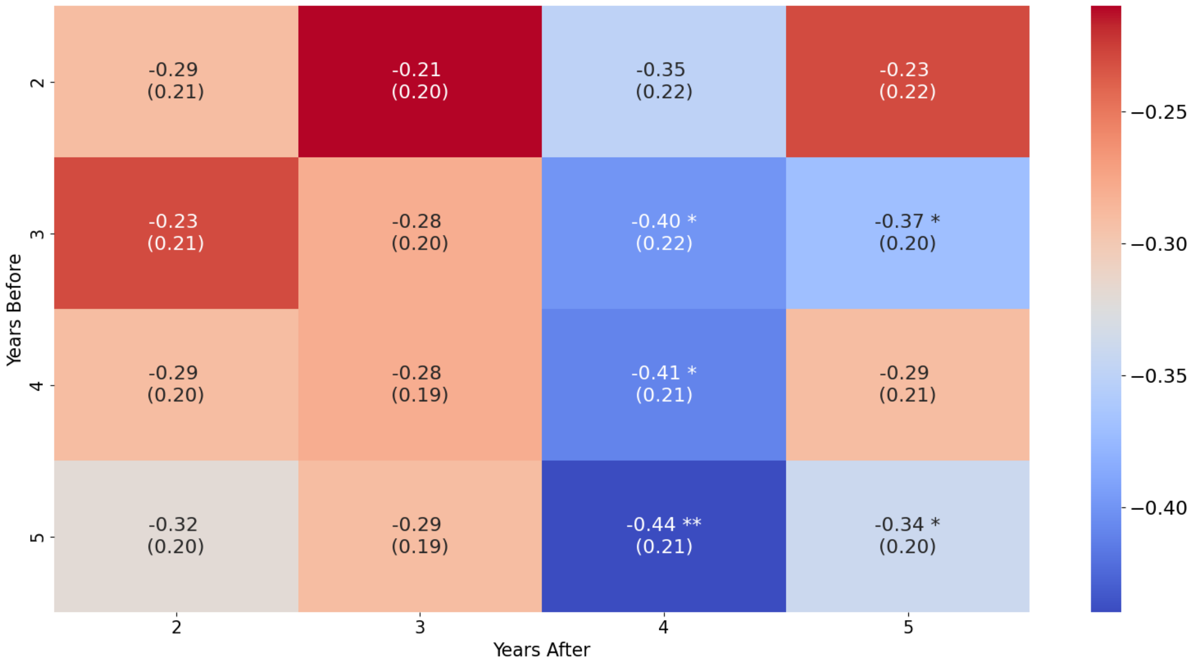

Figure 4.4: Estimated discontinuity in post-merger average patent citation, C_{ij}.

Notes: The entry in row i and column j is the coefficient of θ_{ij} (estimated discontinuity) in Equation 1 where the dependent variable is C_{ij} and we use "Donut" RD estimation with a triangular kernel. The bandwidth is 15. The observations with the absolute value of the dependent variable higher than 3.5 are dropped. The standard errors are clustered at the running variable level and reported inside parentheses. (* ~ P-value < 10%, ** ~ P-value < 5%, *** ~ P-value < 1%)

technological fields and the citation counts, as some fields are more intensive than others. However, this is controlled in patent intensity as a measure, hence will be the main measure to be used in the next parts of the paper, especially for and for sub-Section 4.2.3 and Section 4.3. Nonetheless, we get some statistically significant estimates. For example, the estimated discontinuity in C_{54} is -0.44 log-points, or 36% reduction post-merger citation count, and it is statistically significant at 5% level.

We now consider size as a measure of innovation. The size as a measure is used here and in the following subsections and will not be reported in Section 4.3. The reason for this is that size is the most imprecise measure of innovation out of all the three used in this paper, and it does not account for technological field effects. The size itself is only a quantitative measure and has no qualitative aspects of patents inside it. For example, the size of the patents for a given deal might decrease significantly after the deal, however, their intensity scores or citation counts might increase after the deal which could be the result of adapting a policy that prioritizes quality over quantity. Nonetheless, we report it for the sake of completeness. Figure 4.5 shows the estimated discontinuities for size as a measure. We can see that all the coefficients are negative, however, none is statistically significant and there is no visible trend across year combinations. As we will see in Subsection 4.2.2, there are visible trends and

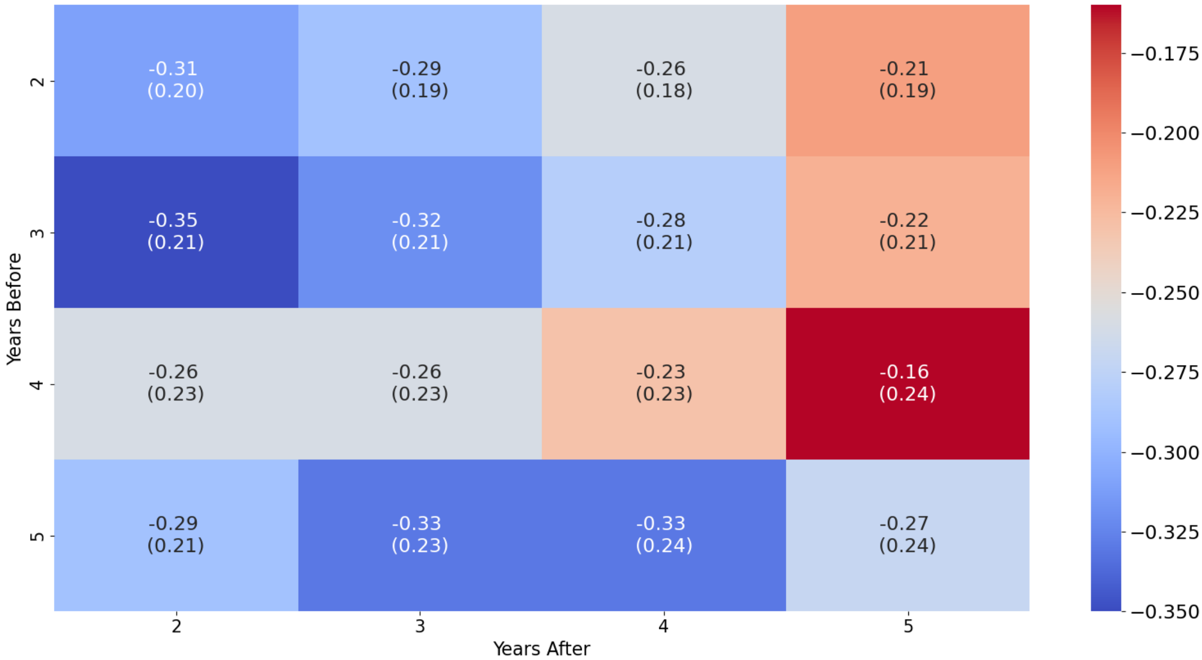

Figure 4.5: Estimated discontinuity in the post-merger patent counts, S_{ij}.

Notes: The entry in row i and column j is the coefficient of θ_{ij} (estimated discontinuity) in Equation 1 where the dependent variable is S_{ij} and we use "Donut" RD estimation with a triangular kernel. The bandwidth is 15. The observations with the absolute value of the dependent variable higher than 3.5 are dropped. The standard errors are clustered at the running variable level and reported inside parentheses. (* ~ P-value < 10%, ** ~ P-value < 5%, *** ~ P-value < 1%)

statistically significant estimates if we consider a smaller bandwidth and assume linearity.

To sum up, we reported the estimated discontinuities, θ_{ij}, from Equation 1 for all three measures assuming a global quadratic relationship between the dependent variable and running variable. In all estimations, the bandwidth was 15 and we weighted the observations with weights given by a triangular kernel and used weighted least squares for estimation. The observations at the threshold were dropped, resulting in "Donut" RD estimation. In the following subsection, we consider a smaller bandwidth and assume linearity between the measure used and the running variable.

4.2.2 Local Linear Regression

In this subsection, we assume linearity between the measure used and the running variable. To this end, we impose $\beta_{2ij} = \beta_{4ij} = 0$ in Equation 1. The choice of bandwidth is an important aspect of local linear regression as the results are usually sensitive to it, and there are several methods developed to choose it optimally, such as the one described in Imbens and Kalyanaraman (2012) and Calonico et al. (2014), depending on the objective of the researcher. Using a significantly smaller bandwidth leads to improved bias at the cost of increasing variance since we use fewer observations. Since we are going to

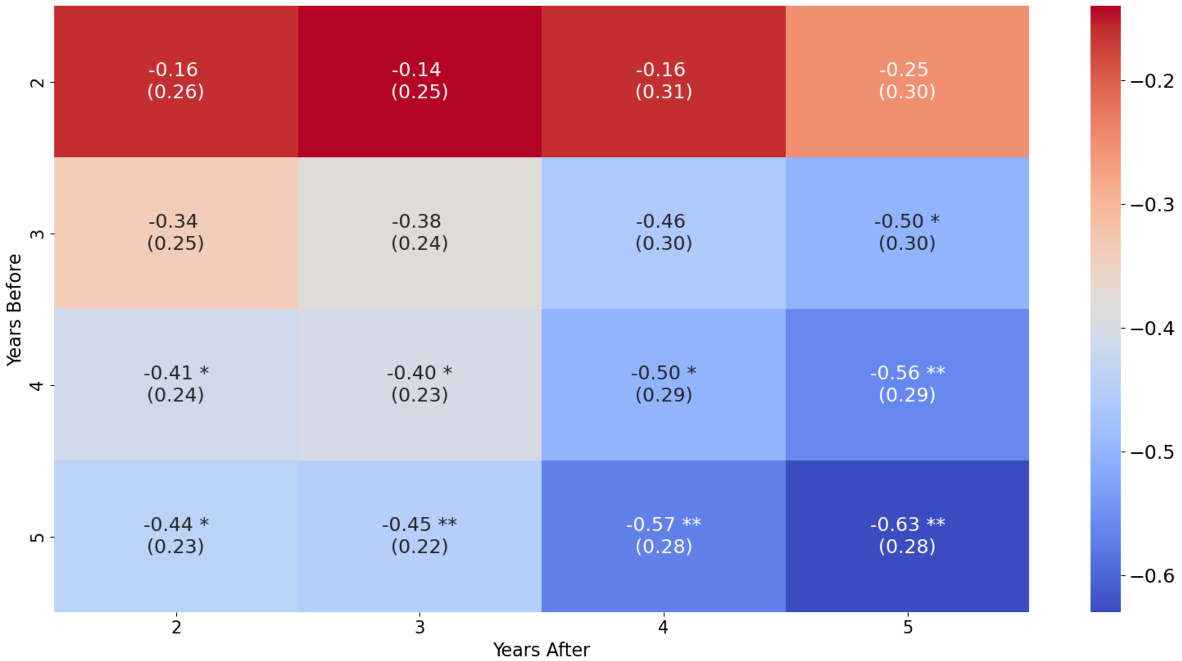

Figure 4.6: Estimated discontinuity in post-merger average patent intensity, P_{ij}.

Notes: The entry in row i and column j is the coefficient of θ_{ij} (estimated discontinuity) in Equation 1 where the dependent variable is P_{ij} and we use "Donut" RD estimation with an Epanechnikov kernel. Linearity is assumed, hence $\beta_{2ij} = \beta_{4ij} = 0$ is imposed on Equation 1. The bandwidth is 4 for both sides. The observations with an absolute value of the dependent variable higher than 3 are dropped. The standard errors are clustered at the running variable level and reported inside parentheses. (* ~ P-value < 10%, ** ~ P-value < 5%, *** ~ P-value < 1%)

use bias-corrected point estimates with CER optimal bandwidths in the following subsection, we use the same bandwidth for both sides of the cutoff and report results for this choice in this subsection. As we will see later the CER optimal bandwidths vary around 2-4 depending on the specification, hence we choose 4 as a bandwidth. All of the estimations are done with weighted least squares with weights given by an Epanechnikov kernel[4] which gives comparatively higher weights to the observations closer to the boundary of bandwidth than a triangular kernel. We choose this kernel as we are considering a smaller bandwidth, hence would like to give even more weight to the observations further away from the cutoff. The results are very similar if we use a uniform kernel or a different bandwidth (3, 5, 7 for example), but it is not reported here for the sake of brevity.

Figure 4.6 reports the estimated discontinuities from local linear regression for patent intensity as a measure. The trend described before exists in this estimation as well, that is, the long-term coefficients larger in magnitude and statistically more significant. To give an example, the estimated ATE when we consider P_{55} as a measure is -0.63 log-points, or 46% reduction, which is very close to what we found in the quadratic regression (41% reduction), and it is statistically significant at 5% level. This trend

[4]The weight of observation with a running variable value equal to z is $\frac{3}{4}(1 - (\frac{z}{h})^2)$ where h is the bandwidth choice which is 4 in this subsection

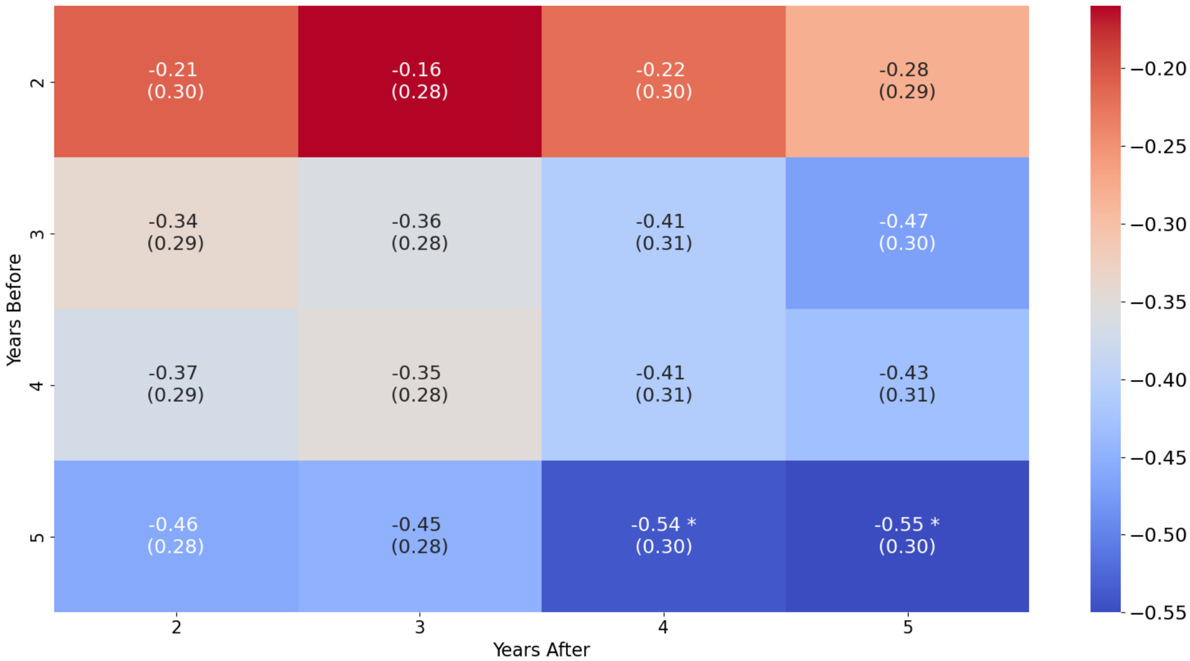

Figure 4.7: Estimated discontinuity in post-merger average patent citation, C_{ij}.

Notes: The entry in row i and column j is the coefficient of θ_{ij} (estimated discontinuity) in Equation 1 where the dependent variable is C_{ij} and we use "Donut" RD estimation with an Epanechnikov kernel. Linearity is assumed, hence $\beta_{2ij} = \beta_{4ij} = 0$ is imposed on Equation 1. The bandwidth is 4 for both sides. The observations with a value of the dependent variable higher than 3.5 are dropped. The standard errors are clustered at the running variable level and reported inside parentheses. (* ~ P-value < 10%, ** ~ P-value < 5%, *** ~ P-value < 1%)

of being large and significant in the long term suggests the idea that the effect of costly notification is more detrimental to innovation in the longer term.

Figure 4.7 shows the estimated ATE's from local linear regression with average citation as a measure of innovation. Even though there is no visible trend, the coefficients in the widest window are bigger in absolute value and statistically significant at 10% level. The discrepancy between the results of intensity and citation for both quadratic and linear regressions suggests that the effects of technological fields are prominent in citation counts and have to be accounted for. Hence, in the following subsection, we only use patent intensity as a measure to be estimated via "rdrobust" command.

We now report the results of size as a measure from local linear regression in Figure 4.8. One thing worth noting again is that all coefficients are negative, and almost all of them are statistically significant at the 5% level. Moreover, for a given column, the coefficients become larger in magnitude across rows. That is, for a given duration after the deal date, ATE of notification on the percentage change in patent size is more intensified if we consider a longer duration before the deal date. Moreover, for a given row, the coefficients become smaller in magnitude across columns. These indicate that, at the notification threshold, the costly notification lowers the number of patents of the merged entity for all

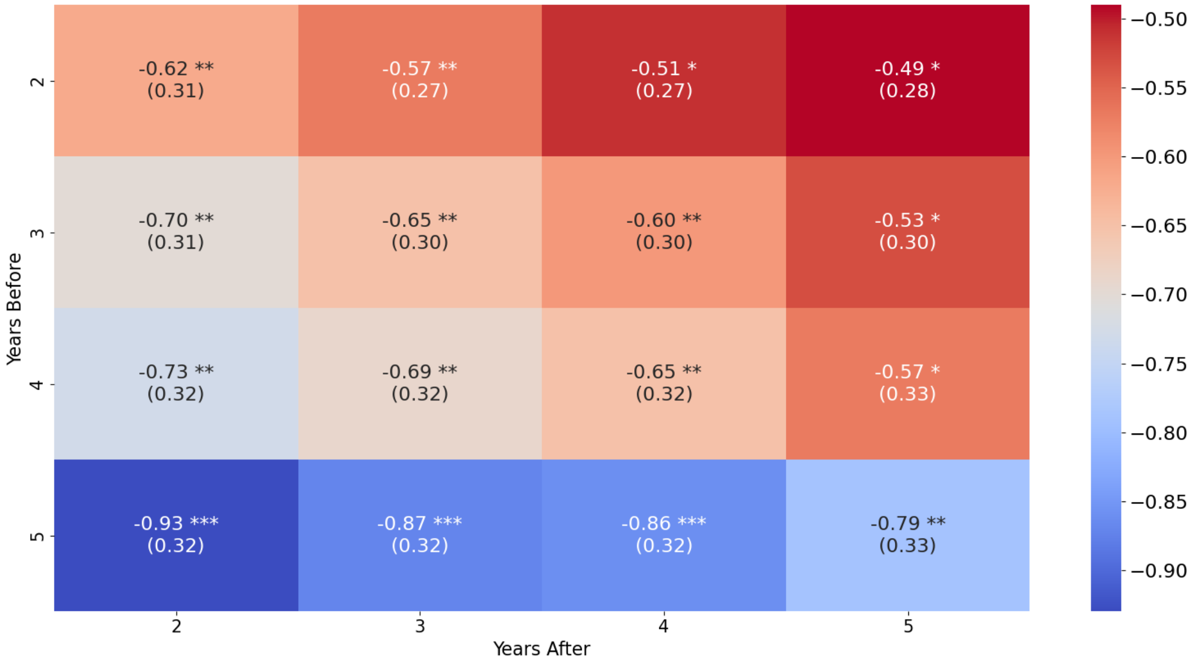

Figure 4.8: Estimated discontinuity in post-merger patent count, S_{ij}.

Notes: The entry in row i and column j is the coefficient of θ_{ij} (estimated discontinuity) in Equation 1 where the dependent variable is S_{ij} and we use "Donut" RD estimation with an Epanechnikov kernel. Linearity is assumed, hence $\beta_{2ij} = \beta_{4ij} = 0$ is imposed on Equation 1. The bandwidth is 4 for both sides. The observations with a value of the dependent variable higher than 3.5 are dropped. The standard errors are clustered at the running variable level and reported inside parentheses. (* ~ P-value < 10%, ** ~ P-value < 5%, *** ~ P-value < 1%)

year combinations, and this effect is more prominent right after the deal than 5 years after the deal, for example. As an example, the estimated discontinuity for S_{22} is -0.62 log-points, or 46% reduction, whereas for S_{25} it is -0.49 log-points, or 38% reduction.

To summarize, we reported the estimates from local linear regression with a bandwidth of 4 and weights given by an Epanechnikov kernel. The results for patent intensity are similar to the results of quadratic regression, that is the ATE is negative in all cases, and it is larger in magnitude and statistically significant in the long term. Even though the results are larger in magnitude than quadratic regression for citation, they are not statistically significant at the 5% level and there is no visible trend across years. The results of size, however, is drastically different than before. We find that the ATE is negative statistically significant and is larger in magnitude in the shorter term. This is probably because the entities in reported deals have considerably higher expenditures in other matters after the deals, hence they devote a smaller budget to R&D which results in a lower number of patents right after the deal, compared to the unreported deals. Once again, as the number of patents is only a quantitative measure, we refrain from using causal language relating the notification and innovation.

4.2.3 RD-Robust Estimation

In this subsection, we estimate ATE with the so-called non-parametric approach described in Calonico et al. (2014). For this method, we only estimate ATE for patent intensity as it is a more precise measure of innovation than citation count or number of patent size because it accounts for technological field and year effects. We drop the observations with an absolute value of the dependent variable higher than 3 as these are outliers and a linear regression and CER-optimal bandwidth choice is quite sensitive to the outliers. This was also done for local linear regression of patent intensity in the previous subsection. Moreover, as all the dependent variables are in natural logarithms, a value of 3 (-3) would mean a 20-fold increase (decrease) which is quite unrealistic and is probably caused by not correctly capturing the patents of a certain deal[5] or there are very few patents either before the deal and after the deal which results in extreme levels of change in patent intensity. While doing this, we only drop between 9 and 18 observations depending on the year combination, and upon further examination of these data points, we find that almost all of them have a very high number of patents after (before) the deal compared to before (after), which suggests that not all patents have been correctly captured for these deals or the deal date is on the boundaries of their patenting activity. Another way of correcting for this would be weighting deals with weights given by an inverse of the minimum between the number of patents before and after the deal. However, this weighting would be additional to kernel weighting which would result in not-so-easily interpretable results. Since we only drop a handful of observations, this does not harm the causal relation significantly.

The estimations are done with "rdrobust" command. It finds a bandwidth that is CER-optimal for each side of the cutoff for estimation, and a higher auxiliary bandwidth to be used for bias correction. For each side of the cutoff, the former is used in local linear regression for point estimation while the latter is used in quadratic regression for bias correction. Calonico et al. (2019) argue that bias-corrected estimation with MSE-optimal bandwidth gives suboptimal confidence intervals in terms of coverage error. They develop a method to calculate the main and auxiliary bandwidth that is CER-optimal, in the sense that it has the fastest decaying coverage error rate. For robust inference purposes, we use these CER-optimal bandwidths as they minimize the coverage error rate of the bias-corrected confidence intervals. These bandwidths take the form of $h_{CER,n} = n^{\frac{-p}{(3+p)(3+2p)}} \times h_{MSE,n}$, where $h_{MSE,n}$ refers to the MSE-optimal bandwidth for a sample size of n and p is the degree of polynomial for estimation, which is 1 in our case. We use Epanechnikov kernel for weigths as explained in the previous subsection, covariates are year dummies, and once again we drop the observations at

[5]This can happen if the name of the newly created entity after the deal is different than the names of the merging parties. Unfortunately, these name changes are not recorded in PatentsView and it is computationally costly to check whether every deal had a name change or not, which results in not correctly capturing the complete patent profile of some deals.

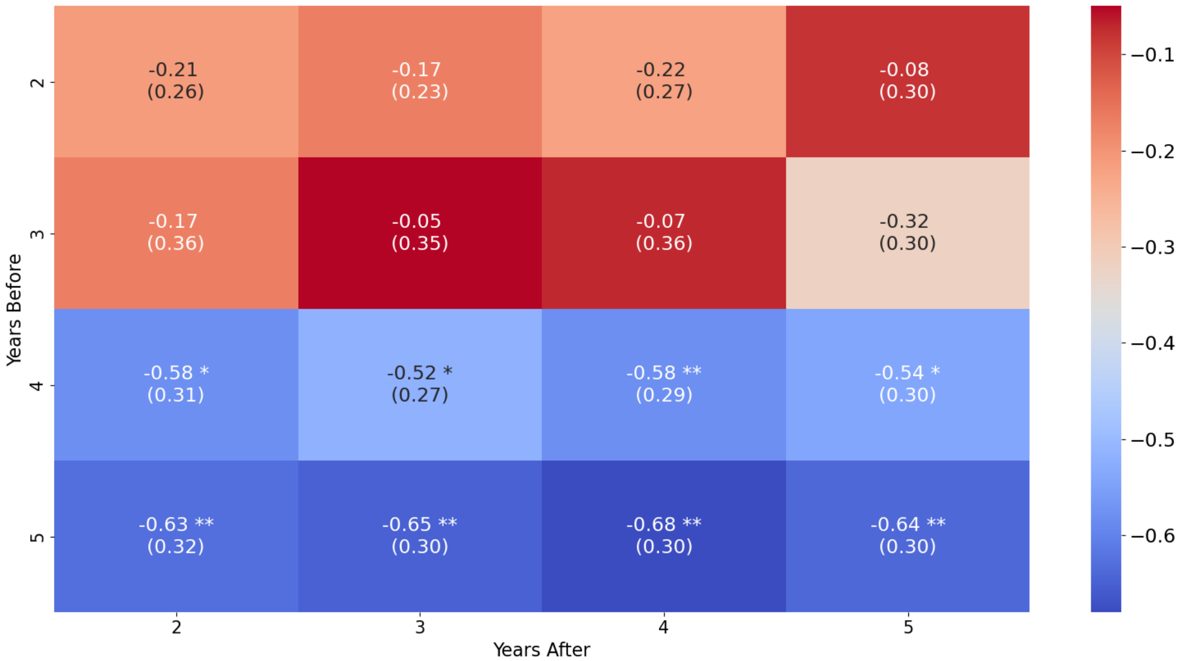

Figure 4.9: Estimated discontinuity in post-merger average patent intensity, P_{ij}.

Notes: The entry in row i and column j is the bias-corrected ATE for patent intensity estimated via user-written "rdrobust" command with Epanechnikov kernel. Linear regression is used for point estimation, while quadratic regression is used for bias correction. The bandwidth choice is CER-optimal and varies between $1.5 - 2.5$ for the left side, whereas it is between $2.5 - 3.5$ for the right side of the cutoff. The observations with an absolute value of the dependent variable higher than 3 are dropped. The standard errors are calculated with NN method with a minimum number of 3 neighbors and reported inside parentheses. (* ~ P-value < 10%, ** ~ P-value < 5%, *** ~ P-value < 1%)

the threshold. The variance-covariance estimation is done with the heteroskedasticity-robust Nearest Neighbor method with a minimum number of 3 neighbors.

Figure 4.9 shows the results of this estimation. Apart from the first row, the trend of being larger in magnitude and statistically more significant in the long term is present here. That is, for a given column, the effect becomes larger in magnitude after the second row. Moreover, when we consider the 5 years before the deal, all four of the coefficients are negative and statistically significant at the 5% level. For comparison, the ATE when we consider P_{55} is -0.64 log points (47% reduction) which is very close to what we found in local linear regression in Figure 4.6 (-0.63 log-points). The fact that the bias-corrected point estimation with robust inference also results in a statistically significant negative effect further supports the idea that the notification lowers the post-merger innovation for the deals that were reported as they were just above the threshold compared to the deals that were not reported as they were just below it in the 5-year window around the deal date. These results are surprising as the objective of the premerger notification program is to capture and block the anti-competitive deals with notification thresholds, hence, we should expect no effect at the threshold as the reported deals are deemed to be not anti-competitive. However, as explained before, the notification is very costly and comes with uncertainty for the merging entities, and the costs may actually outweigh, on average, the benefit of blocking anti-competitive deals, resulting in an overall negative effect. We talk more about this in Chapter 5.

4.3 High-patent producers vs. Low-patent producers

In this section, rather than using the full sample, we estimate ATE for P_{55} for a subsample given by an upper bound for the number of patents in 5-year window around a deal date. This is done to be able to distinguish deals that have a high number of patents from deals that have relatively lower patenting activity. To this end, we only include the deals that have a lower number of patents in 5-year window than a given maximum number, and we reduce this upper bound from 120 to 50 with a decrement of 1, and estimate ATE for the corresponding subsample with non-parametric robust estimation discussed in the previous subsection[6]. The only difference here is that we use MSE-optimal bandwidth choice as CER-optimal bandwidths are smaller and result in a loss of efficiency since we are considering strictly smaller samples in each iteration. Since the main objective is to observe the behavior of the estimated discontinuity rather than a robust inference, the bandwidth choice is not the main concern here. Nonetheless, we report the robust 95% CI's which justifies[7] the use of MSE-optimal bandwidths.

Figure 4.10 shows these estimates. X-axis values denote the upper bounds and are flipped for easier readability, while Y-axis values denote the estimated discontinuity and confidence intervals wrapped around it. The point estimates decline, in general, from left to right, or as we take a smaller subsample, while the upper bound of confidence intervals moves further away from 0 despite using a necessarily lower number of observations along the X-axis. For example, the ATE for the upper bound of 120 (the leftmost point) is -0.73 log-points (51% reduction), while it is -1.40 log-points (75% reduction) for an upper bound of 50 (the rightmost point). To put it differently, the ATE becomes larger in magnitude and statistically more significant as the low-patent producers take up a relatively larger part of the sample. This suggests that the deals with a relatively lower patenting activity tend to suffer more from the costly notification. Moreover, this finding implies that the overall negative effect estimated in the previous sections is largely driven by the low-patent [8] firms. The possible reasons leading to this is

[6]The results for bounds higher than 120 are not reported because of limited variability, and the results for bounds lower than 50 are not reported because of very few numbers of observations in the sample.

[7]MSE-optimal bandwidth is not valid for conventional CI's but is valid for robust CI's as explained in Calonico et al. (2014).

[8]High-patent and low-patent firms refer to the deals with a high or low number of patents in 5-year window around the deal date, respectively.

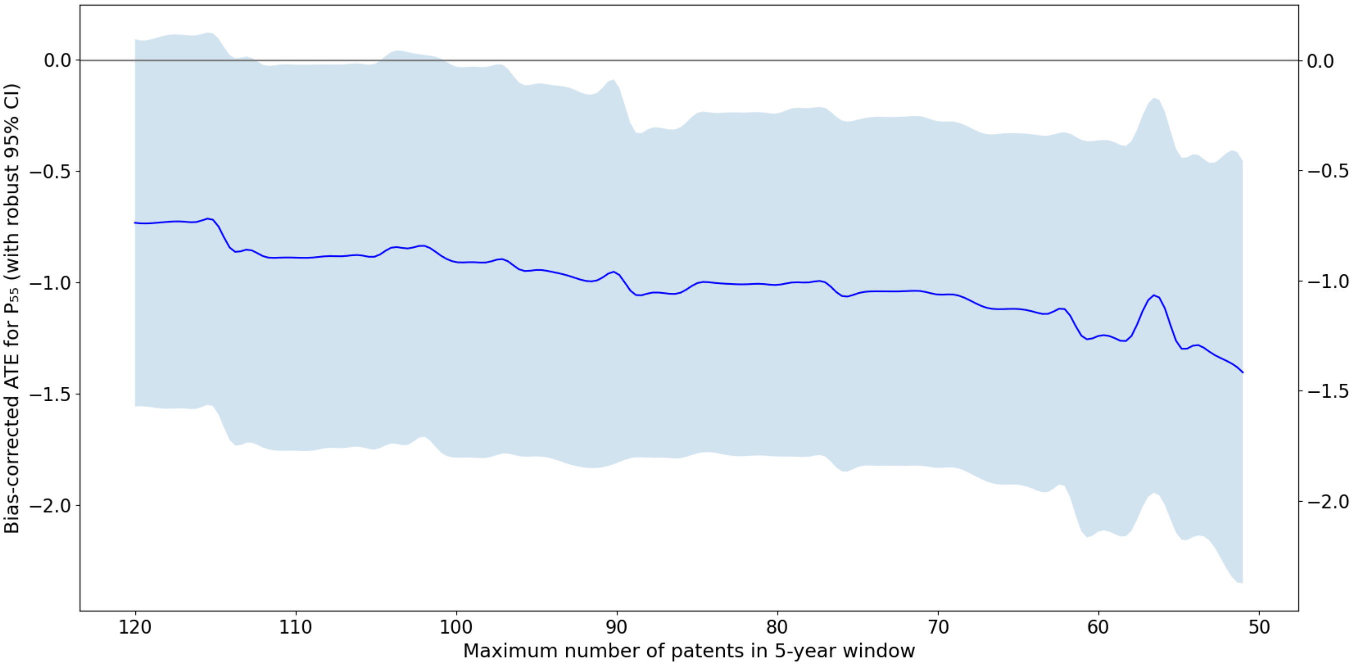

Figure 4.10: Estimated discontinuity in post-merger average patent intensity, P_{55}, for a given upper bound for the number of patents.

Notes: The value for a given level of the maximum number of patents in 5-year window around the deal date is the bias-corrected ATE (for the corresponding subsample) for patent intensity estimated via user-written "rdrobust" command with Epanechnikov kernel. Linear regression is used for point estimation, while quadratic regression is used for bias correction. The bandwidth choice is MSE-optimal and varies between $4-5$ for the left side, whereas it is between $3-4$ for the right side of the cutoff. Observations with an absolute value of the dependent variable higher than 3 are dropped. The robust 95% CI is depicted around a point estimate.

discussed in Chapter 5. It should be noted that a similar trend is observed for other long-term patent intensity coefficients and different kernel choices, but are not reported here for brevity. Moreover, the continuity of the running variable for each estimated is tested and we failed to reject the null of continuity at the 5% level for each estimation.

Chapter 5

Discussion

The premerger notification program in the U.S. requires merging entities to notify the authorities and seek approval if the deal size meets the thresholds outlined by the SOT and SOP tests in the HSR Act. The primary objective of the program is to capture anti-competitive deals and block them to preserve competition. As the processing of notifications is costly for FTC and DOJ and very small deals are unlikely to hurt competition, the thresholds serve to capture the deals that are likely to hurt competition and forego the deals that do not.

However, it can also be argued that having a threshold would also enable some anti-competitive deals to be consummate with no scrutiny as their deal value falls below the threshold. This is well analyzed in Wollmann (2019) and Morzenti (2022), where the former shows the rise of mergers between competitors below the threshold and the latter shows a reduction in innovation for such deals, both using the amendment to the Act in 2001. This paper, on the other hand, measures the impact of notification on innovation at the notification threshold of 50% million (as adjusted). This objective is different than the above papers as it measures ATE exactly at the threshold after the said amendment, and more importantly, it checks for the optimality of the threshold regarding post-merger innovation while the above papers analyze the impact of the amendment, not the threshold choice per se. The optimality in this context means that if the threshold was chosen optimally, then the effect of notification on innovation measures should be null. Contrary to the expected, we find that post-merger patent intensity, the most precise measure of innovation out of all three used in this paper, decreases by 47% in the robust bias-corrected estimation (the effect is similar in magnitude for other specifications as well), which means the premerger notification actually cause reported deals to be less innovative than unreported ones in a small window around the threshold, rendering it suboptimal. This impact is more pronounced in the long-term specifications as the duration between the emergence of an idea and its formation into a patent might take a few years. Albeit being less precise, citation and size also decline for all specifications and are statistically significant for some specifications. Moreover, to distinguish high and low patent producers, we bound the number of patents in 5-year window around a deal date above and found that the impact of costly notification is more detrimental for low patent producers and these are the deals that drive overall effect when we consider patent intensity in the long term as a measure of innovation.

The reason why the impact of notification on post-merger innovation is negative can be attributed to direct financial costs and uncertainty. Premerger notifications involve filing fees and other substantial costs as firms often need to hire legal counsel, economic consultants, and other experts to prepare submissions and guide them through the regulatory process. The amount of required information is sizeable and in the case of the second request, the total direct costs may exceed \$1 million (Howell, 2002). These costs have most probably risen up through the years as the FTC has tightened the requirements for correct notification multiple times after the amendment. The latest change in the requirements came around late June 2023[1] which will increase the burden on merging entities by increasing the number of hours required for documentation by 107 and will likely prevent some pro-competitive deals if enacted.[2]These direct costs divert the resources from R&D which is likely to lower innovation.

In addition to direct financial costs, the notification is also time-consuming and might result in strategic changes in the newly formed entity. The time to find and hire economists and experts, and to seek legal counsel would cost time to the companies in addition to time for preparing documents and a 30-day waiting period. Moreover, the companies have to present a strategy that does not raise a competitive concern for the authorities, and this may be achieved via strategic changes in the firms. As the firms are still subject to scrutiny after the approval, any diversion from the suggested competition and innovation strategy could be challenged by the authorities, with heavy penalties in case the punishment is deemed necessary. Hence, this creates uncertainty for firms regarding their innovation strategy as the new proposals might risk further scrutiny. These financial and opportunity costs, be they direct or indirect, present substantial drawbacks for notification, leaving its effect ambiguous. In fact, as we find in this paper, for a small window around the notification threshold, these costs indeed outweigh the benefits of preventing anti-competitive deals, resulting in the overall negative effect of a premerger notification program on innovation. Because of these findings, the notification threshold is deemed suboptimal. Moreover, this effect is larger in size if we consider low-patent deals as they may prioritize post-merger innovation less than high-patent deals for a given amount of costs.

[1]Proposed amendments to HSR rules.

[2]Killing Deals Softly: FTC Proposes 107-Hour Increase in Hart-Scott-Rodino Burden | Insights | Holland and Knight.

This paper has a few limitations. As we have a high number of private deals in the sample for which there is practically no data available, we are unable to test covariates being determined ”pre-treatment”. However, it is assumed and is likely to hold as explained in Section 3.2. Moreover, we do not observe name changes in the USPTO database which may result in not correctly capturing a full patent profile of a few deals. Since we do not observe financial data for private deals, we also assume that all the deals with a deal value above the threshold are reported to the authorities, resulting in Sharp RD design, rather than a ”fuzzy” one. However, this is unlikely to affect the results of this paper as the SOP thresholds are too lax to not capture a significant amount of deals, as explained in Section 3.2.

5.1 Concluding Remarks & Future Work

To sum up, this paper analyzes the effect of the premerger notification program on innovation by exploiting the notification threshold of \$ 50 million (as adjusted). We apply sharp ”Donut” RDD estimation and find that the the notification is indeed costly to the reported deals in the sense that it lowers their post-merger innovation score defined by patent intensity. To give an example, the post-merger intensity score when we consider 5 years before and after the deal date is reduced by 47% in the nonparametric estimation with CER-optimal bandwidths and robust SE's, and it is statistically significant at 5% significance level. The results are similar across different estimation techniques. Moreover, with the same method, we find that the effect of costly notification is more detrimental to low patent producers compared to companies with relatively higher patenting activity. These results suggest that merger review should consider innovation dimension of the merging entities differently so that the notification rules are not entirely based on financial figures and the size of a deal. Future work is required to have a deeper understanding of the relationship between notification and innovation. A theoretical model can be built to better understand the dynamics and timeline of how parties in an M&A deal plan their notification process and suggest an innovation strategy that does not raise competitive concerns, and how increasing cost burden might actually result in directing resources away from R&D activity, hence lowering post-merger innovation. Similar analysis can be done for other countries with similar notification requirements for comparison. These are left to future research.

Bibliography

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, *74*(1), 235–267. https://doi.org/10.1111/j.1468-0262.2006.00655.x

Barreca, A. I., Lindo, J. M., & Waddell, G. R. (2011). Heaping-induced bias in regression-discontinuity designs. https://doi.org/10.3386/w17408

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, *23*(2), 192–210. https://doi.org/10.1093/ectj/utz022

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). Rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, *17*(2), 372–404. https://doi.org/10.1177/1536867X1701700208

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295–2326. https://doi.org/10.3982/ECTA11757

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, *18*(1), 234–261. https://doi.org/10.1177/1536867X1801800115

Danqing, M. (n.d.). *Technology development and corporate mergers*. https://www.gsm.pku.edu.cn/info/1066/21712.htm

Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, *37*(3), 447–456. https://doi.org/10.1080/07350015.2017.1366909

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209. https://doi.org/10.1111/1468-0262.00183

Hart-scott-rodino antitrust improvements act of 1976 | federal trade commission. (n.d.). https://www.ftc.gov/legal-library/browse/statutes/hart-scott-rodino-antitrust-improvements-act-1976

Howell, A. G. (n.d.). Why premerger review needed reform - and still does. *Wm. Mary L. Rev.*, *43*. https://scholarship.law.wm.edu/wmlr/vol43/iss4/14

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, *79*(3), 933–959. https://doi.org/10.1093/restud/rdr043

Killing deals softly: FTC proposes 107-hour increase in hart-scott-rodino burden | insights | holland & knight. (n.d.). https://www.hklaw.com/en/insights/publications/2023/06/killing-deals-softly-ftc-proposes-107-hour-increase

Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, *108*(8), 2277–2304. https://doi.org/10.1257/aer.20160945

Kuhn, J., Younge, K., & Marco, A. (2020). Patent citations reexamined. *RAND Journal of Economics*, *51*(1), 109–132. https://ideas.repec.org//a/bla/randje/v51y2020i1p109-132.html

Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, *142*(2), 655–674. https://doi.org/10.1016/j.jeconom.2007.05.003

Lerner, J., Sorensen, M., & Strömberg, P. (2011). Private equity and long-run investment: The case of innovation. *The Journal of Finance*, *66*(2), 445–477. https://www.jstor.org/stable/29789786

Morzenti, G. (2022). *Antitrust policy and innovation*. https://economics.unibocconi.eu/events/giovanni-morzenti-antitrust-policy-and-innovation

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, *2*(1), 1–26. https://doi.org/10.2307/1164933

Wollmann, T. G. (2019). Stealth consolidation: Evidence from an amendment to the hart-scott-rodino act. *American Economic Review: Insights*, *1*(1), 77–94. https://doi.org/10.1257/aeri.20180137