

**CEP Discussion Paper No 1447**

**Revised January 2017**  
**(Replaced August 2016 version)**

**Colocation and Knowledge Diffusion: Evidence from  
Million Dollar Plants**

**Christian Fons-Rosen  
Vincenzo Scrutinio  
Katalin Szemerédi**

## **Abstract**

This paper uses the entry of large corporations into U.S. counties during the 1980s and 1990s to analyse the effect of plant opening on knowledge spillovers to local inventors. We use a difference-in-differences identification strategy exploiting information on the revealed ranking of possible locations for large plants in the US. Under the identifying assumption that locations not chosen (losers) are a counterfactual for the chosen location (winner), we find that patents of these large corporations are 68% more likely to be cited in the winning counties relative to the losing counties after entry. The effect materializes after the opening of the plant, rather than after the entry decision itself. The increase in citations is stronger for more recent patents whereas patent quality does not seem to play an important role. We find that the increase in citations is larger from patents belonging to the same technology class of the cited patent.

Keywords: productivity, innovation, knowledge diffusion

JEL codes: O3; R11; R12

This paper was produced as part of the Centre's Growth Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

## **Acknowledgements**

We would like to thank Richard Hornbeck for his support. We are grateful to Alan Manning, Guy Michaels, John Morrow, Gianmarco Ottaviano, Steve Pischke, John Van Reenen, Catherine Thomas, Tommaso Sonno, Giorgio Zanarone and all the participants at the London School of Economics Labour work in progress seminars and participants of the Universitat Pompeu Fabra Applied Economics work in progress seminar for their useful comments and suggestions. Fons-Rosen acknowledges that this project was supported by the Spanish Ministry of Economy and Competitiveness (ECO2014-55555-P).

Christian Fons-Rosen, Universitat Pompeu Fabra, Barcelona, GSE and CEPR. Vincenzo Scrutinio, London School of Economics and Centre for Economic Performance, LSE. Katalin Szemeredi, London School of Economics and Centre for Economic Performance, LSE.

Published by

Centre for Economic Performance

London School of Economics and Political Science

Houghton Street

London WC2A 2AE

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means without the prior permission in writing of the publisher nor be issued to the public or circulated in any form other than that in which it is published.

Requests for permission to reproduce any article or part of the Working Paper should be sent to the editor at the above address.

# 1 Introduction

It is a well established fact that for a region to sustain long-term economic growth and industrial development, it must keep upgrading its knowledge and technology. In addition to the region pursuing its own innovation projects, another relevant factor for its growth is its ability to absorb knowledge produced elsewhere. The entrance of new firms is one key way in which knowledge from the outside can be absorbed and diffused towards local firms. To date, not many studies provide causal estimates of the effect of geographical proximity on knowledge diffusion across local firms. The identification of the effect of location on knowledge diffusion is complex: the location decision of corporations is a choice variable and the endogeneity of this choice therefore results in an identification challenge. Firms select among possible locations and enter the one that provides the highest profit prospects, therefore resulting in location “cherry picking”. It may well be that these large firms choose to open a new plant in a region where knowledge synergies and technology complementarities make it a favourable location.

Due to this endogeneity concern, we exploit the setting developed by Greenstone and Moretti (2004) and Greenstone, Hornbeck and Moretti (2010) and use the revealed location ranking of large plants, Multimillion Dollar Plants (from now on MDP), to identify a valid counterfactual. The authors use the corporate real estate journal Site Selection which includes articles in which description of the location decision of a multinational plant is provided. These articles not only include the U.S. county that the plant chose, the “*winner*”, but also the runner up U.S. counties, the “*losers*”. The winner-loser pair is considered a *case*. Greenstone et al. (2010) find that five years after the new plant opening, the productivity of incumbent plants in winning counties is 12% higher compared to the productivity of plants in losing counties. A key contribution in our paper is to address a channel by which local firms can benefit from the entry of large corporations into their counties.

For our identification strategy, we use the assumption that the winning county (the county where the MDP is located) and the losing counties (where it did not end up locating), would have trended similarly had the MDP not entered. This assumption is supported by our data, as patents in both the winning and losing county show similar trends in citations made to the MDP firm’s patent before the entry of the MDP.<sup>1</sup>

We use data from the US patent office (USPTO) on patent citations as a proxy for knowledge flows from the MDP to US inventors. As the USPTO provides geographic information on the inventors, one can allocate each patent to a US county (and therefore, in our case, we will be able to match inventors- and their respective patents- to the winning and losing counties). Finally we standardise company names applying for a patent and match it to the company names of our sample of MDP firms. After the matching process we end up knowing each MDP firm’s

---

<sup>1</sup>for more evidence on the parallel trend assumption see Greenstone, Hornbeck and Moretti (2010)

stock of patents, how many citations they have received on a yearly basis, and the location of the inventors citing the MDP firm's stock of patents.

Our findings suggest that indeed the presence of the MDP firm leads to higher levels of citations to its patents from patents in the winner county. Patents belonging to the entering firm are 68% more likely to be cited in the winning county with respect to the losing county in the period after the entry. We find that knowledge flow is especially present once the plant opens ("opening year), while the decision itself to enter ("annouement year") does not lead to an increase in citations. The increase in the probability of citation is stronger for patents in the same technology of the patents of the entering firm. The effect is stronger for more recent patents, whereas patent quality does not seem to play a role in citation differences.

The rest of the paper is structured as follows: section 2 provides a literature review; section 3 presents our empirical strategy; section 4 describes the data; section 5 presents our main results; section 6 analyzes inventors citing the MDP patents; section 7 reports several robustness checks for our identification assumption; section 8 analyzes possible margins of heterogeneity; section 9 concludes.

## 2 Literature Review

There are four streams of related literature to which our project contributes. The first stream concerns knowledge flows. Employees from different firms exchange ideas about new products and new ways to produce goods. The ideas on positive production externalities through the sharing of knowledge date back as far as Marshall (1890), and then were extensively studied by Lucas (1988), Grossman and Helpman (1991) and Saxenian (1994) who studied spillovers within firms in Silicon Valley, Glaeser (1999), and Moretti (2004).

Our paper mainly contributes to a second stream of literature related to patent citations as a measure of knowledge flows. Citations appear to be correlated with the value of innovation (Trajtenberg, 1990), and patent citations have proven to be good indicators of spillovers (Jaffe, Trajtenberg and Henderson, 1993; Caballero and Jaffe, 1993). Jaffe et al. (1993) found that inventors cite other inventors living in geographical proximity more than proportionally than those further away. Patent citations have been used as a proxy for knowledge flows by Duguet and MacGarvie (2005), Jaffe et al. (2001), Jaffe et al. (2000). Inventive activity, based on patents, has been studied by Verspagen and Schoenmaker (2004) who use patent citations to measure the spatial concentration of knowledge flows between firms, and within the same firm between different regional locations. Agrawal et al. (2008) use patent citations to estimate knowledge flow production function and find that spatial and social proximity of inventors are substitutes in their influence on access to knowledge. Griffith, Harrison and Van Reenen (2006) examine the

location decisions of R&D of UK firms to the US. They provide evidence of knowledge spillovers associated with “technology sourcing” and find that UK firms particularly benefit from R&D labs located in the US in sectors in which the US inventive activity is high. Finally, Abramovsky, Griffith, Macartney, and Miller (2008) use patents to look at the productivity activities for firms of 15 European countries by looking at the patents that these European countries have applied for at the European Patent Office (EPO).

Given our setting, this paper also provides an indirect contribution to a third stream of the literature on economics of agglomeration economics literature. The positive impact of agglomeration was already put forth by Marshall (1890) who highlighted that in dense clusters, “the mysteries of the trade become no mystery but are, as it were, in the air.” There are various theories justifying this concentration including advantages to natural resources, low transport costs possibly due to cheaper and faster supply of intermediate goods and services (Krugman, 1991a and 1991b; Glaeser and Kohlhase, 2003), intellectual spillovers, better worker-firm match due to large labour markets (Krugman 1991a) leading to better labour productivity, as well as productivity advantages (Ellison and Glaeser, 1999). Researchers also showed that concentration is especially high between industries that are economically similar (Lychagin, Pinske, Slade, Van Reenen (2010) and Ellison and Glaeser (1997). Glaeser and Gottlieb (2009) argued that density, as a means to speed knowledge flows, plays a large role in modern city development. Ottaviano and Puga (1998) provide a survey on these matters.

Greenstone, Hornbeck and Moretti (2010) quantify agglomeration spillovers by estimating how the productivity of incumbent manufacturing plants changes when a new, large corporation opens in their county. They find evidence for local productivity spillovers and prove that economic distance, both in terms of industry similarity and geographical proximity, matters. By using firm-to-firm linkages using the USPTO, our paper provides evidence of a possible channel through which firm entry might benefit incumbent firms, namely through technology diffusion and knowledge flows.

Finally, the fourth stream of literature is the one related to foreign direct investment (FDI). There has been a long-lasting debate on the effect of FDI performed by large multinational enterprises (MNEs) on the local community. Evidence supporting positive spillovers through FDI is mixed, as is argued in a number of survey papers such as Gorg and Strobl (2001), Crespo and Fontoura (2007) and Gorg and Greenaway 2004. Benefits of spillover however are present in the work of Arnold and Javorcik (2009), Blomstrom and Persson (1983), Kokko (1994) and Kokko (1996), Blomstrom and Kokko (1996, 1997), Lipsey (2001a, b) and Javorcik (2004) . These authors mostly present positive spillover from the host country’s perspective, due to productivity or market access. Most of the FDI and knowledge spillover literature uses a production function equation to analyse total factor productivity (TFP) residuals. With this approach, firm-level output or growth rates are regressed on inputs such as capital and labour which lead to TFP

residual. Even though this approach brings valuable insights, it is not the optimal approach to analyse micro-level linkages between firms. Furthermore, price effects caused by possible changes in competition can lead to imprecise coefficient estimates. Therefore, our approach based on patent citations, is complementary to address how large multinationals affect local innovation conditions.

Our approach is similar to Branstetter (2006) and Fons-Rosen (2010). These papers use patent citations to proxy for knowledge flows and find, respectively, that Japanese FDI increases the flow of knowledge spillovers from and to the U.S., and that FDI performed in Central and Eastern Europe (CEE) leads to an increase in knowledge flows to CEE firms.

### 3 Empirical Strategy and specification

Our aim is to estimate the causal effect of firm location on knowledge diffusion. In order to have a reliable estimate we require the identification of a location that is similar to the country where the MDP firm decided to locate. However, the location decision of a firm is not exogenous: firms select among possible locations, carefully considering locations which provide the highest opportunities and benefits for them. Firms might take into account complementarities between their technological endowment and the prevalent technology in the location chosen for the plant, implying a non-random location decision. As a result, the identification of a proper counterfactual location is particularly challenging as the chosen location is unlikely to be similar to another randomly chosen location.

To overcome this identification challenge, we exploit information on the revealed ranking of possible locations for large plants collected by Greenstone and Moretti (2004) and Greenstone, Hornbeck and Moretti (2010). The authors use the corporate real estate journal *Site Selection* which includes articles describing the new plant location decision process of large firms (the "Million Dollar Plant"). These articles not only include the country that the firm chose for its plant, the "*winner*", but also include the runner up U.S. counties, the "*losers*". The winner-loser pair is considered a *case*.

The identification assumption for our empirical analysis will be that losing US counties represent a valid counterfactual with respect to the winning county.

Figure 1 provides supporting evidence for our identifying assumption. The figure plots the number of citations received by the patents of the entering firm by patents produced by inventors located in the county. As it can be seen both levels and trends in citations are very similar in the winning and losing county before the entry decision. After the entry year, however, the average number of citations made to the MDP is consistently above the one of the losing counties. This difference widens considerably after the first four years. This delay can be explained by the time

needed to absorb the technology of the new firm. In addition, it is important to note that there are two different time periods to consider. The first one is the firm's *announcement to enter* which we define as entry year, the second one is the *actual opening* of the new plant. These two dates might not coincide, as the construction of the plant and related infrastructure might take several years from the point of announcement. We exploit this delay in one of our identification checks in section 6. A similar picture is provided by table 1 which reports the total number of citations before and after the entry announcement of the MDP in the Winning and Losing counties. Numbers are reported also in terms of citations per county to account for the different numbers of winning and losing (i.e a case can have more than 1 losing county and this would make the raw number comparison inappropriate). The number of citations per county is very similar in the period before entry: the small discrepancy can be explained with the small drop in citations in winning counties 2 years before entry. There is however a stark difference in the period after entry: the number of citations per county is more than double in winning counties with respect to losing counties.

We use citations to patents of the entering firm as a proxy for knowledge diffusion. More specifically, we restrict our attention to patents of the entering firm with application year before the decision of entry. We do this as the entering firm could have internalized the technology of the chosen location after entry. Hence patents produced by the entrant after entry could be more compatible with local technology and, as a consequence, would be more likely to be cited.

The choice of using patent citations to identify knowledge flows is consistent with previous literature such as Branstetter (2006) and Fons-Rosen (2010). Patent citation represent an attractive element for several reasons: first, they allow to clearly identify references to already existing technology; second, information are likely to be accurate. It is required that every new patent cites the existing technology on which it builds upon in order to identify possible infringement in the property rights of other patents owners. For this reasons patents are carefully reviewed by the UPSTO and possible missing citations are included by examiners (Thompson, 2006). Hence, citations represent an accurate link between new and past knowledge. At the same time, however, we also acknowledge that patent citations might represent only part of the information flows to local firms after the opening of a new plant. Local firms could, in fact, benefit from knowledge of different practices of the new firms such as production organization, inventory and human resource management as well as many others. Due to the lack of suitable data, we ignore these possible additional margins. It could be argued, however, that the use of a firm technology for the development of new technology represents a particularly demanding form of technology flows and hence our effects could be considered a lower bound for the actual knowledge diffusion.

In terms of the econometric technique, we have a wide range of possibilities. As the the number of citations received by a patent is a count variable, a poisson regression would be the

most appropriate tool. However, this presents some difficulties due to the large set of fixed effects that we would like to use in our model (which might lead to incidental parameter bias invalidating our estimates) and to possible convergence issues. For these reasons as our main specification to estimate the propensity of local patents to cite patents of MDP firms, we use a linear probability model and provide results with the Poisson estimator in the Appendix. This choice also allows to avoid any assumptions regarding the underlying distribution of the error term.

In our initial specification we aggregate the data at the case-county-year level. The dependent variable is a count of the number of citations received and the baseline specification is:

$$\begin{aligned} NumCitations_{tcf} = & \alpha + \beta_1 Post_{tc} + \beta_2 Winner_{cf} + \beta_3 Winner_{cf} X Post_{tc} + \\ & + \gamma Z_{ft} + \eta_c + \theta_f + \lambda_t + \mu_{cf} + \varepsilon_{tcf} \end{aligned} \quad (1)$$

where  $t$  represents the year,  $c$  the case and  $f$  the county.  $NumCitations_{tcf}$  is the total number of citations by patents produced in county  $f$  in year  $t$  to the pool of pre-existing patents of the entering firm, belonging to case  $c$ .  $Post_{tc}$  is a dummy equal to one after the year in which the entry is announced.  $Winner_{cf}$  is a dummy variable that takes value one if the MDP firm belonging to case  $c$  enters county  $f$  at any point in time. The interaction term  $Winner_{cf} X Post_{tc}$  is the classic difference in difference term. In addition, we add time-varying controls at county level ( $Z_{ft}$ ) and a rich set of fixed effects: year ( $\lambda_t$ ), county ( $\theta_f$ ), case ( $\eta_c$ ). Standard errors are clustered at the county level.

Our main coefficient of interest is the interaction between  $Winner_{cf}$  and  $Post_{tc}$ , that is,  $\beta_3$ . This represents the increase in the number of citations received by the stock of patents of the MDP firm after its entry decision in the winning county with respect to citations from the losing county.

After these preliminary results, we move to a richer framework which allows the analysis of citations received by each patent of the entering firm. Hence, this specification allows for a disaggregated measure of the previous specification. This specification delivers a more complete description of the the extensive margin, by capturing not only if there has been at least one citation from the patents produced in the county but also if different patents of the entering firm were cited. We present results for a linear probability model with dependent variable, a dummy equal to 1 if the patent belonging to the entering firm receives at least one citation from patents produced in the county. Results using the number of citations are consistent and reported in Appendix. The model is then:

$$\begin{aligned}
Citation_{ptcf} = & \alpha + \beta_1 Post_{tc} + \beta_2 Winner_{cf} + \beta_3 Winner_{cf} X Post_{tc} + \\
& + \delta X_{pt} + \gamma Z_{ft} + \eta_c + \theta_f + \lambda_t + \mu_{pcf} + \varepsilon_{ptcf}
\end{aligned} \tag{2}$$

notation is similar to 3 but includes  $p$  which identifies the patent.  $Citation_{ptcf}$  is a dummy equal to one if patent  $p$  of the entering firm, belonging to case  $c$ , received at least one citation by patents produced in county  $f$ , in year  $t$ , and takes on the value of zero otherwise. In addition with respect to the previous specification, we add time-varying controls at patent ( $X_{pt}$ ) level and add patent-county-case fixed effects ( $\mu_{pcf}$ ). These fixed effects are particularly relevant as they allow to account for any time invariant unobserved complementarity between the technology used in case  $c$ , county  $f$  and patent  $p$ , which could bias our estimates. Standard errors are again clustered at county level.

Also in this case, our main coefficient of interest is the interaction between  $Winner_{cf}$  and  $Post_{tc}$ , that is,  $\beta_3$ . In this setting, however, it can be interpreted as the incremental probability of the MDP patent  $p$  being cited in the winning county with respect to the probability of citation by patents in the losing county. A positive coefficient would be consistent with technology diffusion.

In order to perform the above regression, we construct a panel for patents of the MDP firm produced before the decision of entering the winning county. The panel tracks the number of citations received from county  $f$ , in case  $c$ , by patent  $p$  in each year  $t$ . Additional details on the panel construction are reported in Appendix A.

## 4 Data

### 4.1 Plant opening decision

We use data for the opening of a new plant by large corporations in the US between 1982 and 1992 reported in Greenstone and Moretti (2004). The authors exploited information regarding the selection of counties for a new plant reported in the real estate journal “The Site Selection” in a series of articles called “The Million Dollar Plant”. These articles reported the final location chosen together with runner up locations, that is to say suitable counties which were discarded at the last step of the decision process. Each plant opening decision is defined as a “case” and it involves the company opening the plant, the county where the plant should have been located (“winning county”) and the set of runner up counties (“losing counties”). Note that in some cases the firm finally decided not to open the plant or we do not find any news related to the plant

opening. These cases were excluded from our main analysis and used in a placebo identification check in section 6.

Starting from the original set of 82 plant openings we reduce our sample to 49 usable cases. In particular, we require a case to satisfy the following 3 criteria:

1. MDP firms in either manufacturing, service or utility sector (14 cases lost).
2. We exclude cases where we were unable to confirm the opening of the plant or we found evidence that it did not take place (15 cases lost).<sup>2</sup>
3. Since our method is based on patent citations, we only keep cases with MDP firms with a positive patent stock (4 cases lost). More specifically:
  - We exclude MDP firms that had no patents in the period 1975-2010.
  - The MDP must have produced at least one patent before the entry decision.

An example of a missed entry is the case of Volvo (case 62 of Greenstone and Moretti, 2004). The company planned to open a plant for automobiles in 1991 in Chesapeake county (Virginia), but the project was eventually discarded. According to an article reported in the website of Don Beyer, a US automobile dealer:

*“At the time, it was slated to be built in Chesapeake, VA ... Due to a myriad of factors Volvo decided to open a plant in South Carolina instead. Now after all these years, we will see a US built Volvo roll off the assembly line in 2018”.* (Beyer Auto, 2015)

An example of a clear entry is that of TRW (case 13 of Greenstone and Moretti, 2004), an electronic company, which opened a plant in Fairfax county in 1987. The dedication of the plant was described in an article in the *Washington Post* in 1987:

*“The occasion was the dedication of the first building of the security-conscious industrial giant’s TRW Federal System Park on a wooded 120-acre site between Rte. 50 and I-66 in fair Lakes”.* (Washington Post, 1987)

Table 2 shows descriptive statistics on cases. The top panel reports the number of “winners” and “losers”. In our sample we have 49 cases, 50 winning counties, and 80 losing counties totaling 113 locations involved in our bidding process.<sup>3</sup> The second panel describes the number

---

<sup>2</sup>Plant openings were confirmed through local journal articles, books and websources

<sup>3</sup>In one case, the MDP firm was located on the border of two counties, therefore we assigned both counties as winners.

of cases by entry year. The bidding cases take place between 1982 and 1993, with substantial variation across these years.

## 4.2 Patents

For our research we track patent citations. For this purpose, we use the data collected by Lai et al. (2014) based on U.S. Patent and Trademark Office data (USPTO) and the NBER patent database (Hall, Jaffe and Trajtenberg, 2005; Jaffee and Trajtenberg, 2002). These datasets contain detailed information on the universe of U.S. patents granted between January 1963 and December 2010 and all citations made to those patents between 1975 and 2010. For our analysis, we use information on the application year of the patent, the technological area to which it belongs, the inventors (e.g. their geographical location), and the assignee of the patent.

From the full set of US patents we identify the subset of patents (7.53% of all the US patents) belonging to our MDP firm.<sup>4</sup> It should be noted that patents are not evenly distributed across firms. As described in Figure 2, out of the 48 MDP firms,<sup>5</sup> more than 50% of the pool of patents belongs to the top 5 MDP firms.

## 4.3 Citations

From all citations by patents registered in the US (43,288,130), 5.64% are made to patents belonging to our MDP firms. We restrict our sample to patents satisfying the following conditions.

First, we only use *citing* patents of US inventors that were applied for in the winning and losing counties.<sup>6</sup> Second, we are interested in citations in which the *citing* (county of inventors) and *cited* patent (pool of patents of the entering firm) belong to the same *case*. Third, we only look at citations made to the pool of patents of the MDP firm with application year *before the year of the decision to enter*. To clarify, the MDP firm can produce patents after entry which align with the technology of the county in which it locates. As these patents might be directly related to the local technology of the county, patents produced post entry by the MDP could be cited more. This increase in citations will falsely be attributed to the MDP entry.<sup>7</sup> Fourth, self citations are excluded from the analysis. Once we restrict our sample of citations based on these four criteria, we are left with 1710 citations.

---

<sup>4</sup>Patents are assigned to firms based on the name of the assignee at registration. Unfortunately, it is not possible to keep track of ownership change of the patent using UPSTO data. This limitation might lead to a negative bias on our estimates and should lead to interpret them as a lower bound of technology spillover

<sup>5</sup>There is one firm involved in two different cases.

<sup>6</sup>We are able to identify the location of 99.2% of all US inventors.

<sup>7</sup>The MDP pool of patents, from now on, will refer strictly to the patents produced by the MDP prior to the entry date.

## 5 Results

### 5.1 Basic result

Results for equation 3 - with specification on county-year-case level - are reported in Table 3. Column (1) shows that the entry of the MDP firm is related to an increase in the average number of citations received by its stock of patents. The coefficient is relatively stable across specifications: it implies that, on average, patents by the entering firm receive 0.7 more citations in the winning county with respect to the losing county after entry. The effect represents a 97% increase in the number of citations with respect to a baseline of 0.74 citations per year from patents in the losing county<sup>8</sup> in the same period. Throughout the rest of our paper, the probability of citation in the losing county after entry will be our baseline probability to which our coefficient of interest is compared to. Column 2 includes years fixed effects together with the log of active population (the number of individuals between 25 and 54 years of age) and the number of citations received by the entering firm patents from other counties in the US. Column 3 includes county and year fixed effects to control for county and year shocks and column 4 adds interactions between year and case fixed effects to control for any possible time pattern of citations within case effects. Finally, column 5 includes county/case fixed effects, which allow to control for any time fixed complementarity between the technology of the county and of the entering firm corresponding to case  $c$ . While this result suggests that indeed the entry of the firm leads to an increase in the number of citations, this specification is characterized by several shortcomings: first, it does not allow to control for patent characteristics and hence we cannot exclude that these results are driven by heterogeneity at the patent level; second, it is not possible to disentangle between the *intensive* margin, i.e. if the patents of the MDP firm that were already cited before the entry of the MDP are the ones which continue to be cited, and the *extensive* margin, i.e. if patents of the MDP firm that were not being cited prior to the entry of the MDP firm are now being cited. The latter case would be more relevant as it would point to the diffusion of different technologies rather than a more intense use of those patents already cited before entry. For these reasons, we move to our second and preferred specification which is based on patent level information.

In Table 4, we present basic estimates of the effect of the MDP's location decision on citations received by its pre existing patents as described in equation 2. For the sake of clarity, coefficients are multiplied by 1000. Column (1) includes *Winner* and *Post* dummies and their interaction, and shows that patents of the MDP firms are more likely to be cited in the winning county after the entry decision of the firm. The effect corresponds to a 0.0327 percentage points increase over the baseline of 0.0497 percentage points per year from patents in the losing county. This represents a substantial relative increase: our effect represents a 68% increase in

---

<sup>8</sup>Calculated as the average citation in the losing counties in the period after entry (when *Winner*=0, and *Post*=1)

the probability of citation. Column (2) includes a set of fixed effects for the number of years after the application year of the MDP patent to control for the time pattern of citations of the patent, and year fixed effects to control for any year shocks. We also include the number of citations received by the MDP patent from counties not involved in the specific case, *Other citations*. This is particularly relevant as it controls for possible patterns in the use of a particular technology in the overall economy.<sup>9</sup> In addition, we control for log of the population of working age individuals (25-54) in the county, *log(ActivePopulation)*, to normalize citations for the size of the county. The estimated coefficient for the interaction term remains large and statistically significant. In column (3) we add county and case fixed effects to control for any time invariant county and case variation. Column (4) further includes year-case fixed effects to control for any time-case specific shock. Column (5) adds county, case and patent fixed effects interactions which account for any possible time invariant complementarity between each patent and the county technology characteristics. Considering the advantages of this regression model over the previous one, this will be our preferred specification throughout the rest of the paper. It is worth emphasising that, despite the inclusion of this large set of controls and fixed effects, the magnitude of our coefficient of interest remains close to the one in the simplest specification in column (1).

Controls have the expected sign. The effect of (the log of) active population size is positive and statistically significant in column (2) but becomes insignificant once county fixed effects are included. This might be related to the fact that population does not change substantially over the period considered and hence the inclusion of county fixed effects absorbs most of the relevant variation. The number of other citations received in the US is also positive and highly statistically significant in all the specifications. This allows to control for possible time trends in the use of a particular patent by tracking the general use of the patent in the rest of the US. Estimates using the number of citations as dependent variable with a simple OLS regression or a Poisson model are consistent with results reported. The corresponding tables are reported in the appendix in section Appendix B.

These findings are consistent with several possible mechanisms. First, the announcement of a plant opening by a large firm in an area could increase inventors' awareness of that company's stock of patents and lead them to search for possible useful patents related to the company. This would imply that the announcement of the plant opening could be sufficient to increase the number of citations. Second, firms integrated in the production chain of the MDP firm could develop complementary technologies to the ones of the MDP firm. Third, the opening of a plant can allow local inventors to meet with inventors working at the MDP firm. This would improve knowledge of the firm's technological endowment. Fourth, inventors might work temporarily for

---

<sup>9</sup>Controlling non-parametrically for the number of other citations using a set of dummy variable leads only to a small change of our coefficient of interest, which remains positive and highly significant (result not reported).

the MDP firm and then produce patents alone or for different firms exploiting their accumulated knowledge. It should be noted that all these mechanisms are not mutually exclusive.

## 5.2 Citations from MDP firm to local patents

Results in the section above show that the MDP firm's entry leads to an increase of citations to its stock of patents from inventors located in the winning county. The information flow can also be bilateral, i.e after entry, the MDP firm can become acquainted with patents of local inventors due to geographic proximity. We explore this possibility by developing a framework similar to the one described above but now considering citations made by inventors of MDP to pre existing patents by inventors in winning and losing counties.

The number of citations from the pool of MDP patents to local patents is reported in figure 3. MDP inventors do not site the stock of patents more in the winning county than in the losing county. This is of no surprise, as MDP firms are in general large and international and hence are already aware of any relevant technology for their production process. This suggests that the MDP location is not driven by any specific complementarity or relationship between the county and the MDP firm's technology set.

## 6 Addressing Identification Concerns

In this section, we perform various checks to validate our identification assumptions. More specifically we perform 4 different checks. First, we evaluate if the period before the entry was characterized by different trends in citations to entering firm's patents in the losing and winning county; the regression framework used also allows to assess the time pattern of the effect of entry on citations. Second, we exploit information on plant opening dates to evaluate if the increase in citations materializes also in the period between the announcement of the entry decision of the firm and the actual opening of the plant. The absence of any effect before the opening would suggest a direct channel for the technology diffusion such as workforce interaction or supply/demand channels. In addition, we can consider the period between the entry and the opening as a placebo period where the choice of the county has already been made but the presence of the company is not yet established with a functioning plant. The absence of any difference in citations to the entering firm's patents between winning and losing county would provide additional evidence in favour of our parallel trend assumption. Indeed, in presence of different trends in winning and losing counties, we would observe an increase in citations before the opening of the plant as this trend would have been present at the time of the entry decision of the firm. Then, we assess if the number of runner-up counties is related to the magnitude of the effect. If the cases with more losing counties are characterized by larger

effects, this could suggest that the control group in these cases might be contaminated by the presence of some less suitable controls as the elimination process of more suitable locations was stopped earlier. Finally, we use cases excluded from the analysis as we were not able to confirm the opening of the plant or we found evidence that the plant did not open to perform a standard placebo test.

## 6.1 Pre-trend and time pattern

In order to evaluate the possible presence of different trends in the winning and losing counties, we use a standard variation in the difference in difference framework. We estimate the following equation:

$$\begin{aligned}
 Citation_{ptcf} = & \alpha + \beta_1 Post_{tc} + \beta_2 Winner_{cf} + \sum_{j=-6}^{-1} \delta_j I(distance_{tc} = j) + \sum_{j=1}^8 \delta_j I(distance_{tc} = j) + \\
 & + \sum_{j=-6}^{-1} \gamma_j Winner_{cf} XI(distance_{tc} = j) + \sum_{j=1}^8 \gamma_j Winner_{cf} XI(distance_{tc} = j) + \\
 & + \delta X_{pt} + \gamma Z_{ft} + \eta_c + \theta_f + \lambda_t + \mu_{pcf} + \varepsilon_{ptcf}
 \end{aligned} \tag{3}$$

where  $distance_{cf}$  is the difference between year  $t$  and the year of entry of the firm in the winning county in the specific case.<sup>10</sup> The first two sums includes a set of time fixed effects for years before the entry (first sum) and after entry (second sum). The two following sums include interaction terms between these fixed effects and the winner dummy. The year of entry of the firm is used as the baseline. Hence, the interaction terms describe the difference over time between the winning and losing counties with respect to the year of entry. Ideally, we would like to observe coefficients for the interaction terms for years before the entry (i.e those corresponding to years with  $distance < 0$ ) stable in size and close to zero whereas coefficients for the interaction terms for years after entry (i.e those corresponding to years with  $distance > 0$ ) should be positive and statistically significant. This would support our assumption of parallel trends in absence of the firm's entry. In addition, this specification allows to assess in which period after the entry decision the increase in citation takes place and how persistent it is. Figure 4 reports the coefficients for the interaction terms together with 95% confidence intervals.

The estimation shows no difference in the citation dynamic in periods before entry between winning and losing counties. The coefficients on the interaction terms are virtually zero before entry and they progressively increase over the years. Coefficients after entry are consistently larger with respect to those of years before the entry. In terms of significance, the coefficient for the year following entry and those for 5 or more years after entry are significant at 5%. This delay could be explained by the time needed by inventors to incorporate the technology of

---

<sup>10</sup>Suppose that in case  $c$  the MDP firm enters the winning county in 1992. Then, the year 1990 will correspond to  $distance = -2$ , the year 1991 to  $distance = -1$  and so on for all the other years.

entering firms or by delay in the plant opening. Plant construction and related delays can create a time gap between the entry decision and the actual opening of the plant. This would translate in a delayed effect on citations if indeed production chain relationship or direct contact between inventors are crucial for knowledge diffusion. We explore this possibility in the next section.

## 6.2 Announcement and opening dates

As mentioned above, although the articles in the Site Selection Journal report the year in which the firm *announced to open* a plant in a particular location (entry year), in several cases this year does not coincide with the *opening date* of the plant. In particular, when dealing with large plants as those in our setting, the construction time for the plant can be non-negligible. We were able to identify the year in which the plant opened in several cases.<sup>11</sup> In cases where the year of opening (or expected opening) was not specified, we assume that the plant opened in the year of entry (or in the following year if construction started in the second half of the year). Overall, in about 60% of our cases the entry year differs from the year of the plant opening. This delay allows to analyze if the opening of the plant is the event triggering the increase in citations rather than the decision to enter.

Figure 5 reports the number of citations based on the actual opening year rather than the decision to entry year. It shows that the opening itself corresponds to a large increase in the average number of citations received. This graphical intuition is confirmed in a regression framework. More specifically, we estimate our preferred specification including now an additional term for the period after the opening and an interaction term with the dummy for the winner county. We would expect the interaction between *Post Announcement* and *Winner* to be insignificant, whereas the interaction between *Post Opening* and *Winner* should be significant and possibly larger in magnitude with respect to what we found in Table 4.

Figure 5 reports the number of citations based on the actual opening year rather than the decision to entry year. It shows that the opening itself corresponds to a large increase in the average number of citations received. This graphical intuition is confirmed in a regression framework. More specifically, we estimate our preferred specification including now an additional term for the period after the opening and an interaction term with the dummy for the winner county. We would expect the interaction between *Post Announcement* and *Winner* to be insignificant, whereas the interaction between *Post Opening* and *Winner* should be significant and possibly larger in magnitude with respect to what we found in Table 4.

Table 5 confirms our prediction: the interaction term between winner and the period after opening is positive and highly statistically significant after the inclusion of county and case fixed

---

<sup>11</sup>In 12 cases we find that the plant opened 1 year after the announcement. In 11 cases, it was 2 years. In 5 cases, it was 3 or more years.

effects in column (3); *WinnerXPost Announcement* is instead never statistically significant and the size of the coefficient is substantially smaller. Additionally, the coefficient for *WinnerXPost Opening* is about 12% larger with respect to the *WinnerXPost* term in Table 4. These results underline the relevance of the plant opening for technology diffusion and support the idea that it could be channeled through supply chain relationships or direct contact between inventors, rather than through awareness about the company's existence.

### 6.3 Number of losing counties

Our identification assumption is based on the idea that losing counties are a good counterfactual for counties where the firm decided to locate. The number of losing counties varies across cases and goes from 1 to 7. In cases with a higher number of losing counties it is possible that the firm stopped the elimination of non-suitable counties early on. This might imply that our set of counterfactuals is contaminated by the presence of counties that are not similar to the winning county. We evaluate whether this is the case by splitting the dataset in two different subsets of cases. These sets are defined according to the number of losing counties reported. More specifically, we run our preferred specification first on a subsample where every winning county is matched to only one losing county and the second subset where there is more than one losing county. If indeed cases with more than one loser are characterized by a less suitable appropriate control group, we would expect the coefficient for the latter group to be larger than the one where every winning county is matched with one losing county.

Results are reported in Table 6. These results show that the coefficient is larger for cases with more than one losing county. The coefficient for cases with one losing county (column 2) is 25% smaller than the one for cases with more than 2 losing counties (column 3). the latter is also significant at 5% with respect to 10% in column 2. However, a formal test for statistical difference between these two results with a triple interaction strategy (column 4) does not allow to reject the null hypothesis of no difference between the two. All balanced, this test does not provide compelling reasons to conclude that our results are driven by the presence of not appropriate counterfactuals.

### 6.4 Placebo with excluded cases

Finally, we use 15 cases involving manufacturing and service firms for which we were not able to confirm the plant opening or we find evidence that the firm explicitly renounced to open the new plant. In table 7, we present results for our preferred specification using this set of cases. As previously, the dependent variable is a dummy taking value 1 if the patent receives at least one citation by patents produced in county  $f$  in year  $t$ . Column (1) presents the simplest

specification without any fixed effects; the estimated coefficient of the interaction term is close to zero and far from statistically significant. The coefficient becomes even negative in the following columns after regressors and fixed effects are included but it is never statistically significant. This is an additional strong indication that geographical proximity is relevant for the diffusion of knowledge, and that results are not driven by news or announcements about future plan openings.

## 7 Inventors

In this section, we look at inventors in winning and losing counties. The original USPTO database only provides information concerning the name and surname of the inventor, which, due to spelling issues, abbreviation and homonymy, make it difficult identify patents belonging to the same inventor. We therefore rely on the disambiguation dataset developed by Lai et al. (2014). This is based on probabilistic matching approach based on inventor and patent information. This dataset is the building block for our panel of inventors. Inventors appear in our dataset once they register their first patent and are assumed to quit being inventor after their last patent in the dataset. Although this assumption might present some criticalities, it is probably the most reasonable with the data at hand. We drop few observations which implied that an inventor had been patenting for more than 60 years (this accounts 350 out of more than 3,000,000 inventor-county-year observations). The location of inventors across the years is based on the residence information of the inventor reported in their patents. As inventors do not register patents every year, assumptions have to be made concerning the inventor location when no patent is registered. Suppose that an inventor register two patents in two different years and these patents are reported in different counties. Given the limited knowledge, we cannot observe the true county of residence for the inventor in the years elapsed between the application years of the two patents. It should be noted that this concerns only a small fraction of inventors : about 90% of inventors are reported in a single county in the dataset and only 2.10% appear in more than 2 counties. In this section we assume that inventors change residence in the year in which they appear in the new county. If several observations for the same inventors with different locations are present in the same year, it is assumed that the inventor moved during the year in the new county. We exclude inventors for which more than 2 locations (counties) are reported in the same year. This leads to a loss of about 0.82% of the observations. These observations might be more likely to be misclassified and hence we exclude them to minimize measurement error. The overall loss of data is anyway negligible.

The panel of inventors allows us to compare inventors in winning and losing counties. The more similar the inventors in the respective counties are before entry, the more evidence we have that our MDPs did not enter based on inventor characteristics. For our analysis we

identify the set of inventors present in the winning and losing counties up to six years prior to entry and compare their observable characteristics. Results are reported in table (8). The table shows in column (1) to column (4) that there is no difference between the two groups in experience (number of years since their first patent), the number of patents produced by year, the number of yearly patents produced and weighted by quality<sup>12</sup>, the number of patents produced in the past (cumulative). Column (5) and column(6) show that there is a statistical differences when considering the cumulated number of patents weighted by quality (both in levels and logs). However, the size of the difference is negligible once compared to the standard deviation of the variable of interest and therefore it is unlikely to play a major role in the location decision. Most interestingly, column (7) and column (8) show that inventors in the winning and losing counties are not different with respect to the number of yearly registered patents or the cumulated number of patents in the technology class of the MDP. This provides evidence that the firm did not choose the location due to a concentration of inventors in its technology field. This further supports our claim that the losing counties are an appropriate counterfactual for winning counties.

As a first step, we describe inventors citing the MDP patents in both the winning and losing counties. We classify inventors according to whether they appear in our dataset before or only after the date of entry of MDP. In table 9 we observe the distribution of inventors. The table shows that the composition of the inventors citing the MDP is similar in both the winning and losing counties. Almost half of our inventors are New Inventors. These are inventors who only appear in our dataset after the entry date and their first reported residence is in the winning county, respectively losing county. The second largest group are stayers. Stayers are incumbent inventors, that is, they appear in the respective county both prior and after entry date. It is worth emphasizing that the number of inventors among the stayers in the losing county who cite the MDP patents before the MDP entry date is the same as the number of stayers who cite after the MDP entry date, while in the winning county, among the stayers most of the stayers cite the MDP firm only after entry date. This shows, that the presence of MDP plays an important role in the winning county. The last group, immigrants, are those inventors which do not appear in the winning county before the entry. They represent about 11 percent of our inventors. The remaining inventors only appear in the winning county prior to MDP entry. These are divided into two groups: Quitters who disappear from the dataset after entry, and Leavers who remain in our dataset, however move to a different county from the winning one.

We now look at the difference in characteristics of inventors who cite our MDP and who do not cite our MDP in the winning county. In table 10 reports coefficients of regressions comparing the two groups of inventors (those that cite and do not cite) according to experience and quality. Quality is proxied by different measures of patents produced up to the year in which

---

<sup>12</sup>Quality is defined as the number of citations received within five years from application normalized by the number of patents received by patents with the same application year

they cite the MDP. The first row compares the two sets of inventors. In the subsequent rows, we continue the comparison by dividing the inventors according to the groups mentioned in the previous paragraph. We find that citing inventors tend to be less experienced than non-citing inventors, but they tend to be of higher quality especially in the main technology class of the MDP. This suggests that those benefiting from the entry are the inventors who are in the early part of their career: the lack of a pre existing set of projects and knowledge might allow them to be more flexible and prone to absorb the newly available technology. At the same time, these inventors appear also to be the most productive, considering the number of patents produced both weighting by citations received and not. This shows that the top part of the quality distribution of inventors is the most receptive to MDP technology.

Finally, we can exploit our data in order to investigate some possible channels through which inventors might get acquainted with the MDP stock of knowledge. For this purpose we define two sets of inventors: MDP inventors and collaborators of MDP inventors. The former are the inventors who are authors of a patent belonging to the MDP; the latter, instead, are inventors who are coauthors in a patent together with a MDP inventor, that is to say an inventor who in the past was author of one of the MDP patents. In this second case, the patent registered does not belong to the MDP. To simplify our setting, the two categories will be mutually exclusive, that is to say that a MDP inventor cannot be classified as a MDP collaborator. Figures 6 and 7 provide a summary of our data for MDP inventors and collaborators. Figure 6 reports the share of MDP inventors in winning and losing counties with respect to the year of entry (announcement year, left panel) and the year of the opening of the plant (right panel). As it can be seen, reassuringly for our identification assumption, the share of MDP inventors in the winning and losing county is extremely similar in the years before the entry announcement but it is in most of the cases larger in the winning county after the entry announcement is made. It is worth noting that the share of inventors who patented at least once for the MDP increases in winning counties a few years before the opening of the plant. This might be related to two different dynamics: first of all, inventors might be employed by the MDP *extra moenia*, for example through financing research and development in research institutions; a second possibility is that the opening of the MDP is related to the inflow of MDP inventors from other countries which causes the observed increase in the share of MDP inventors. A similar dynamic can be observed for collaborations in figure 7: in the period before entry winning and losing counties show a remarkable similarity and the proportions of collaborations picks up after the plant opening, remaining at a higher level for the rest of the sample. Both these elements might be related to the technology diffusion of the MDP: an inventor who works for the MDP will have better knowledge of the technology internally used by the MDP and he will have more opportunities to cite patents suitable for his research interest. Similarly, by coauthoring a patent with a MDP inventor, a local inventor might collect relevant information concerning the MDP technology which might be used later on for the production of further patents. In table 11, we look at whether inventors in winning and losing

counties directly worked or collaborated with a MDP inventor before the citing the MDP firm patents for the first time. These very direct channels of contact between the local inventor and the entering firm technology accounts only for a small share of the inventors citing the MDP (12% of citing inventors). It is, however, noteworthy that among the 23 stayer inventors who cited the entering firm patents after the entry 16% worked for the firm prior citing the its patents after the plant opened in the county. This suggests that indeed the common labour market for inventors is indeed playing a role in the technology diffusion although it remains rather marginal in terms of number of inventors. All these elements suggest that other channels which might be more complex to quantitatively identify in absence of richer data (e.g supply/demand relationship between the entering firm and the firm where local inventors work, competition or more simply a common environment which allows for informal contact and information transmission).

## 8 Heterogenous Effects

The effect of the plant opening may vary across several dimensions depending on characteristics of the patent, of the entering firm and of the individuals or companies which could produce patents. In this section, we explore several possible margins. First, we focus on the characteristics of patents that is to say how many years before the entry decision the patent was registered, its quality and its technology class. Then, we move to the difference across the sector of the entering firm. Finally, we evaluate if the effect is stronger for inventors employed by companies or for standalone inventors, that is inventors who do not work for a firm and whose patent does not belong to any firm when the patent application is filed. Finally, we check the sensitivity of our estimates to the exclusion of single cases.

### 8.1 Patents

#### Old vs new patents

The effect of proximity to patent citations could depend on how recent the patent is with respect to the plant opening of the firm. More recent patents represent the latest technology development and could hence be more valuable for inventors for the development of new patents. In addition, more recent patents could be less widely known and proximity could then play a more important role in the technology diffusion. We divide patents into 3 categories depending on the time distance between their application year and year of the decision to enter. We group all the patents with application years at least 16 years prior to the entry in one category. The other categories encompass the following range of application years: 1-6 and 7-15 years prior to MDP entry. We run our preferred specification on each of these groups and plot the resulting effects in Figure 8.

Results are consistent with our expectations. The estimated coefficient is close to 0 for very old patents (application year more than 15 years before the year of entry). Hence, our results seems to be driven by patents with more recent application dates.

## Patent quality

Another possibility is that the effect varies according to patent quality. In order to proxy for patent quality, we use the fixed effect methodology developed in Hall, Jaffe and Trajtenberg (2001). We count the number of citations that each patent received in the first five years after its application year and then regress this variable on a full set of year fixed effects.<sup>13</sup> We then normalize citations received by patent with application year  $t$  by the corresponding fixed effect and get a normalized index of citations for the patent. We identify two groups of high quality patents: the first one is composed of patents with an index value above the median; the second is composed of patents with an index value above the 90<sup>th</sup> percentile of the distribution of the quality index. We estimate our specification with year-case fixed effects, and include dummies for these two categories and interaction terms.<sup>14</sup>

The first 3 columns of Table 12 report estimates for the model with year-case fixed effects: column (1) includes a dummy for patents with quality above the median (*Above median*) and its interaction terms; column (2) a dummy for patents above the 90<sup>th</sup> percentile (*Above 90<sup>th</sup>*) and its interaction term; column (3) includes both dummies and interaction terms. The same structure is followed for the next 3 columns using our preferred specification that includes patent-county-case fixed effects. The analysis shows that there seem to be no differential effect for patents in the top half of the quality distribution. However, there is some evidence that patents in the top decile of the distribution experience a much larger increase in citations: the effect is large but statistically significant only at 10%. This result can be explained by the fact that the highest quality patents represent more relevant technology development and hence local inventors can benefit more by becoming acquainted with these patents rather than the average patent.

---

<sup>13</sup>Results are similar if we use interactions between years and technology class.

<sup>14</sup>Note that we estimate this specification using only patents with application year at least 5 years before opening. This avoids a possible tautological definition of the quality indicators as quality measures would also be based on citations received after entry.

## 8.2 Technology class

Finally, we focus on the pattern of citations according to technology class.<sup>15</sup> We expand our dataset to obtain a panel of patent, technology class, case and county code. The dependent variable is modified accordingly and is now a dummy equal to one if patent  $p$  receives at least one citation from patents produced in county  $f$  in technology class  $j$  and in year  $t$ . Similarly, the variable *Other citations* will now represent the number of other citations received in the US by patents in technology class  $j$  excluding citations in counties involved in case  $c$ .

This structure leads to a eightfold increase in the number of observations and can lead to a change in the estimated coefficients in our preferred specification.<sup>16</sup> We estimate the following equation:

$$\begin{aligned} Citation_{pijtcf} = & \alpha + \beta_1 Post_{tc} + \beta_2 Winner_{cf} + Winner_{cf} X Post_{tc} + \\ & + \delta X_{pt} + \gamma Z_{ft} + \mu_{pcf} + \varepsilon_{pijtcf} \end{aligned} \quad (4)$$

where  $i$  is the technology class of the MDP firm's patent and  $j$  is the technology class of the patents citing patent  $p$ . The rest of the notation is unaltered with respect to equation 2. This regression model is estimated separately for  $i = j$  and for  $i \neq j$ . Results are reported in table 13.

Column (1) reports our basic specification pooling together citations from any technology class. This is equivalent to our regression in Section 5 with the additional dimension of the technology class of the citing patents. It can be seen that coefficients are largely consistent with previous analysis and the effect is positive and statistically significant. The coefficient represents a 0.0046 percentage point increase in the probability that a patent  $p$  is cited in year  $t$  by patents in technology class  $j$ . This represent a 76% increase with respect to the baseline probability of 0.00607 percentage points in the losing county in the period after entry. In column (2) we move to the analysis of citations from patents in the same technology class of the MDP firm patent. Our coefficient of interest is substantially larger and statistically significant at the 1% significance level. It represents a 52% increase in the baseline probability of citation (probability of being cited in the losing county in the *same* technology class). Column (3) finally reports the effect for citations from patents in a different technology class. The coefficient is

---

<sup>15</sup>Technology classes are: Biotech, Chemical, Architecture, Video/Security, Communications, Semiconductors, Construction, and Manufacturing

<sup>16</sup>Previously, the dependent variable took on the value of 1 if it received 1 citation regardless of the technology class of the citing patent. In this new specification, a patent can receive citations from patents belonging to 8 different types of technology class. The dependent variable takes on a value of 1 for technology classes that cite the patent (that is, if 3 different technology classes cite the patent, those patent observations will take on a value of 1, while the rest of the 5 patent observations will take on a value of 0).

still positive and statistically significant; although it is smaller in absolute terms, it represents a larger proportional effect (+122%) if compared to the baseline citation probability (probability of being cited in the losing county in *another* technology class). This finding could be explained by the fact that the baseline citation probability from patents in the same technology class is already large and so an extremely strong knowledge or technology shock is required for a high relative increase. Additionally, it is reasonable that inventors had higher pre-existing knowledge about MDP patents in the same technology class of the patents they produced. Thus the effect on patents from other technology classes might be proportionally larger.

### 8.3 Types of MDP firms

In this section we check if technology diffusion is related to the sector of the MDP firm. We estimate our preferred specification dividing cases according to the sector of the entering firm. Results are in Table 14. Column (1) reports basic estimates for the full sample, whereas column (2) and (3) divide the sample in manufacturing and non-manufacturing firms. Our coefficient of interest is positive and statistically significant in both cases and larger for non-manufacturing firms. These results suggests that technology diffusion is not specific to the case of manufacturing firms and the effects is indeed stronger for Service firm. This could be related to the fact that patents from non-manufacturing companies (for example, companies which develop electronics technology) can be applied in a wider range of settings.

### Influence of single cases

Given the high variability across cases in terms of the number of observations, and that the number of citations to MDP patents is small, it is a legitimate concern that our results might be driven by outlier cases. We address this concern by evaluating the influence of a single case on our estimates. In order to assess the effect of each case on our estimates, we run our preferred specification (the one including patent-county-case fixed effects) iteratively excluding in each round a different case. Then we plot the resulting coefficients together with their confidence interval at 90% and 95% in Figure 9. Results suggest that our estimates are robust to the exclusion of single cases: the effect remains positive and statistically significant. Most of the cases only have a marginal impact on the coefficient and only a few of them appear to be relatively important.

## Type of citation: firm or individual

In this section, we decompose citations to patents by type of assignee of the *citing* patent. A patent can be registered under the name of a company which then own legal right over the patent or by standalone inventors who register the patent as private individuals. We would expect the effect to be stronger for inventors working for firms. This could be due to several possible explanations. First of all, if the technology diffusion is mediated through demand and supply relations, inventors employed by firms should be more affected by the opening of the new plant. Second, inventors working for firms might have more funds and input available which might allow them to produce more patents. Third, a similar pattern could be attributed to sorting; inventors who are developing new technologies related to the MDP firm might be more attractive for firms which would then have higher incentives to hire them to acquire property rights on the new patents and reap their economic benefits.

We run our preferred specification separately for these two types of assignee with patent fixed effects as in column 5 of table 4. Results are reported in table 15. Column (1) presents the baseline specification in which all citing patents are bundled together. The estimated coefficient for the interaction term between the winner dummy and the post opening dummy is 0.33 and strongly statistically significant. In column (2) we only keep citing patents coming from individuals employed by companies, universities or research centres and results look very similar. The effect is still positive but not statistically significant for inventors not employed any firm or institution (column (3)). These results are consistent with our predictions.

## 9 Conclusions

This paper evaluates the degree of knowledge diffusion from large entering plants to the local economy. The major challenge for our analysis is the endogeneity of the location of the new plant. We use revealed ranking of preferred locations by large companies for the opening of a new plant to identify a valid counterfactual to the winning county as in Greenstone, Hornbeck, and Moretti (2010). This dataset, originally developed by Greenstone and Moretti (2004) reports both the county chosen for the new plant opening, the "*winner*", and the runner up U.S. counties, the "*losers*". Therefore, our identification assumption is that citations to pre-existing patents of the entering firm would have been similar in the winning and losing county without the opening of the new plant.

Using patent citations as a proxy for knowledge flows, we find that patents of the entering firm are 68% more likely to be cited in the winning county rather than in the losing counties after entry. There is no evidence of differences in trends in citations between the winning and losing county before the firm entry. The increase in citations is driven by the plant opening itself: no

difference in citations is observed between the year of the entry announcement "announcement year" and the plant opening year "opening year". In the winning county the inventors citing the MDP appear to be at an earlier stage of their career and to be of higher quality. This suggests that the entry mostly benefits the most flexible and successful inventors. The increase in the probability of citations is stronger (i) for more recent patents of the MDP firm and (ii) from patents that share the same technology class as the technology class of patents of the entering firms. Patent quality of the MDP firm does not seem to play a role in citation differences.

The present work could be further developed in several directions. A particularly promising one would be to track inventors over time. This will allow us to answer (i) whether winning counties attract more and better inventors from other counties; (ii) whether the increase in citations is driven by local inventors or rather migrating inventors; (iii) whether incumbent inventors start direct collaborations with MDP inventors; (iv) whether incumbent inventors increase the quality of citations made or whether there is full crowding out of other equally relevant citations.

## 10 References

Abramovsky, L., R. Griffith and G. Macartney and H. Miller, (2008). "The location of innovative activity in Europe," IFS Working Papers W08/10, Institute for Fiscal Studies

Agrawal, A., D. Kapur, and J. McHale (2008). "How Do Spatial and Social Proximity Influence Knowledge Flows? Evidence From Patents Data". *Journal of Urban Economics*, vol. 64, pp. 258-269.

Arnold, J., and B. Javorcik, (2009). "Gifted Kids or Pushy Parents? Foreign Direct Investment and Plant Productivity in Indonesia". *Journal of International Economics*, Vol. 79, No 1.

Blomstrom, M. and A. Kokko, (1996). "Multinational corporations and spillovers", CEPR Discussion Papers 1365, CEPR Discussion Papers

Blomstrom, M. and A. Kokko, (1997). "Regional integration and foreign direct investment", CEPR Discussion Papers 1659, CEPR Discussion Papers

Blomstrom, M. and A. Kokko, (1998). "Multinational corporations and spillovers", *Journal of Economic Surveys*, Vol. 12, No. 3: 247-277

Branstetter, L. (2006). "Is Foreign Direct Investment a Channel of Knowledge Spillovers? Evidence from Japan's FDI in the United States", *Journal of International Economics*, Vol. 68, pp. 325-344.

Caballero J. and Jaffe A. (1993). "How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth," NBER Chapters, in: NBER Macroeconomics Annual 1993, Volume 8, pages 15-86 National Bureau of Economic Research, Inc.

Crespo, N. and M. P. Fontoura (2007). "Determinant Factors of FDI Spillovers - What Do We Really Know?", *World Development*, Vol. 35, No. 3: 410-425.

Duguet E. and Garvie M. (2005). "How well do patent citations measure flows of technology? Evidence from French innovation surveys," *Economics of Innovation and New Technology*, Taylor and Francis Journals, vol. 14(5), pages 375-393, July.

Ellison. G., Glaeser E. (1997) "Geographic Concentration in US manufacturing Industries: A Dartboard Approach", Journal of Political Economy, October, Vol. 105, No 5: 889-927

Fons-Rosen C., (2010). "Knowledge Flows Through FDI: the Case of Privatizations in Central and Eastern Europe", mimeo, 2010.

Gorg H. and Greenaway D., (2004). "Much Ado about Nothing? Do Domestic Firms Really Benefit from Foreign Direct Investment?", The World Bank Research Observer, Vol. 19, No. 2: 171-197.

Glaeser E., (1999). "Learning in Cities", Journal of Urban Economics, Vol. 46 (September): 254 -77.

Glaeser E. and Kohlhase J., (2003). "Cities, Regions and the Decline of Transport Costs", Papers Regional Sci. 83 (January): 197-228.

Glaeser E. and Gottlieb J. (2009) "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States" Journal of Economic Literature, American Economic Association, vol. 47(4), pages 983-1028, December

Gorg H. and Strobl E., (2001). "Multinational Companies and Productivity Spillovers: A Meta-Analysis", Economic Journal, vol. 111, pp. 723-739

Greenstone M. and Moretti E., (2004). "Bidding for Industrial Plants: Does Winning a 'Million Dollar Plant' Increase Welfare?", National Bureau of Economic Research Working Paper No. 9844.

Greenstone M., Hornbeck R., and Moretti E., (2010). "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings", Journal of Political Economy, Vol. 118, No. 3: 536-598.

Griffith, R., R. Harrison, J. Van Reenen (2006). "How Special is the Special Relationship? Using the Impact of U.S. R&D Spillovers on U.K. Firms as a Test of Technology Sourcing", The American Economic Review, vol. 96, no. 5; pp. 1859-1875

Grossman G. and Helpman E., (1991). "Trade, Knowledge Spillovers, and Growth", Working Paper No. 3485. National Bureau of Economic Research, Cambridge, Massachusetts.

Hall B. , Jaffe A. and Trajtenberg M., (2005). "Market Value and Patent Citations", RAND Journal of Economics, The RAND Corporation, vol. 36, No 1: 16-38, Spring.

Hall B., Jaffe A. and Trajtenberg M., (2001). "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools", CEPR Discussion Papers 3094, C.E.P.R. Discussion Papers

Jaffe, A., M. Trajtenberg, and M. Fogarty (2000). "The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees", NBER Working Paper No. 7631

Jaffe, A., M. Trajtenberg, and R. Henderson (1993). "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations", Quarterly Journal of Economics, vol. 108, No. 3: 577-598.

Javorcik, B.S. (2004). "Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward productivity of domestic firms?", The American Economic Review, 94, No 3: 605-627, June.

Krugman P. (1991a). "Geography and Trade", Cambridge, MA: MIT Press.

Krugman, Paul (1991b). "Increasing Returns and Economic Geography", Journal of Political Economy, 99, 483-499.

Lai R., D'Amour A., Yu A., Sun Y., Torvik V. and Fleming L. (2014). "Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010)", Research Policy, Vol. 43, No. 6: 941-955.

Lipsey, R.E. (2001a). "Foreign direct investment and the operations of multinational firms: concepts, history and data", NBER Working Papers 8084, National Bureau of Economic Research.

Lipsey, R.E. (2001b). "Foreign direct investors in three financial circles", NBER Working Papers 8665, National Bureau of Economic Research.

Lychagin, S., J. Pinkse, M. E. Slade and J. Van Reenen (2010). "Spillovers in Space: Does Geography Matter?", CEP Discussion Paper No 991 Marshall, A. (1890, 1920). Principle of Economics, Macmillan.

Lucas, R. (1988). "On the Mechanics of Economic Development", *Journal of Monetary Economics*, Vol. 22, No. 1: 3-42.

Marshall, A. (1980). "Principles of Economics", London: Macmillan and Co.

McCarty Moretti, E. (2004). "Estimating the External Return to Higher Education: Evidence from Cross-Sectional and Longitudinal Data", *Journal of Econometrics* 120 (July-August): 175-212.

Ottaviano, G. and D. Puga (1998). "Agglomeration in the global economy: A survey of new economic geography", *World Economy*, Vol. 21, Oxford (UK), 1998, pp. 707-731.

Santos-Silva, J.M.C. and Tenreyro S. (2006). "The Log of Gravity", *The Review of Economics and Statistics*, November 2006, Vol. 88, No. 4: 641-658.

Saxenian, A. (1994). "Regional Advantage: Culture and Competition in Silicon Valley and Route 128", Cambridge, MA: Harvard University Press.

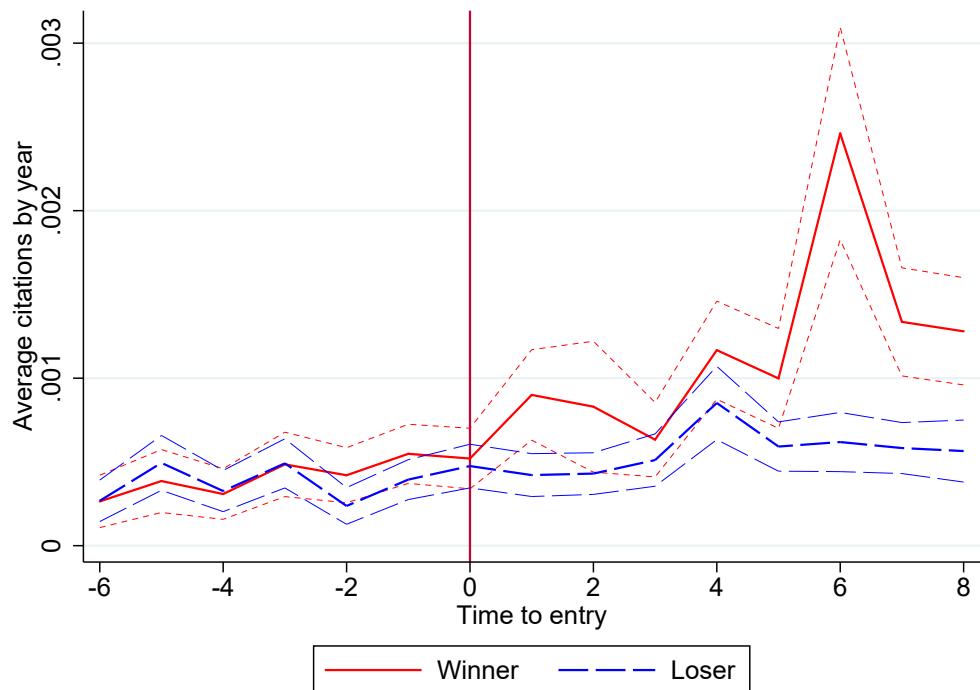
Thompson, P. (2006). "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner- Added citations", *The Review of Economics and Statistics*, May 2006, Vol. 88, No. 2: 383-388.

Trajtenberg, M. (1990). "A Penny for Your Quotes: Patent Citations and the Value of Innovations", *The Rand Journal of Economics*, Spring 1990, Vol. 21, No. 1: 172-187.

Verspagen, B. and Schoenmakers, W., (2004) "The spatial dimension of patenting by multinational firms in Europe", *Journal of Economic Geography*, Oxford University Press, Vol. 4, No. 1: 23-42, January.

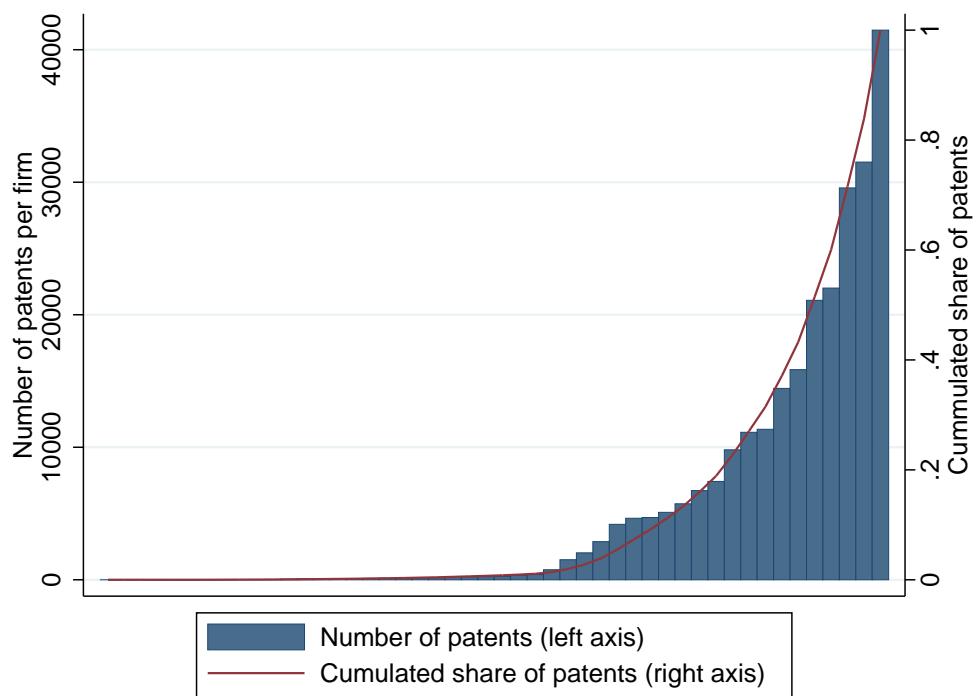
## Figures

Figure 1: Citations by year to entering firm patents



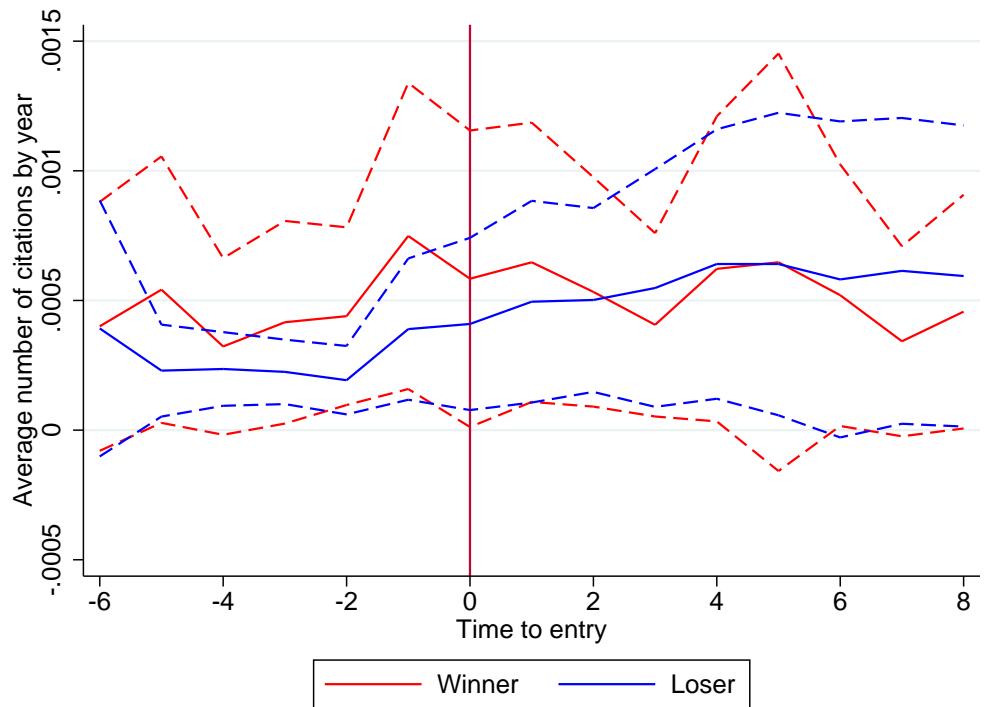
Note: Average number of citations in winning (red) and losing counties (blue). Zero corresponds to the year of the announcement of entry in the winning county. Dashed line is corresponds to confidence interval at 95% confidence level

Figure 2: Number of patents by MDP firms



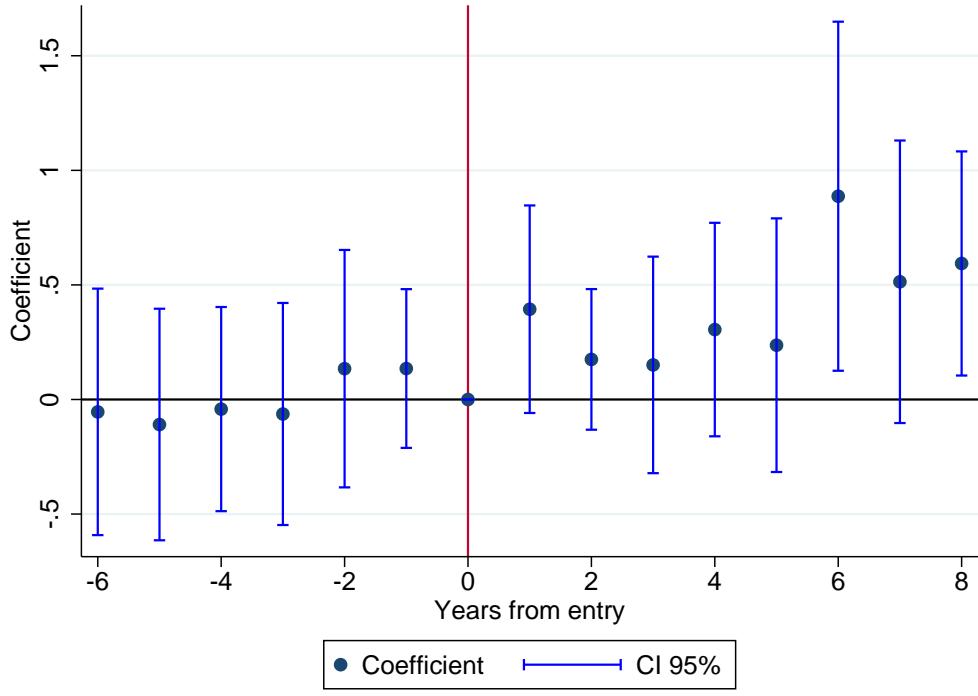
Note: Number of patents produced by MDP firms in the sample before the entry decision. Left axis reports the number of patents (histogram) whereas the right axis reports the cumulative distribution function (red line). Patents belonging to multiple firms are counted for each assignee

Figure 3: Citations to local patents by MDP patents



Note: Number of citations by year per patent produced by local inventors in winning and losing county before entry year. Red line is citations to patents from winning county and blue line represent citations to patents from losing county. Dashed line correspond to 95% confidence interval.

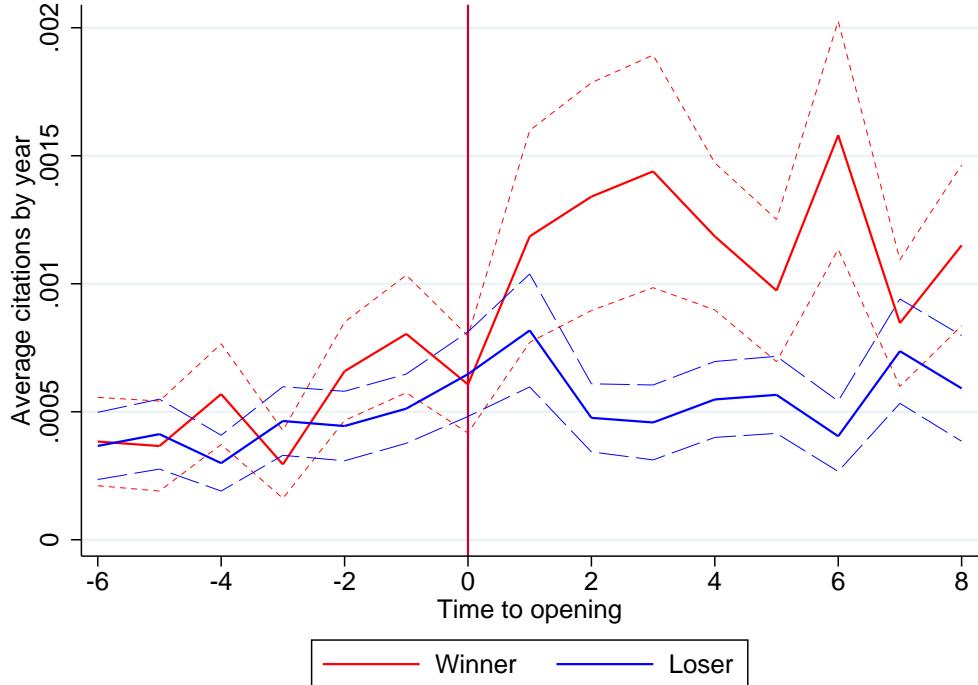
Figure 4: Time pattern of the effect of entry



Note:

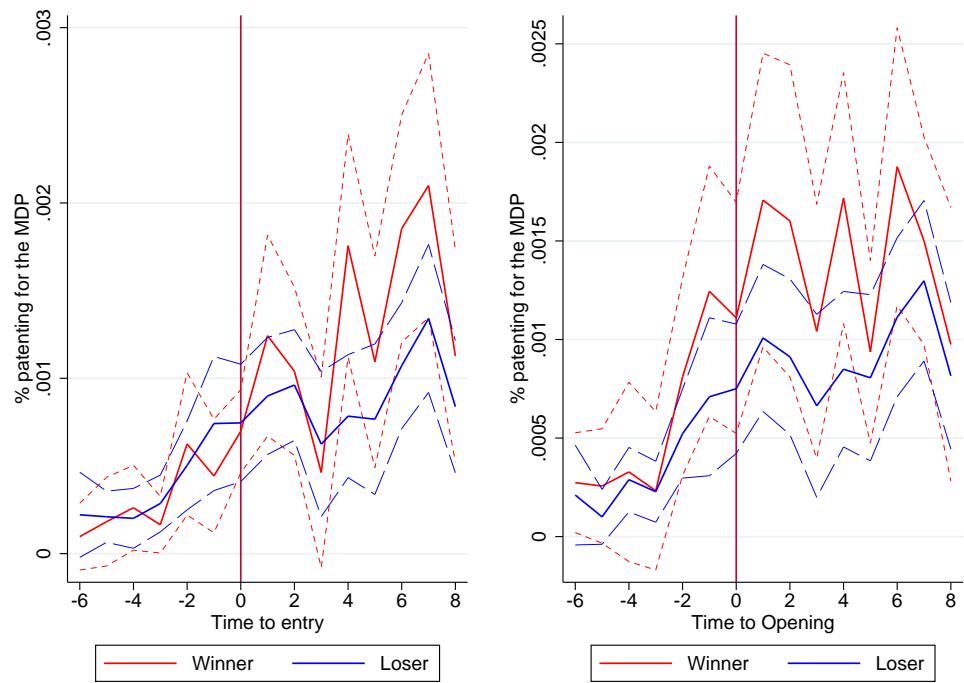
Coefficients from regression of patent being cited at least once by a patent registered in county  $c$  at time  $t$  on controls in preferred specification and a dummy for the year of the announcement of entry and its leads and lags. Coefficients reported together with 95% confidence interval

Figure 5: Average number of citations and time to plant opening



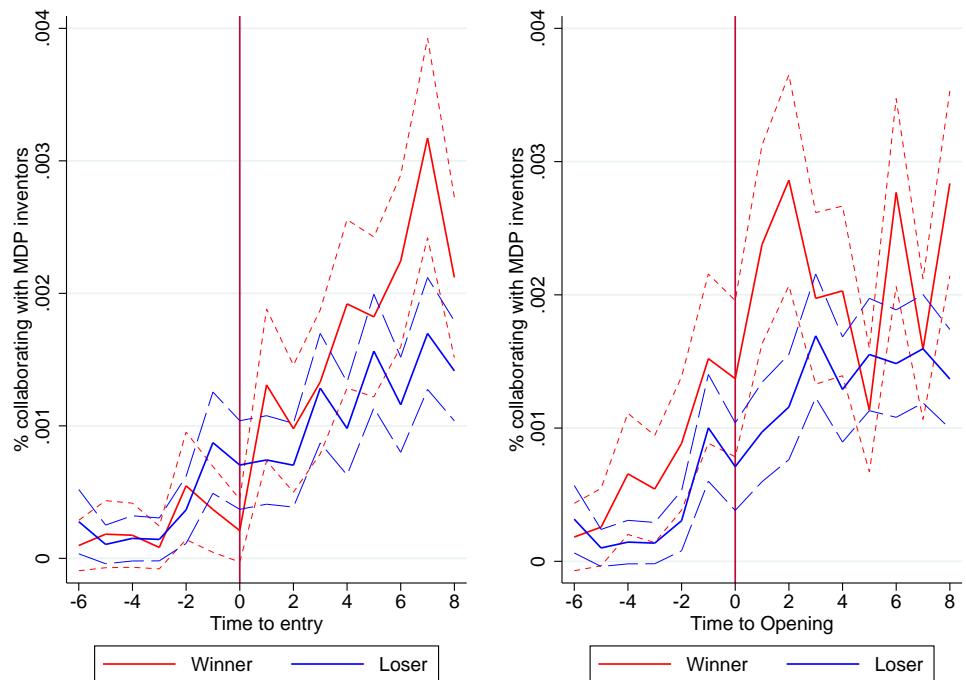
Note: Average number of citations in winning (red) and losing counties (blue). Zero corresponds to the year of opening of the plant in the winning county. Dashed line is corresponds to confidence interval at 95% confidence level.

Figure 6: Share of inventors in Winning and Losing counties patenting for the MDP firm



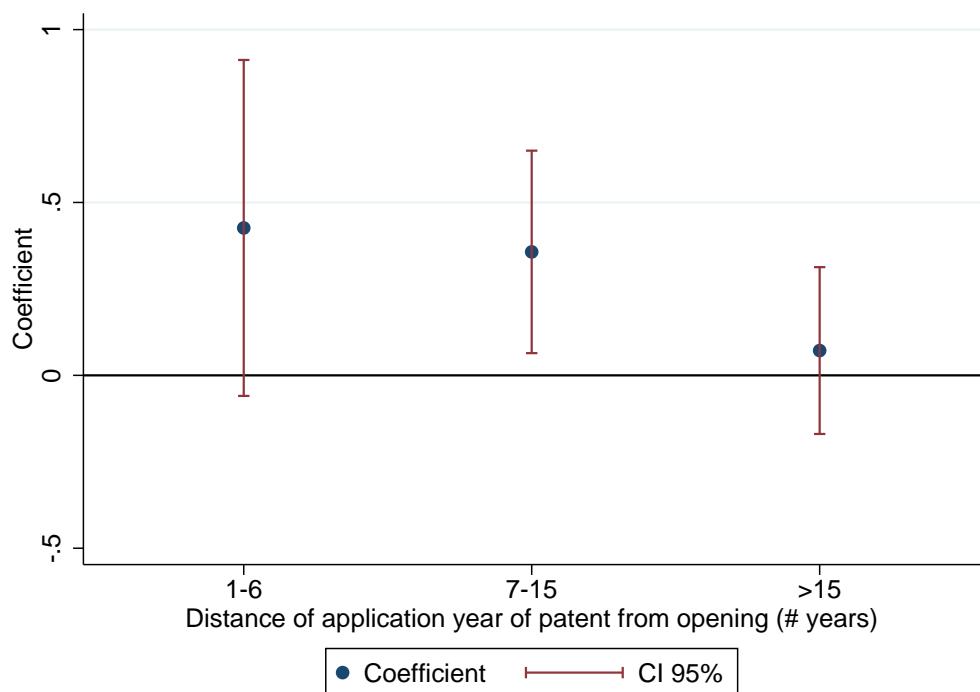
Note: Share of inventors in winning (red) and losing counties (blue) who are authors of a patent for the MDP firm. This implies that the MDP firm is the registered *assignee* for the patent. Left panel reports the series with respect to the announcement date whereas left panel reports series with respect to the opening date of the plant. Dashed line is corresponds to confidence interval at 95% confidence level.

Figure 7: Share of inventors in Winning and Losing counties patenting for the MDP firm



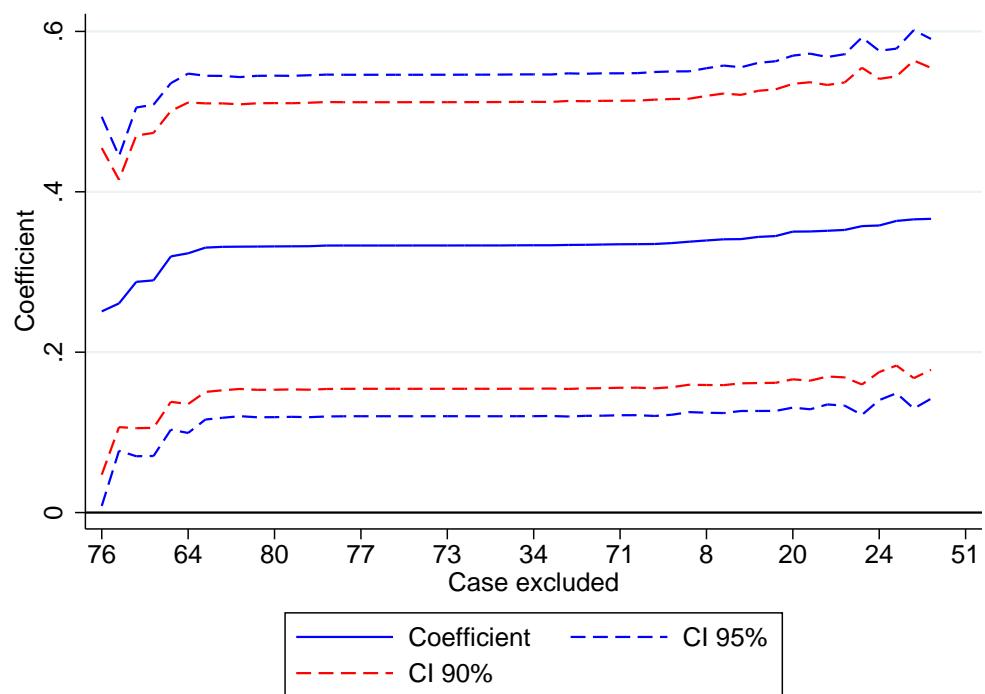
Note: Share of inventors in winning (red) and losing counties (blue) who are coauthors with MDP firm inventor. A inventor is defined as a MDP inventor if he was among the authors of a MDP patent in the past. This implies that the MDP firm is the registered *assignee* for the patent. Left panel reports the series with respect to the announcement date whereas left panel reports series with respect to the opening date fo the plant. Dashed line is corresponds to confidence interval at 95% confidence level.

Figure 8: Application distance from entry and effect on citations



Note: Coefficients of linear probability model with patent/case/county fixed effects and classical controls together with 95% confidence interval. Patents divided in 3 groups according to the number of years elapsed between their registration and the year of the plant opening: patents registered more than 15 years before the opening; patents registered between 15-7 years before the opening and patents registered 6 or less years before the plant opening in the winning county. Standard errors clustered at county level.

Figure 9: Exclusion of single cases and effect of opening on citations



Note: Coefficients of linear probability model with patent/case/county fixed effects and classical controls together with 95% confidence interval. Coefficients obtained from regression including all cases but *case x*. Standard errors clustered at county level.

## Tables

Table 1: Total citations in winning and losing counties before and after the entry announcement.

	Before	After
Loser: citations	195	563
Loser: citations per county	2.436	7.038
Winner: citations	139	720
Winner: citations per county	2.780	14.400

Note: Total number of citations in winning and losing counties before and after the entry date of the MDP. Period “Before” includes years from the year 6 years before the entry up to the year before entry. Period “After” includes years from the year of entry up to 8 years after the entry announcement of the firm.

Table 2: Descriptive statistics

Number of counties	
Winning <sup>1</sup>	50
Losing	80

Decomposition of MDP starting year	
1982	3
1983	2
1984	2
1985	4
1986	4
1987	5
1988	7
1989	3
1990	4
1991	4
1992	6
1993	5
Total	49

Note: 1 In case of a firm located on the border of two counties, we set those two counties as winners.

Table 3: Effect of location on probability of citation at the case-county-year level

VARIABLES	(1) Count	(2) Count	(3) Count	(4) Count	(5) Count
Winner	0.050 (0.226)	0.205 (0.214)	0.190 (0.511)	-1.174 (2.052)	
Post	0.358** (0.149)	0.066 (0.272)	-0.322 (0.226)		-0.322 (0.219)
WinnerXPost	0.715* (0.414)	0.702** (0.322)	0.724** (0.335)	0.825** (0.361)	0.724** (0.323)
Other citations		0.002*** (0.001)	0.003** (0.001)		0.003** (0.001)
log(ActivePopulation)		0.582*** (0.169)	0.048 (0.790)	-0.909 (2.588)	0.032 (0.770)
Year FE	N	Y	Y	Y	Y
County FE	N	N	Y	Y	Y
Case FE	N	N	Y	Y	Y
YearXCase FE	N	N	N	Y	N
CountyXCase	N	N	N	N	Y
Observations	2,055	2,055	2,055	2,055	2,055

Note: OLS model where at county case level. Dependent variable is the number of citations received by patents of the entering firm before entry. Standard errors are clustered at county level. Level of significance: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 4: Effect of location on probability of citation at patent level

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Winner	0.045 (0.189)	-0.005 (0.186)	-0.527*** (0.071)	-0.029 (0.556)	
Post Announcement	0.140* (0.075)	0.188 (0.224)	-0.165 (0.128)		-0.135 (0.126)
WinnerXPost Announcement	0.327** (0.153)	0.411*** (0.137)	0.385*** (0.128)	0.377*** (0.097)	0.337*** (0.124)
log(ActivePopulation)		0.490*** (0.134)	0.397 (0.471)	0.280 (0.727)	0.475 (0.478)
Other citations		1.099*** (0.351)	1.094*** (0.351)	1.089*** (0.349)	0.855*** (0.261)
Year FE	N	Y	Y	Y	Y
Year since Patent application FE	N	Y	Y	Y	Y
County FE	N	N	Y	Y	Y
Case FE	N	N	Y	Y	Y
YearXCase FE	N	N	N	Y	N
PatentXCountyXCase FE	N	N	N	N	Y
Observations	2,505,649	2,505,649	2,505,649	2,505,649	2,505,649

Note: Linear probability model where the dependent variable is dummy equal to 1 if patent  $p$  receives at least one citation by patents produced in county  $f$  and year  $t$ . Coefficients multiplied by 1000. Citation frequency in the losing county is 0.000497. Standard errors are clustered at county level. Level of significance: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 5: Plant opening and citation probability

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Winner	0.045 (0.189)	-0.007 (0.187)	-0.567*** (0.078)	-0.221 (0.546)	
Post Announcement	0.093 (0.146)	0.196 (0.162)	-0.062 (0.124)		-0.039 (0.124)
Post Opening	0.070 (0.148)	0.036 (0.198)	-0.209** (0.096)		-0.207** (0.093)
WinnerXPost Opening	0.187 (0.280)	0.360 (0.245)	0.395** (0.173)	0.424** (0.166)	0.369** (0.184)
WinnerXPost Announcement	0.190 (0.222)	0.151 (0.199)	0.114 (0.157)	0.086 (0.112)	0.085 (0.170)
log(ActivePopulation)		0.496*** (0.135)	0.307 (0.443)	0.055 (0.708)	0.391 (0.446)
Other citations		1.100*** (0.351)	1.095*** (0.351)	1.089*** (0.349)	0.855*** (0.261)
Year FE	N	Y	Y	Y	Y
Year since Patent application FE	N	Y	Y	Y	Y
County FE	N	N	Y	Y	Y
Case FE	N	N	Y	Y	Y
YearXCase FE	N	N	N	Y	N
PatentXCountyXCase FE	N	N	N	N	Y
Observations	2,505,649	2,505,649	2,505,649	2,505,649	2,505,649

Note: Linear probability model where the dependent variable is dummy equal to 1 if patent  $p$  receives at least one citation by patents produced in county  $f$  and year  $t$ . Citation frequency in the losing county is 0.000497. Coefficients multiplied by 1000. Standard errors are clustered at county level. Level of significance: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 6: Citation probability and number of losing counties

VARIABLES	(1) Overall	(2) Only 1 losing	(3) 2 or more losing	(4) Overall
Post opening	-0.136 (0.100)	-0.138 (0.160)	-0.024 (0.152)	-0.116 (0.150)
WinnerXPost opening	0.333*** (0.109)	0.297* (0.156)	0.397** (0.181)	0.264* (0.147)
More than 2 losing X Post opening				-0.031 (0.117)
More than 2 losing X Winner X Post opening				0.137 (0.243)
log(ActivePopulation)	0.534 (0.395)	0.002 (0.708)	0.830* (0.478)	0.554 (0.384)
Other citations	0.726*** (0.199)	1.092** (0.431)	0.627*** (0.203)	0.726*** (0.199)
Year FE	Y	Y	Y	Y
Year since Patent	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Case FE	Y	Y	Y	Y
PatentXCountyXCase FE	Y	Y	Y	Y
Observations	2,602,894	1,124,686	1,478,208	2,602,894
Number of cases	49	32	17	49

Note: Linear probability model where the dependent variable is dummy equal to 1 if patent  $p$  receives at least one citation by patents produced in county  $f$  and year  $t$ . Column 1 reports estimates for the baseline specification. Column 2 restricts the sample to cases where only one losing county is present and Column 3 includes all the cases with more than 1 losing counties. Standard errors are clustered at county level. Level of significance: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 7: Citation probability and cases excluded

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Winner	1.535 (1.421)	2.406* (1.381)	1.568* (0.830)	0.509 (0.440)	
Post Announcement	-0.017 (0.187)	0.581 (0.678)	-0.032 (0.222)		-0.054 (0.213)
WinnerXPost Announcement	0.038 (0.173)	-0.012 (0.180)	-0.155 (0.103)	-0.157 (0.105)	-0.089 (0.090)
log(ActivePopulation)		1.028* (0.566)	0.801 (0.856)	0.221 (0.234)	0.678 (0.763)
Other citations		2.361 (1.864)	2.326 (1.848)	2.328 (1.849)	2.084 (1.547)
Year FE	N	Y	Y	Y	Y
Year since Patent	N	Y	Y	Y	Y
County FE	N	N	Y	Y	Y
Case FE	N	N	Y	Y	Y
YearXCase FE	N	N	N	Y	N
PatentXCountyXCase FE	N	N	N	N	Y
Observations	360,678	360,678	360,678	360,678	360,678

Note: Linear probability model where the dependent variable is dummy equal to 1 if patent  $p$  receives at least one citation by patents produced in county  $f$  in year  $t$ . Estimation of the baseline specification on the 15 excluded cases. Standard errors are clustered at county level. Level of significance: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8: Comparison of inventor characteristics in winning and losing counties

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Experience	Patents	Patents (w)	Cum patents	Cum patents (w)	Cum patents (w & log)	Patents (w) - Main tech	Cum patents (w) - Main tech
Winner	-0.147 (0.158)	-0.020 (0.016)	-0.035 (0.028)	-0.234 (0.201)	-0.569** (0.282)	-0.088* (0.049)	0.011 (0.018)	-0.005 (0.124)
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Case FE	Y	Y	Y	Y	Y	Y	Y	Y
YearXCase FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	193,829	193,829	193,829	193,829	193,829	193,829	193,829	193,829
SD dependent	4,889	1,038	1,908	6,494	9,313	1,404	474	4,559

Note: OLS regression for inventors located in winning and losing county in the six years prior to entry. Winner is a dummy taking value 1 if the inventor is located in the winning county. (w) identifies columns where patents were weighted by a quality index based on normalized citations. Main tech is the main technology class of the MDP firm defined as the technology class with the highest number of patents. Standard errors clustered at county level. Level of significance: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 9: Decomposition by type of inventors citing MDP patents in winning and losing counties

Type	Winning	Losing
<b>Stayers</b>	234	215
<i>Stayers: Cite before and after opening date</i>	34	22
<i>Stayers: Cite only before opening date</i>	78	97
<i>Stayers: Cite only after opening date</i>	122	96
<b>New Inventors</b>	249	267
<b>Immigrants</b>	52	59
<i>Immigrants: started before opening date</i>	36	40
<i>Immigrants: started after opening date</i>	16	19
<b>Quitters</b>	76	98
<b>Leavers</b>	9	6
<b>Total</b>	620	645

Note: Inventors considered only if citing MDP when they are in winning county. Type of Inventors: Stayers are inventors who were present in the winning county both before and after the entry of the MDP; New Inventors are inventors who only appeared in our dataset after entry of MDP in the winning county; Immigrants are inventors who were either i.) present in our dataset before entry date in another county and moved into winning county after entry or ii.) they only appeared in the dataset in some county after the entry date, and at one point they move to the winning county; Quitters are inventors who were present in the winning county before entry, but after entry are not in our dataset any more; Leavers are those inventors who were present in the winning county before entry of MDP, and after entry they moved to a different county.

Table 10: Comparison of Characteristics of Citing and Non-Citing Inventors in Winning Counties

	(1) Experience	(2) Cum. Patents	(3) Cum. Patents (Norm)	(4) Cum. Patents (Norm+Weights)	(5) Main Tech	(6) Not Main Tech
Citing vs Not Citing (All)	-0.985* (0.479)	3.156*** (0.820)	0.561*** (0.057)	1.273*** (0.221)	0.859*** (0.135)	0.051*** (0.024)
Citing vs Not Citing (Stayers)	-1.584*** (0.560)	5.688*** (1.361)	0.750*** (0.083)	1.924*** (0.449)	1.265*** (0.180)	0.081* (0.043)
Citing vs Not Citing (New Inventor)	-0.177* (0.087)	1.087*** (0.320)	0.465*** (0.089)	0.873*** (0.283)	0.564 (0.335)	0.039 (0.040)
Citing vs Not Citing (Immigrant: appear bef. opening)	0.998* (0.545)	5.420*** (2.570)	0.358*** (0.165)	0.784*** (0.363)	0.798*** (0.278)	-0.002 (0.016)
Citing vs Not Citing (Immigrant: appear aft. opening)	-0.062 (0.341)	-0.619 (0.538)	-0.107 (0.124)	0.439 (0.617)	0.884* (0.464)	-0.056* (0.030)
Citing vs Not Citing (Quitters)	-0.664 (0.462)	-0.134 (0.309)	0.106*** (0.034)	0.339 (0.218)	0.185* (0.095)	0.019 (0.027)
Citing vs Not Citing (Leavers)	-2.104 (1.244)	0.189 (0.939)	0.519 (0.375)	0.800*** (0.195)	0.360*** (0.060)	0.055* (0.027)
Year FE	Y	Y	Y	Y	Y	Y
Case FE	Y	Y	Y	Y	Y	Y
YearXCase FE	Y	Y	Y	Y	Y	Y

Note: Inventors considered only if citing MDP when they are in winning county. Type of Inventors: Stayers are inventors who were present in the winning county both before and after the entry of the MDP; New Inventors are inventors who only appeared in our dataset after entry of MDP in the winning county; Immigrants are inventors who were either i.) present in our dataset before entry date in another county and moved into winning county after entry or ii.) they only appeared in the dataset in some county after the entry date, and at one point they move to the winning county; Quitters are inventors who were present in the winning county before entry, but after entry are not in our dataset any more; Leavers are those inventors who were present in the winning county before entry of MDP, and after entry they moved to a different county. Standard errors clustered at case level. Level of significance: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 11: Citations, Collaborations and Patenting for the MDP in Winning counties

Type	Winning	Collaboration before citing	Patent for MDP before citing
Stayers: Cite before and after entry date	34	1	2
Stayers: Cite only before opening date	78	1	4
Stayers: Cite only after opening date	122	3 (0)	23 (19)
New Inventors	249	12	8
Immigrants: started before opening date	36	3	5
Immigrants: started after opening date	16	1	1
Total	535	21	43

Note: Inventors considered only if citing MDP when they are in winning county. For classification of inventors we remand to note in table 9. Inventors patent for the MDP if they are among authors of a patent belonging to the MDP based on the assignee file. They are collaborators of the MDP if they are coauthors of a patent and one of the coauthors was among the authors of a MDP patent in the past. The current patent should not belong to the MDP. Number in parentheses for *Stayers: cite only after the opening date* report the number of inventors whose collaboration or patenting year is after the opening date of the plant.

Table 12: Citation probability and quality of patents

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Winner	0.102 (0.416)	0.092 (0.386)	0.096 (0.410)			
WinnerXPost opening	0.215** (0.103)	0.222*** (0.082)	0.212** (0.104)	0.205** (0.093)	0.208** (0.088)	0.204** (0.093)
Above Median	0.371*** (0.134)		0.175*** (0.061)			
Above MedianXWinner	-0.110 (0.208)		-0.007 (0.130)			
Above MedianXPost Opening	-0.176* (0.100)		-0.055 (0.047)	-0.168* (0.098)		-0.050 (0.044)
Above MedianXWinnerXPost Opening	0.188 (0.216)		0.022 (0.166)	0.177 (0.230)		0.009 (0.177)
log(ActivePopulation)	0.286 (0.491)	0.282 (0.495)	0.283 (0.495)	0.583 (0.373)	0.591 (0.377)	0.599 (0.375)
Other citations	0.874*** (0.257)	0.857*** (0.255)	0.855*** (0.255)	0.659*** (0.186)	0.660*** (0.186)	0.660*** (0.186)
Above 90 <sup>th</sup> pc		1.174** (0.530)	1.081** (0.513)			
Above 90 <sup>th</sup> pcXWinner		-0.580 (0.648)	-0.577 (0.610)			
Above 90 <sup>th</sup> pcXPost Opening		-0.681 (0.454)	-0.651 (0.450)		-0.649 (0.451)	-0.621 (0.448)
Above 90 <sup>th</sup> pcXWinnerXPost Opening		0.886* (0.528)	0.874* (0.493)		0.883 (0.545)	0.878* (0.503)
Post opening				-0.066 (0.095)	-0.088 (0.090)	-0.066 (0.095)
Year FE	Y	Y	Y	Y	Y	Y
Year since Patent	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y
Case FE	Y	Y	Y	Y	Y	Y
YearXCase FE	Y	Y	Y	N	N	N
PatentXCountyXCase FE	N	N	N	Y	Y	Y
Observations	2,151,198	2,151,198	2,151,198	2,151,198	2,151,198	2,151,198

Note: Linear probability model where the dependent variable is dummy equal to 1 if patent  $p$  receives at least one citation by patents produced in county  $f$  and year  $t$ . Above median identifies patents in the top half of the quality distribution of MDP patents. Above 90<sup>th</sup> pc identifies patents in the top decile of the quality distribution. Quality defined as normalized index of citations in the first five years after the patent application. Standard errors are clustered at county level. Level of significance: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 13: Citation probability and technology class

VARIABLES	(1) Overall	(2) Same tech	(3) Different tech
Post Opening	-0.016 (0.013)	-0.035 (0.075)	-0.014 (0.008)
WinnerXPost Opening	0.046*** (0.014)	0.168** (0.071)	0.029*** (0.010)
log(ActivePopulation)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Other citations	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
Year FE	Y	Y	Y
Year since Patent	Y	Y	Y
County FE	Y	Y	Y
Tech FE	Y	Y	Y
Case FE	Y	Y	Y
PatentXCountyXCase FE	Y	Y	Y
Mean in losing	0.0607	0.321	0.0236
Observations	20,823,152	2,602,894	18,220,258

Note: Linear probability model where the dependent variable is dummy equal to 1 if patent  $p$  receives at least one citation by patents produced in county  $f$  and year  $t$ , and belonging to technology class  $j$ . Column 1 reports estimates for the baseline specification. Column 2 restricts the sample to cases where only one losing county is present and column 3 includes all the cases with more than 1 losing counties. Coefficients multiplied by 1000. Standard errors are clustered at county level. Level of significance: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 14: Citation probability and firm type

VARIABLES	(1) Overall	(2) Manufacturing	(3) Not manufacturing
Post opening	-0.136 (0.100)	-0.119 (0.095)	-0.183 (0.228)
WinnerXPost opening	0.333*** (0.109)	0.216** (0.098)	0.842*** (0.112)
log(ActivePopulation)	0.534 (0.395)	0.408 (0.602)	0.247 (1.275)
Other citations	0.726*** (0.199)	1.049*** (0.242)	0.495** (0.225)
Year FE	Y	Y	Y
Year since Patent	Y	Y	Y
County FE	Y	Y	Y
Case FE	Y	Y	Y
PatentXCountyXCase FE	Y	Y	Y
Observations	2,602,894	1,991,938	610,956

Note: Linear regression model (OLS) using as dependent variable the a dummy equal to 1 if patent  $p$  received at least one citation from patents produced county  $f$  in year  $t$ . Panel of patents starting from minimum between year of application for patent or six years before entry up to 8 years after entry. Patents can then have a minimum of 10 observations (patent produced at -1) per county and a maximum of 15 observations (patent produced at -6 or before). Analysis using only patents with application year at least 5 years before firm entry in the county. Standard errors clustered at county level. Level of significance: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 15: Citation probability and type of citation

VARIABLES	(1) Overall	(2) Firms	(3) Individuals
Post opening	-0.136 (0.100)	-0.114 (0.089)	-0.028 (0.023)
WinnerXPost opening	0.333*** (0.109)	0.283*** (0.103)	0.055 (0.045)
log(ActivePopulation)	0.534 (0.395)	0.642* (0.365)	-0.119 (0.143)
Other citations	0.726*** (0.199)	0.685*** (0.189)	0.046** (0.021)
Year FE	Y	Y	Y
Year since Patent	Y	Y	Y
County FE	Y	Y	Y
Case FE	Y	Y	Y
YearXCase FE	N	N	N
PatentXCountyXCase FE	Y	Y	Y
Observations	2,602,894	2,602,894	2,602,894

Note: Linear probability model where the dependent variable is dummy equal to 1 if patent  $p$  receives at least one citation by patents produced in county  $f$  and year  $t$ . *Individuals* refers to the subset of patents with the assignee being an individual. *Firms* includes all other categories of patent assignees, including firms, research centers, etc. Standard errors are clustered at county level. Level of significance: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## Appendix A Construction of panel of MDP patents

We construct a panel for each MDP patent. To illustrate, suppose the MDP firm is considering whether to open a plant in county A or county B. Also, assume a patent  $p$  was produced by the MDP in year  $t$ , which took place before the entry decision. Our dataset will have a panel from year  $t$  onwards for patent  $p$  in both county A and county B. The panel for each patent finishes 8 years after the MDP entry. For patents with application date between the MDP's entry decision and 6 years prior to that, we keep the full set of years. For patents with applications dates 7 or more years prior to the MDP entry, we start counting 6 years prior to the MDP entry. Observations are uniquely identified by the patent identifier, county, year and case considered.

## Appendix B Number of citations and Poisson Estimation

Results are not very different with the number of citations as dependent variable (Table B1). Also in this case, the coefficient in our preferred model (0.00049) is small in absolute terms but large if we consider that patents receive on average 0.000561 citations in the losing county in the same period. This implies that entry leads to a 87.34% increase in the number of citations in the winning county with respect to the losing county (column 5). Coefficients for other controls are consistent with previous results. Similarly, results also hold with the Poisson estimator (see Tables B2 and B3). Poisson estimates are obtained using the procedure developed by Santos-Silva and Tenreyro (2006)

Table B1: Effect of location on number of citations

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Winner	0.049 (0.199)	-0.011 (0.207)	-0.925*** (0.111)	-0.305 (1.015)	
Post Announcement	0.192* (0.101)	0.173 (0.258)	-0.251 (0.178)		-0.204 (0.171)
WinnerXPost Announcement	0.515** (0.231)	0.623*** (0.201)	0.598*** (0.201)	0.601*** (0.155)	0.495*** (0.178)
log(ActivePopulation)		0.568*** (0.151)	0.503 (0.747)	0.099 (1.316)	0.822 (0.702)
Other citations		1.851*** (0.696)	1.849*** (0.701)	1.842*** (0.699)	1.274*** (0.435)
Year FE	N	Y	Y	Y	Y
Year since Patent application FE	N	Y	Y	Y	Y
County FE	N	N	Y	Y	Y
Case FE	N	N	Y	Y	Y
YearXCase FE	N	N	N	Y	N
PatentXCountyXCase FE	N	N	N	N	Y
Observations	2,505,649	2,505,649	2,505,649	2,505,649	2,505,649

Note: Linear regression model where the dependent variable is the number of citations received by patent  $p$  in county  $f$  and year  $t$ . Number of citations in the losing county is 0.000561 in the period after entry. Standard errors are clustered at county level. Level of significance: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table B2: Number of citations and announcement: Poisson regressions

VARIABLES	(1)	(2)	(3)	(4)	(5)
Winner	0.12393 (0.48343)	0.19279 (0.32628)	-4.99295** (2.11066)	-0.76943 (0.87415)	-5.51540** (2.67200)
Post Announcement	0.41991* (0.25422)	0.38338 (0.49148)	-0.23389 (0.25085)	0.59897 (0.47584)	-0.20987 (0.25184)
WinnerXPost Announcement	0.57199* (0.31475)	0.28121 (0.26350)	0.35967** (0.14054)	0.41355** (0.17468)	0.32233** (0.15909)
log(ActivePopulation)		1.09417*** (0.18343)	2.74892** (1.09391)	0.43063 (0.93048)	2.52122** (1.20151)
Other Citations		0.10843*** (0.01026)	0.09966*** (0.00739)	0.09854*** (0.00891)	0.08389*** (0.02774)
Year FE	N	Y	Y	Y	Y
Year since Patent	N	Y	Y	Y	Y
County FE	N	N	Y	Y	Y
Case FE	N	N	Y	Y	Y
Tech FE	N	N	Y	Y	Y
YearXCase FE	N	N	N	Y	N
PatentXFipsXCase FE	N	N	N	N	Y
Observations	2,505,649	2,491,939	1,856,024	1,310,870	14,407

Table B3: Number of citations and opening: Poisson regressions

VARIABLES	(1)	(2)	(3)	(4)	(5)
Winner	0.25538 (0.39028)	0.20406 (0.35269)	-4.87004** (1.91905)	-1.00044 (0.75140)	-5.16926** (2.53790)
Post Opening	0.38740*** (0.12541)	0.18928 (0.45600)	-0.20919 (0.18728)	0.43304 (0.40696)	-0.16473 (0.17222)
WinnerXPost Opening	0.50158* (0.25703)	0.32796 (0.29038)	0.54452*** (0.20411)	0.73812*** (0.13377)	0.54673** (0.22530)
log(ActivePopulation)		1.08223*** (0.18062)	2.62895*** (1.00394)	0.31741 (0.82800)	2.35357** (1.12471)
Other Citations		0.10858*** (0.01020)	0.09946*** (0.00755)	0.09854*** (0.00891)	0.08473*** (0.02754)
Year FE	N	Y	Y	Y	Y
Year since Patent	N	Y	Y	Y	Y
County FE	N	N	Y	Y	Y
Case FE	N	N	Y	Y	Y
Tech FE	N	N	Y	Y	Y
YearXCase FE	N	N	N	Y	N
PatentXFipsXCase FE	N	N	N	N	Y
Observations	2,505,649	2,491,939	1,856,024	1,310,870	14,407

**CENTRE FOR ECONOMIC PERFORMANCE**  
**Recent Discussion Papers**

1446 Grace Lordan  
Jörn-Steffen Pischke Does Rosie Like Riveting? Male and Female Occupational Choices

1445 Stephen J. Redding  
David E. Weinstein A Unified Approach to Estimating Demand and Welfare

1444 Anna Valero  
John Van Reenen The Economic Impact of Universities: Evidence from Across the Globe

1443 Marta De Philippis STEM Graduates and Secondary School Curriculum: Does Early Exposure to Science Matter?

1442 Thierry Mayer  
Marc J. Melitz  
Gianmarco I.P. Ottaviano Product Mix and Firm Productivity Responses to Trade Competition

1441 Paul Dolan  
Georgios Kavetsos  
Christian Krekel  
Dimitris Mavridis  
Robert Metcalfe  
Claudia Senik  
Stefan Szymanski  
Nicolas R. Ziebarth The Host with the Most? The Effects of the Olympic Games on Happiness

1440 Jörn-Steffen Pischke Wage Flexibility and Employment Fluctuations: Evidence from the Housing Sector

1439 Brian Bell  
John Van Reenen CEO Pay and the Rise of Relative Performance Contracts: A Question of Governance

1438 Fadi Hassan  
Paolo Lucchino Powering Education

1437 Evangelia Leda Pateli Local and Sectoral Import Spillovers in Sweden

1436	Laura Kudrna Georgios Kavetsos Chloe Foy Paul Dolan	Without My Medal on My Mind: Counterfactual Thinking and Other Determinants of Athlete Emotions
1435	Andrew B. Bernard Andreas Moxnes Yukiko U. Saito	Production Networks, Geography and Firm Performance
1434	Zack Cooper Stephen Gibbons Matthew Skellern	Does Competition from Private Surgical Centres Improve Public Hospitals' Performance? Evidence from the English National Health Service
1433	Nicholas Bloom Raffaella Sadun John Van Reenen	Management as a Technology?
1432	Andrew B. Bernard Toshihiro Okubo	Product Switching and the Business Cycle
1431	Fabrice Defever Alejandro Riaño	Protectionism through Exporting: Subsidies with Export Share Requirements in China
1430	Andrew B. Bernard Renzo Massari Jose-Daniel Reyes Daria Taglioni	Exporter Dynamics and Partial-Year Effects
1429	Cletus C. Coughlin Dennis Novy	Estimating Border Effects: The Impact of Spatial Aggregation
1428	Alan Manning	The Elusive Employment Effect of the Minimum Wage
1427	Decio Coviello Andrea Guglielmo Giancarlo Spagnolo	The Effect of Discretion on Procurement Performance

**The Centre for Economic Performance Publications Unit**  
**Tel 020 7955 7673 Fax 020 7404 0612**  
**Email [info@cep.lse.ac.uk](mailto:info@cep.lse.ac.uk) Web site <http://cep.lse.ac.uk>**