



## WP 36\_08

**MICHAEL GREENSTONE**

Massachusetts Institute of Technology, UK

**RICHARD HORNBECK**

Massachusetts Institute of Technology, UK

**ENRICO MORETTI**

University of California, Berkeley, USA  
and The Rimini Centre of Economics Analysis, Italy

# “IDENTIFYING AGGLOMERATION SPILLOVERS: EVIDENCE FROM MILLION DOLLAR PLANTS”

Copyright belongs to the author. Small sections of the text, not exceeding three paragraphs, can be used provided proper acknowledgement is given.

*The Rimini Centre for Economic Analysis (RCEA)* was established in March 2007. RCEA is a private, non-profit organization dedicated to independent research in Applied and Theoretical Economics and related fields. RCEA organizes seminars and workshops, sponsors a general interest journal *The Review of Economic Analysis*, and organizes a biennial conference: *Small Open Economies in the Globalized World (SOEGW)*. Scientific work contributed by the RCEA Scholars is published in the RCEA Working Papers series.

The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Rimini Centre for Economic Analysis.

Identifying Agglomeration Spillovers: Evidence from Million Dollar Plants  
Michael Greenstone, Richard Hornbeck, and Enrico Moretti  
NBER Working Paper No. 13833  
March 2008  
JEL No. J0,R0

**ABSTRACT**

We quantify agglomeration spillovers by estimating the impact of the opening of a large new manufacturing plant on the total factor productivity (TFP) of incumbent plants in the same county. Articles in the corporate real estate journal *Site Selection* reveal the county where the "Million Dollar Plant" ultimately chose to locate (the "winning county"), as well as the one or two runner-up counties (the "losing counties"). The incumbent plants in the losing counties are used as a counterfactual for the TFP of incumbent plants in winning counties in the absence of the plant opening. Incumbent plants in winning and losing counties have economically and statistically similar trends in TFP in the 7 years before the opening, which supports the validity of the identifying assumption.

After the new plant opening, incumbent plants in winning counties experience a sharp relative increase in TFP. Five years after the opening, TFP of incumbent plants in winning counties is 12% higher than TFP of incumbent plants in losing counties. Consistent with some theories of agglomeration, this effect is larger for incumbent plants that share similar labor and technology pools with the new plant. We also find evidence of a relative increase in skill-adjusted labor costs in winning counties, indicating that the ultimate effect on profits is smaller than the direct increase in productivity.

Michael Greenstone  
Department of Economics  
Massachusetts Institute of Technology  
50 Memorial Drive E52-359  
Cambridge, MA 02142-1347  
and NBER  
mgreenst@mit.edu

Enrico Moretti  
Department of Economics  
University of California, Berkeley  
549 Evans Hall  
Berkeley, CA 94720-3880  
and NBER  
moretti@econ.berkeley.edu

Richard Hornbeck  
Department of Economics  
Massachusetts Institute of Technology  
50 Memorial Drive  
Cambridge, MA 02142-1347  
hornbeck@mit.edu

## Introduction

In most countries, economic activity is spatially concentrated. While some of this concentration is explained by the presence of natural advantages that constrain specific productions to specific locations, Ellison and Glaeser (1999) and others argue that natural advantages alone cannot account for the observed degree of agglomeration. Spatial concentration is particularly remarkable for industries that produce nationally traded goods, because the areas where economic activity is concentrated are typically characterized by high costs of labor and land. Since at least Marshall (1890), economists have speculated that this concentration of economic activity may be explained by cost or productivity advantages enjoyed by firms when they locate near other firms. The list of potential sources of these agglomeration advantages includes: cheaper and faster supply of intermediate goods and services; proximity to workers or consumers; better quality of the worker-firm match in thicker labor markets; lower risk of unemployment for workers and lower risk of unfilled vacancies for firms following idiosyncratic shocks; and knowledge spillovers.<sup>1</sup>

The possibility of documenting that productivity advantages through agglomeration are real is tantalizing, because it could provide insights into a series of important questions. Why are firms that produce nationally traded goods willing to locate in cities like New York, San Francisco or London, characterized by extraordinary production costs? In general, why do cities exist and what explains their historical development? Why do income differences persist across regions and countries?

Beside its obvious interest for urban and growth economists, the existence of agglomeration spillovers has tremendous practical relevance. Increasingly, local governments compete by offering substantial subsidies to industrial plants to locate within their jurisdictions. The main economic rationale for these incentives depends on whether the attraction of new businesses generates some form of agglomeration externalities. In the absence of positive externalities, it is difficult to justify the use of taxpayer money for subsidies based on efficiency grounds. The optimal magnitude of these incentives depends on the magnitude of agglomerations spillovers, if they exist.<sup>2</sup>

Despite their enormous theoretical and practical relevance, the existence and exact magnitude of agglomeration spillovers are considered open questions by many. To date, there are two primary approaches for testing for spillovers. The first tests for an unequal distribution of firms across the country. These “dartboard” style tests reveal that firms are spread unevenly across the country and that

---

<sup>1</sup> The literature on this topic is enormous, and can not be fully summarized here. Examples include, but are not limited to, Lucas (1988), Krugman (1991a, 1991b), Henderson (2001a, 2001b, 2003), Davis and Henderson (2004), Davis and Weinstein (2002), Henderson and Black (1999), Rosenthal and Stange (2001, 2004), Duranton and Puga (2004), Audretsch and Feldman, (1996, 2004), Moretti (2004a, 2004b, 2004c), Dumais, Ellison and Glaeser (2002), Glaeser (1999), Ottaviano and Thisse (2004).

<sup>2</sup> We discuss in more detail the policy implications of local subsidies in Greenstone and Moretti (2004). See also Card, Hallock, and Moretti (2007) and Glaeser (2001).

coagglomeration rates are higher between industries that are economically similar (Ellison, Glaeser and Kerr, 2007). This approach is based on equilibrium location decisions and does not provide a direct measure of spillovers.

The second approach uses micro data to assess whether firms' total factor productivity (TFP) is higher when similar firms are located nearby. A notable example is Henderson (2003), which estimates plant level production functions for machinery and high-tech industries as a function of the scale of other plants in the same and different industries.<sup>3</sup> The challenge for both approaches is that firms base their location decisions on where their profits will be highest, and this could be due to spillovers, natural advantage, or other cost shifters. A causal estimate of the magnitude of spillovers requires a solution to this problem of identification.

This paper tests for and quantifies agglomeration spillovers by estimating how the productivity of incumbent manufacturing plants changes when a new, large plant opens in their county. We estimate augmented Cobb-Douglas production functions that allow TFP to depend on the presence of the new plant, using plant-level data from the Annual Survey of Manufacturers. Because the location decision is made to maximize profits, the chosen county is likely to differ substantially from an average or randomly chosen county both at the time of opening and in future periods. Valid estimates of the plant opening's spillover effect require the identification of a county that is identical to the county where the plant decided to locate in the determinants of incumbent plants' TFP. These determinants are likely to include factors that affect the new plant's TFP.

This paper's solution is to rely on the reported location rankings of profit-maximizing firms to identify a valid counterfactual for what would have happened to incumbent plants' TFP in winning counties in the absence of the plant opening. These rankings come from the corporate real estate journal *Site Selection*, which includes a regular feature titled "Million Dollar Plants" that describes how a large plant decided where to locate. When firms are considering where to open a large plant, they typically begin by considering dozens of possible locations. They subsequently narrow the list to roughly 10 sites, among which 2 or 3 finalists are selected. The "Million Dollar Plants" articles report the county that the plant ultimately chose (i.e., the 'winner'), as well as the one or two runner-up counties (i.e., the 'losers'). The losers are counties that have survived a long selection process, but narrowly lost the competition.

The identifying assumption is that the incumbent plants in the losing county form a valid counterfactual for the incumbents in the winning counties, after conditioning on differences in pre-existing trends, plant fixed effects, industry-by-year fixed effects, and other control variables. Compared to the rest of the country, winning counties have higher rates of growth in income, population, and labor

---

<sup>3</sup> Moretti (2004b) takes a similar approach to estimate agglomeration externalities generated by human capital spillovers.

force participation. But compared to losing counties in the years before the opening of the new plant, winning counties have similar trends in most economic variables. This finding is consistent with both our presumption that the average county is not a credible counterfactual and our identifying assumption that the losers form a valid counterfactual for the winners.

We first measure the effect of the new Million Dollar Plant (MDP) on total factor productivity (TFP) of all incumbent manufacturing plants in winning counties. In the 7 years before the MDP opened, we find statistically equivalent trends in TFP for incumbent plants in winning and losing counties, which supports the validity of the identifying assumption. After the MDP opened, incumbent plants in winning counties experienced a sharp relative increase in TFP. Five years later, the MDP opening is associated with a 12% relative increase in incumbent plants' TFP.<sup>4</sup> This effect is statistically significant and economically substantial; on average incumbent plants' output in winning counties is \$430 million higher 5 years later (relative to incumbents in losing counties), holding constant inputs. We interpret this finding as evidence of meaningful productivity spillovers generated by increased agglomeration.

Having found evidence in favor of the existence of agglomeration spillovers, we then try to shed some light on the possible mechanisms. We follow Moretti (2004b) and Ellison, Glaeser, and Kerr (2007) and investigate whether the magnitude of the spillovers depends on economic linkages between the incumbent plant and the MDP. Specifically, we test whether incumbents that are geographically and economically linked to the MDP experience larger spillovers, relative to incumbents that are geographically close but economically distant from the MDP. We use several measures of economic links including input and output flows, measures of the degree of sharing of labor pools, and measures of technological linkages.<sup>5</sup>

We find that spillovers are larger for incumbent plants in industries that share worker flows with the MDP industry. A one standard deviation increase in our measure of worker transitions is associated with a 7 percentage point increase in the magnitude of the spillover. Similarly, the measures of technological linkages indicate statistically meaningful increases in the spillover effect. Surprisingly, we find little support for the importance of input and output flows in determining the magnitude of the spillover. Overall, this evidence provides some support for the notion that spillovers occur between firms that share workers and use similar technologies.

To guide the analysis and interpret the results, we set out a straightforward Roback (1982) style model that incorporates spillovers between producers and derives an equilibrium allocation of firms and workers across locations. In the model, the entry of a new firm produces spillovers. This leads new firms

---

<sup>4</sup> Notably, naïve estimates that control for observables but do not use the MDP research design find negative productivity effects.

<sup>5</sup> We are deeply indebted to Glenn Ellison, Edward Glaeser, and William Kerr for providing their data for five of these measures of economic distance.

who are interested in gaining access to the spillover to enter. This process of entry leads to competition for inputs so that incumbent firms face higher prices for labor, land, and other local inputs. In the model, firms produce nationally traded goods, so they cannot raise the price of their output in response to the higher input prices. Thus, the long run equilibrium is obtained when the value of the increase in output due to spillovers is equal to the increased costs of production due to the higher input prices

The data support two predictions derived from this model. First, we find positive net entry in winning counties, which the model predicts will occur if there are sufficiently large positive spillovers. Second, we find increases in quality-adjusted wages following MDP openings. These higher wages are consistent with the documented increase in economic activity in the winning counties (which presumably is a response to the spillovers). Furthermore, the higher wages support the model's prediction that the productivity gains from agglomeration do not necessarily translate into higher profits for incumbents in the long run.

The remainder of the paper is organized as follows. Section I presents a simple model. Section II discusses the identification strategy. Section III presents the data sources. Section IV presents the econometric model. Sections V and VI presents the empirical findings and interpret them. Section VII concludes.

## **I. Theories of Agglomeration and Theoretical Framework**

We are interested in identifying how the opening of a new plant in a county affects the productivity, profits, and input use of existing plants in the same county. We begin by reviewing the theories of agglomeration. We then present a simple theoretical framework that guides the subsequent empirical exercise and aids in interpreting the results.

### **A. Theories of Agglomeration**

Economic activity is geographically concentrated (Ellison and Glaeser, 1997). What are the forces that can explain such agglomeration of economic activity? Here we summarize five possible reasons for agglomeration, and briefly discuss what each of them implies for the relationship between productivity and the density of economic activity.

(1) First, it is possible that firms (and workers) are attracted to areas with a high concentration of other firms (and other workers) by the size of the labor market. There are at least two different reasons why larger labor markets may be attractive. First, a thick labor market is beneficial in the presence of search frictions, if jobs and workers are heterogeneous. In the presence of frictions, a worker-firm match will be on average more productive in areas where there are many firms offering jobs and many workers

looking for jobs.<sup>6</sup>

Alternatively, it is possible that large labor markets are more desirable because they provide insurance against idiosyncratic shocks, either on the firm side or on the worker side (Krugman 1991a). If moving is costly for workers and firms are subject to idiosyncratic and unpredictable demand shocks that lead to lay-offs, workers will prefer to be in areas with thick labor markets to reduce the probability of being unemployed. Similarly, if finding new workers is costly, firms will prefer to be in areas with thick labor markets to reduce the probability of having unfilled vacancies.<sup>7</sup>

These two hypotheses have different implications for the relationship between concentration of economic activity and productivity. If the size of the labor market leads only to better worker-firm matches, we should see that firms located in denser areas are more productive than otherwise identical firms located in less dense areas. The exact form of this productivity gain depends on the shape of the production function.<sup>8</sup>

On the other hand, if the only effect of thickness in labor market is a lower risk of unemployment for workers and a lower risk of unfilled vacancies for firms, there should not be differences in productivity between dense and less dense areas. While productivity would not vary, wages would vary across areas depending on the thickness of the labor market, although the exact effect of density on wages is *a priori* ambiguous.<sup>9</sup> This change in relative factor prices will change the relative amount of labor and capital used. Unlike the case of improved matching described above, the production function does not change: for the same set of labor and capital inputs, the output of firms in denser areas should be similar to the output of firms in less dense areas.

(2) A second reason why the concentration of economic activity may be beneficial has to do with transportation costs (Krugman 1991a and 1991b, Glaeser and Kohlhase, 2003). Because in this paper we focus on firms that produce nationally traded goods, transportation costs of finished products are unlikely to be the relevant cost in this paper's setting. Only a small fraction of buyers of the final product is likely

---

<sup>6</sup> For a related point in a different context, see Petrongolo and Pissarides (2005).

<sup>7</sup> A third alternative hypothesis has to do with spillovers that arise because of endogenous capital accumulation. For example, in Acemoglu (1996), plants have more capital and better technology in areas where the number of skilled workers is larger. If firms and workers find each other via random matching and breaking the match is costly, externalities will arise naturally even without learning or technological externalities. The intuition is simple. The privately optimal amount of skills depends on the amount of physical capital a worker expects to use. The privately optimal amount of physical capital depends on the number of skilled workers. If the number of skilled workers in a city increases, firms in that city, expecting to employ these workers, will invest more. Because search is costly, some of the workers end up working with more physical capital and earn more than similar workers in other cities.

<sup>8</sup> For example, it is possible that the productivities of both capital and labor benefit from the improved match in denser areas. It is also possible that the improved match caused by a larger labor market benefits only labor productivity. This has different implications for the relative use of labor and capital, but total factor productivity will be higher regardless.

<sup>9</sup> Its sign depends on the relative magnitude of the compensating differential that workers are willing to pay for lower risk of unemployment (generated by an increase in labor supply in denser areas) and the cost savings that firms experience due to lower risk of unfilled vacancies (generated by an increase in labor demand in denser areas).

to be located in the same area as our manufacturing plants. The relevant costs are the transportation costs of suppliers of local services and local intermediate goods. Firms located in denser areas are likely to enjoy cheaper and faster delivery of local services and local intermediate goods. For example, a high-tech firm that needs a specialized technician to fix a machine is likely to get service more quickly and at lower cost if it is located in Silicon Valley than in the Nevada desert.

This type of agglomeration spillover does not imply that the production function varies as a function of density of economic activity: for the same set of labor and capital inputs, the output of firms in denser areas should be similar to the output of firms in less dense areas. However, production costs should be lower in denser areas.

(3) A third reason why the concentration of economic activity may be beneficial has to do with knowledge spillovers. There are at least two different versions of this hypothesis. First, economists and urban planners have long speculated that the sharing of knowledge and skills through formal and informal interaction may generate positive production externalities across workers.<sup>10</sup> Empirical evidence indicates that this type of spillover may be important in some high-tech industries. For example, patent citations are more likely to come from the same state or metropolitan area as the originating patent (Jaffe et al. 1993). Saxenian (1994) argues that geographical proximity of high-tech firms in Silicon Valley is associated with a more efficient flow of new ideas and ultimately causes faster innovation.<sup>11</sup> Second, it is also possible that proximity results in sharing of information on new technologies and therefore leads to faster technology adoption. This type of social learning phenomenon applied to technology adoption was first proposed by Griliches (1958).

If density of economic activity results in intellectual externalities, the implication of this type of agglomeration model is higher productivity. In particular, we should see that firms located in denser areas are more productive than otherwise identical firms located in less dense areas. As with the search model, this higher productivity could benefit both labor and capital, or only one of the two factors, depending on the form of the production function. On the other hand, if density of economic activity only results in faster technology adoption and the price of new technologies reflects their higher productivity, there should be no relationship between productivity and density, after properly controlling for quality of capital.

(4) It is possible that firms concentrate spatially not because of any technological spillover, but because local amenities valued by workers are concentrated. For example, skilled workers may prefer

---

<sup>10</sup> See, for example, Marshall (1890), Lucas (1988), Jovanovic and Rob (1989), Grossman and Helpman (1991), Saxenian (1994), Glaeser (1999), and Moretti (2004a, 2004b and 2004c).

<sup>11</sup> The entry decisions of new biotechnology firms in a city depend on the stock of outstanding scientists there, as measured by the number of relevant academic publications (Zucker et al., 1998). Moretti (2004b) finds stronger human capital spillovers between pairs of firms in the same city that are economically or technologically closer.



certain amenities more than unskilled workers. This would lead firms that employ relatively more skilled workers to concentrate in locations where these amenities are available. In this case, we should not see any difference in productivity between dense areas and less dense areas, although we should see differences in wages that reflect the compensating differential.

(5) Finally, spatial concentration of some industries may be explained by the presence of natural advantages. For example, the oil industry is concentrated in a limited number of states because those states have the most accessible oil fields. Similarly, the wine industry is concentrated in California because that is where good weather and suitable land are to be found. For some manufacturing productions, the presence of a harbor may be important. The presence of natural advantages has the implication that firms located in areas with high concentration of similar firms are more productive, but of course this correlation has nothing to do with agglomeration spillovers. Since most natural advantages are fixed over time, this explanation is not particularly relevant for our empirical estimates, which exploit variation over time in agglomeration.

## B. A Simple Model

We begin by considering the case where incumbent firms are homogenous in size and technology. Later we consider what happens when incumbent firms are heterogeneous. Throughout the paper, we focus on the case of factor-neutral spillovers.

**(a) Homogeneous Incumbents.** We assume that all incumbent firms use a production technology that uses labor, capital, and land to produce a nationally traded good whose price is fixed and is normalized to 1. Incumbent firms choose their amount of labor,  $L$ , capital,  $K$ , and land,  $T$ , to maximize the following expression:

$$\text{Max}_{L,K,T} \{ f[A, L, K, T] - wL - rK - qT \}$$

where  $w$ ,  $r$  and  $q$  are input prices and  $A$  is a productivity shifter (TFP). Specifically,  $A$  includes all factors that affect the productivity of labor, capital, and land equally, such as technology and agglomeration spillovers, if they exist. In particular, to explicitly allow for agglomeration effects, we allow  $A$  to depend on the density of economic activity in an area:

$$(1) \quad A = A(N)$$

where  $N$  is the number of firms that are active in a county, and all counties have equal size. We define factor-neutral agglomeration spillovers as the case where  $A$  increases in  $N$ :

$$\delta A / \delta N > 0$$

If instead  $(\delta A / \delta N) = 0$ , we say that there are no factor-neutral agglomeration spillovers.

Let  $L^*(w,r,q)$  be the optimal level of labor inputs, given the prevailing wage, cost of capital, and cost of industrial land. Similarly, let  $K^*(w,r,q)$  and  $T^*(w,r,q)$  be the optimal level of capital and land,

respectively. In equilibrium,  $L^*$ ,  $K^*$ , and  $T^*$  are set so that the marginal product of each of the three factors is equal to its price.

We assume that capital is internationally traded, so its price does not depend on local demand or supply conditions. However, we allow for the price of labor and land to depend on local economic conditions. In particular, we allow the supply of labor and land to be less than infinitely elastic at the county level.

We attribute the upward sloping labor supply curve to the existence of moving costs. Like in the standard Roback (1982) model, we assume that workers' indirect utility depends on wages and cost of housing, and that in equilibrium workers are indifferent across locations. Workers are mobile across locations, but unlike the standard Roback (1982) model we allow for moving costs. For simplicity, we ignore labor supply decisions within a given location and assume that all residents provide a fixed amount of labor.

To illustrate this, consider that there are  $m$  workers in county  $c$  before the opening of the new plant. In particular,  $m$  is such that, given the distribution of wages and the housing costs across localities, the marginal worker in another county is indifferent between moving to county  $c$  and staying. When a new plant opens in county  $c$ , wages there start rising, and some workers find it optimal to move to  $c$ . The number of workers who move, and therefore the slope of the labor supply function, depend on the shape of the mobility cost function. Let  $w(N)$  be the inverse of the reduced-form labor supply function that links the number of firms,  $N$ , active in a county to the local nominal wage level,  $w$ .

Similarly, we allow the supply of industrial land to be less than infinitely elastic at the county level. For example, it is possible that the supply of land is fixed because of geography or land-use regulations. Alternatively, it may not be completely fixed, but it is possible that the best industrial land has already been developed, so that the marginal land is of decreasing quality or more expensive to develop. Irrespective of the reason, we call  $q(N)$  the (inverse of the) reduced form land supply function that links the number of firms,  $N$ , to the price of land,  $q$ . We can therefore write the equilibrium level of profits,  $\Pi^*$ , as

$$\Pi^* = f[ A(N), L^*(w(N), r, q(N)), K^*(w(N), r, q(N)), T^*(w(N), r, q(N)) ] - w(N) L^*(w(N), r, q(N)) - r K^*(w(N), r, q(N)) - q(N) T^*(w(N), r, q(N))$$

where we now make explicit the fact that TFP, wages, and land prices depend on the number of firms active in a county.

Consider the total derivative of incumbents' profits with respect to a change in the number of firms:

$$(2) \quad d\Pi^*/dN = (\delta f/\delta A \delta A/\delta N) + \delta w/\delta N \{ [\delta L^*/\delta w (\delta f/\delta L - w) - L^*] + [\delta K^*/\delta w (\delta f/\delta K - r)] + [\delta T^*/\delta w (\delta f/\delta T - q)] \}$$

$$+ \delta q / \delta N \{ [\delta L^* / \delta q (\delta f / \delta L - w)] + [\delta K^* / \delta q (\delta f / \delta K - r)] + [\delta T^* / \delta q (\delta f / \delta T - q) - T^*] \}$$

If all firms are price takers and all factors are paid their marginal product, equation (2) simplifies considerably and can be written as:

$$(3) \quad d\Pi^* / dN = (\delta f / \delta A \delta A / \delta N) - [\delta w / \delta N L^* + \delta q / \delta N T^*].$$

Equation (3) makes clear that the effect of an increase in  $N$  is the sum of two opposite effects. First, if there are positive spillovers, the productivity of all factors increases. In equation (3), this effect on TFP is represented by the first term,  $(\delta f / \delta A \delta A / \delta N)$ . This effect is unambiguously positive, because it allows an incumbent firm to produce more output using the same amount of inputs. Formally,  $\delta f / \delta A > 0$  by assumption and, if there are positive spillovers,  $\delta A / \delta N > 0$ .

The second term,  $- [\delta w / \delta N L^* + \delta q / \delta N T^*]$ , represents the negative effect from increases in the cost of production, specifically the prices of labor and land. Formally, this term is negative because we have assumed that  $\delta w / \delta N > 0$  and  $\delta q / \delta N > 0$ . (The magnitudes depend on the elasticity of the supply of labor and land.) Intuitively, an increase in  $N$  is an increase in the level of economic activity in the county and therefore an increase in the local demand for labor and land.

Unlike the beneficial effect of agglomeration spillovers, the increase in factor prices is costly for incumbent firms, because they now have to compete for locally scarce resources with the new entrant. The increase in wages and land prices has two effects on incumbents. First, for a given level of input utilizations, it mechanically raises production costs. Second, it leads the firm to re-optimize and to change its use of the different production inputs. In particular, given that the price of capital is not affected by an increase in  $N$ , the firm is likely to end up using more capital than before:

$$\delta K^* / \delta N \Rightarrow 0.$$

By contrast, the effect on the use of labor and land is ambiguous. On the one hand, the productivity of all factors increases. On the other hand, the price of labor and land might increase. The net effect depends on the magnitude of the factor price increases, as well as on the exact shape of the production function (i.e., the strength of technological complementarities between labor, capital, and land).

It is instructive to apply these derivations to the case of a MDP opening that causes positive spillovers. We initially consider the case where for incumbent firms  $d\Pi^* / dN \leq 0$ . This would occur when the agglomeration spillover is smaller than the increase in production costs. In this case, the MDP's opening would not lead to any entry and could cause some existing firms to exit.<sup>12</sup>

The alternative case is that  $d\Pi^* / dN > 0$ , which occurs when the magnitude of the spillover due to the MDP's opening exceeds the increase in factor prices due to the MDP's demand for local inputs. In

---

<sup>12</sup> Similarly, if the spillovers are zero or negative, one might expect exit of incumbent firms and a reduction in local economic activity.

the short run, profits will be positive for new entrants. These positive profits will disappear over time as the price of local factors, like land and possibly labor, is bid up.

In the long run, there is an equilibrium such that firms and workers are indifferent between the county where the new plant has opened and other locales. Since the amount of land is fixed, the higher levels of productivity are likely to be capitalized into land prices. It is also likely that wages will increase. This may occur due to moving costs as noted above.<sup>13</sup> These adjustments make workers indifferent between the county with the new plant and other counties. Similarly, the changes in factor prices mean that firms earn the same profits in the county with the new plant (even in the presence of the spillovers) and in other locations. From a practical perspective, it is impossible to know when the short run ends and the long run begins.

There are two empirical predictions that apply when there are positive spillovers. First, if the magnitude of the spillovers is large enough, new firms will enter the MDP's county to gain access to the spillover. This prediction of increased economic activity holds at any point after potential new entrants have had sufficient time to respond. The second prediction is that the prices of locally traded inputs will rise as the MDP and the new entrants bid for these inputs.<sup>14</sup>

**(b) Heterogeneous Incumbents.** What happens if the population of incumbent firms is non-homogeneous? Consider the case where there are two types of firms: high-tech and low-tech. Assume that for technological reasons, the type of workers employed by high-tech firms,  $L_H$ , differs to some extent from the type of workers employed by low-tech firms,  $L_L$ , although there is some overlap. Assume that the new entrant is a high-tech firm. Equations (4) and (5) characterize the effect of the new high-tech firm on high-tech and low-tech incumbents:

$$(4) \quad d\Pi_H^*/dN_H = (\delta f_H/\delta A_H \delta A_H/\delta N_H) - [\delta w_H/\delta N_H L_H^* + \delta q/\delta N_H T^*]$$

<sup>13</sup> Even with zero moving costs and an infinitely elastic supply of labor, wages will increase if there are land price increases as workers will demand higher wages as compensation for the higher land rents for their homes (Roback 1982).

<sup>14</sup> This paper focuses on the case where the productivity benefits of the agglomeration spillovers are distributed equally across all factors. What happens when agglomeration spillovers are factor biased? Assume, for example, that agglomeration spillovers raise the productivity of labor, but not the productivity of capital. Like before, the technology is  $f[A, L, K, T]$ , but now  $L$  represents units of effective labor. In particular,  $L = \theta H$ , where  $H$  is the number of physical workers and  $\theta$  is a productivity shifter. We define factor-biased agglomeration spillover as the case where the productivity shifter  $\theta$  depends positively on the density of the economic activity in the county  $\theta = \theta(N)$  and  $\delta\theta/\delta N > 0$ . If  $\delta A/\delta N = 0$  and factors are paid their marginal product, then the effect of an increase in the density of the economic activity in a county on incumbent firms simplifies to  $d\Pi^*/dN = (\delta f/\delta H \delta\theta/\delta N) H - [\delta w/\delta N H^* + \delta q/\delta N T^*]$ . The effect on profits can be decomposed in two parts. The first term represents the increased productivity of labor. It is the product of the sensitivity of output to labor ( $\delta f/\delta H > 0$ ), times the magnitude of the agglomeration spillover ( $\delta\theta/\delta N > 0$  by definition), times the number of workers. The second term is the same as in equation (3), and represents the increase in the costs of locally supplied inputs. The increase in  $N$  changes the optimal use of the production inputs. Labor is now more productive, and its equilibrium use increases:  $\delta L^*/\delta N > 0$ . Land is equally productive but its price increases. Its equilibrium use declines:  $\delta T^*/\delta N < 0$ . Neither the price nor the productivity of capital is affected by an increase in  $N$ . Its equilibrium use depends on technology. Specifically, it depends on the elasticity of substitution between labor and capital.

$$(5) \quad d\Pi_L^*/dN_H = (\delta f_L/\delta A_L \delta A_L/\delta N_H) - [\delta w_L/\delta N_H L^*_L + \delta q/\delta N_H T^*]$$

It is plausible to expect that the beneficial effect of agglomeration spillovers generated by a new high-tech entrant is larger for high-tech firms than for low-tech firms:

$$(5') \quad (\delta f_H/\delta A_H \delta A_H/\delta N_H) > (\delta f_L/\delta A_L \delta A_L/\delta N_H)$$

At the same time, one might expect that the increase in labor costs is also higher for the high-tech incumbents, given that they are now competing for workers with an additional high-tech firm:

$$\delta w_H/\delta N_H > \delta w_L/\delta N_H$$

The effect on land prices should be similar for both firm types, since the assumption of a single land market seems reasonable.

There are two takeaways here. First, it may be reasonable to expect larger spillovers to firms that are economically “close” to the new plant. Second the relative impact of the new plant on profits is unclear, because the economically “closer” plants are likely to have bigger spillovers and larger increases in production costs.

### C. Empirical Predictions

The simple theoretical framework above generates four predictions that we bring to the data. Specifically if there are positive spillovers, then:

1. the opening of a new plant will increase the TFP of incumbent plants.
2. the increase in TFP may be larger for firms that are economically “closer” to the new plant.
3. the density of economic activity in the county will increase as firms move in to gain access to the positive spillovers (if the spillovers are large enough).
4. the price of factors of production that are traded locally will increase. We test for changes in the price of quality-adjusted labor, which is arguably the most important locally supplied factor of production for manufacturing establishments.

## II. Plant Location Decisions and Research Design

In testing the four empirical predictions outlined above, the main econometric challenge is the fact that firms do not choose their location randomly. Firms are profit maximizers and choose to locate where their expectation of the present discounted value of the stream of future profits is greatest. This net present value varies tremendously across locations, depending on many factors, including transportation infrastructure, the availability of workers with particular skills, subsidies, etc. These factors are frequently unobserved. Further, they are likely to be correlated with the TFP of existing plants.

Therefore, a naïve comparison of the TFP of incumbents in counties that experience a plant opening with the TFP of incumbents in counties that do not experience a plant opening is likely to yield

biased estimates of productivity spillovers. Credible estimates of the impact of a plant opening on TFP of incumbent plants require the identification of a location that is similar to the location where the plant decided to locate in the determinants of incumbent plants' TFP.

This section provides a case study for how BMW picked the location for one of its plants.<sup>15</sup> The intent is to demonstrate the empirical difficulties that arise when estimating the effect of plant openings on the TFP of incumbent plants. Further, it illustrates informally how our research design may circumvent these difficulties.

After overseeing a worldwide competition and considering 250 potential sites for its new plant, BMW announced in 1991 that they had narrowed the list of potential candidates to 20 counties. Six months later, BMW announced that the two finalists in the competition were Greenville-Spartanburg, South Carolina, and Omaha, Nebraska. In 1992, BMW announced that they would site the plant in Greenville-Spartanburg and that they would receive a package of incentives worth approximately \$115 million funded by the state and local governments.

Why did BMW choose Greenville-Spartanburg? Two factors were important in this decision. The first was BMW's expected future costs of production in Greenville-Spartanburg, which are presumably a function of the county's expected supply of inputs and BMW's production technology. According to BMW, the characteristics that made Greenville-Spartanburg more attractive than the other 250 sites initially considered were: low union density, a supply of qualified workers, the numerous global firms, including 58 German companies, in the area; the high quality transportation infrastructure, including air, rail, highway, and port access; and access to key local services.

For our purposes, the important point to note here is that these county characteristics are a potential source of unobserved heterogeneity. While these characteristics are well documented in the BMW case, they are generally unknown and unobserved. If these characteristics also affect the growth of TFP of existing plants, a standard regression that compares Greenville-Spartanburg with the other 3,000 United States counties will yield biased estimates of the effect of the plant opening. A standard regression will overestimate the effect of plant openings on outcomes if, for example, counties that have more attractive characteristics (e.g., improving transportation infrastructure) tend to have faster TFP growth. Conversely, a standard regression would underestimate the effect if, for example, incumbent plants' declining TFP encouraged new entrants (e.g., cheaper availability of local inputs).

A second important factor in BMW's decision was the value of the subsidy it received. Presumably Greenville-Spartanburg was willing to provide BMW with \$115 million in subsidies because it expected economic benefits from BMW presence. According to local officials, the facility's ex-ante

---

<sup>15</sup> This plant is in Greenstone and Moretti's (2004) set of 82 MDP plants. Due to Census confidentiality restrictions, we cannot report whether this plant is part of this paper's analysis.

expected five-year economic impact on the region was \$2 billion. As a part of this \$2 billion, the plant was expected to create 2,000 jobs *directly* and another 2,000 jobs *indirectly*. In principle, these 2,000 additional jobs could reflect the entry of new plants or the expansion of existing plants caused by agglomeration economies. (The empirical section tests whether this is indeed the case on average). Thus, the subsidy is likely to be a function of the expected gains from agglomeration for the county.<sup>16</sup>

This possibility is relevant for this paper's identification strategy, because the magnitude of the spillover from a particular plant depends on the level and growth of a county's industrial structure, labor force, and a series of other unobserved variables. For this reason, the factors that determine the total size of the potential spillover (and presumably the size of the subsidy) represent a second potential source of unobserved heterogeneity. If this unobserved heterogeneity is correlated with incumbent plants' TFP, standard regression equations will be misspecified due to omitted variables, just as described above.

In order to make valid inferences in the presence of the heterogeneity associated with the plant's expected local production costs and the county's value of attracting the plant, knowledge of the exact form of the selection rule that determines plants' location decisions is generally necessary. As the BMW example demonstrates, the two factors that determine plant location decisions are generally unknown to researchers and, in the rare cases where they are known, are difficult to measure. Thus, the effect of a plant opening on incumbents' TFP is very likely to be confounded by differences in factors that determine the plants' profitability at the chosen location.

As a solution to this identification problem, we rely on the reported location rankings of profit-maximizing firms to identify a valid counterfactual for what would have happened to incumbent plants in winning counties in the absence of the plant opening. We implement the research design using data from the corporate real estate journal *Site Selection*. Each issue of this journal includes an article titled the "Million Dollar Plants" that describes how a large plant decided where to locate. These articles always report the county that the plant chose (i.e., the 'winner'), and usually report the runner-up county or counties (i.e., the "losers").<sup>17</sup> As the BMW case study indicates, the winner and losers are usually chosen from an initial sample of "semi-finalist" sites that in many cases number more than a hundred.<sup>18</sup> The losers are counties that have survived a long selection process, but narrowly lost the competition.

---

<sup>16</sup> The fact that business organizations such as the Chambers of Commerce support these incentive plans (as was the case with BMW) suggests that incumbent firms expect such increases. Greenstone and Moretti (2004) present a model that describes the factors that determine local governments' bids for these plants and whether successfully attracting a plant will be welfare increasing or decreasing for the county.

<sup>17</sup> In some instances the "Million Dollar Plants" articles do not identify the runner-up county. For these cases, we did a Lexis/Nexis search for other articles discussing the plant opening and in 4 cases, among the original 82, we were able to identify the losing counties. The Lexis/Nexis searches were also used to identify the plant's industry when this was unavailable in *Site Selection*. Comprehensive data on the subsidy offered by winning and losing counties is unavailable in the *Site Selection* articles.

<sup>18</sup> The names of the semi-finalists are rarely reported.

We use the losers to identify what would have happened to the productivity of incumbent plants in the winning county in the absence of the plant opening. Specifically, we assume that incumbent firms' TFP would have trended identically in the absence of the plant opening in winning and losing counties within a case. In practice, we adjust for covariates so our identifying assumption is weaker. The subsequent analysis provides evidence that supports the validity of this assumption. Even if this assumption fail to hold, we presume that this pairwise approach is more reliable than using regression adjustment to compare the TFP of incumbent plants in counties with new plants to the other 3,000 United States counties or to using a matching procedure based on observable variables.<sup>19</sup>

### III. Data Sources and Summary Statistics

#### A. Data Sources

The “Million Dollar Plants” articles typically reveal the county where the new firm (the “Million Dollar Plant”) ultimately chose to locate (the “winning county”), as well as the one or two runner-up counties (the “losing counties”). The articles tend to focus on large plants that are the target of local government subsidies. Important limitations of these articles are that the magnitude of the subsidy offered by the winning counties is in many cases unobserved and that the bid is almost always unobserved for losing counties.

We identify the Million Dollar Plants in the Standard Statistical Establishment List (SSEL) – which is the Census Bureau’s “most complete, current, and consistent data for U.S. business establishment”<sup>20</sup> – and matched the plants to the Annual Survey of Manufactures (ASM) and the Census of Manufactures (CM) from 1973-1998.<sup>21</sup> Of the 82 MDP openings in Greenstone and Moretti (2004), we identified 47 genuine and useable MDP openings in the manufacturing data. In order to qualify as a genuine and useable MDP manufacturing opening, we imposed the following criterion: 1) there had to be a new plant in the manufacturing sector, owned by the reported firm, appearing in the SSEL within 2 years before and 3 years after the publication of the MDP article; 2) the plant identified in the SSEL had to be located in the county indicated in the MDP article; and 3) there had to be incumbent plants in both winning and losing counties that were there for each of the previous 8 years. Among the 35 MDP

---

<sup>19</sup> Propensity score matching is an alternative approach (Rosenbaum and Rubin 1983). Its principal shortcoming relative to our approach is its assumption that the treatment (i.e., winner status) is “ignorable” conditional on the observables. As it should be clear from the example, adjustment for observable variables through the propensity score is unlikely to be sufficient.

<sup>20</sup> The SSEL is confidential and was accessed in a Census Data Research Center. The SSEL is updated continuously and incorporates data from all Census Bureau economic and agriculture censuses and current business surveys, quarterly and annual Federal income and payroll tax records, and other Departmental and Federal statistics and administrative records programs.

<sup>21</sup> The sample is cut at 1998 because sampling methods in the ASM changed for 1999. The sample begins in 1973 because of minor known inconsistencies with the 1972 CM.



openings that did not qualify, roughly 20 were outside of the manufacturing sector. (We cannot report the exact number because of the Census Bureau's confidentiality rules).

To obtain information on incumbent establishments in winner and loser counties, we use the *ASM* and *CM*. The *ASM* and *CM* contain information on employment, capital stock, total value of shipments, plant age, and firm identifiers. The 4-digit SIC code and county of location are also reported and these play a key role in the analysis. Importantly, the manufacturing data contain a unique plant identifier, making it possible to follow individual plants over time.<sup>22</sup> The sample that we use includes plants that were continuously present in the *ASM* in the 8 years preceding the year of the plant opening plus the year of the opening. Additionally, we drop observations on plants that have the same owner as the MDP plants. In this period, the *ASM* sampling scheme was positively related to firm and plant size. Any establishment that was part of a company with manufacturing shipments exceeding \$500 million was sampled with certainty, as were establishments with 250 or more employees.

There are a few noteworthy features of this sample of potentially affected plants. First, the focus on existing plants allows for a test of spillovers on a fixed sample of pre-existing plants, which eliminates concerns related to the endogenous opening of new plants and compositional bias. Second, it is possible to form a genuine panel of manufacturing plants. Third, a disadvantage is that the results may not be externally valid to smaller incumbent plants that are not sampled with certainty throughout this period. Nevertheless, it is relevant that this sample of plants accounts for 54% of county-wide manufacturing shipments in the last *CM* before the MDP opening.

Besides testing for an average spillover effect, we also test whether the estimated agglomeration effects are larger in industries that are more closely linked to the MDP based on some measure of economic distance. We use six measures of economic distance in three categories. First, to measure supplier and customer linkages, we use data on the fraction of each industry's manufactured inputs that come from each 3-digit industry and the fraction of each industry's outputs sold to manufacturers that are purchased by each 3-digit industry. Second, to measure the frequency of worker mobility between industries, we use data on labor market transitions from the *Current Population Survey (CPS)* outgoing rotation file. In particular, we measure the fraction of separating workers from each 2-digit industry that move to firms in each 2-digit industry. Third, to measure technological proximity, we use data on the fraction of patents manufactured in a 3-digit industry that cite patents manufactured in each 3-digit industry. We also use data on the amount of R&D expenditure in a 3-digit industry that is used in other 3-digit industries.

Finally, one further data issue merits attention. We have two sources of information on the date of the plant opening. The first is the MDP articles, which often are written when ground is broken on the

---

<sup>22</sup> See the appendix in Davis, Haltiwanger, and Schuh (1996) for a more thorough description of the *ASM* and *CM*.

plant but other times are written when the location decision is made or the plant begins operations. The second source is the SSEL, which in principle reports the plant's first year of operation. However, it is known that plants occasionally enter the SSEL after their opening.

Thus, there is uncertainty about the date of the plant's opening. Further, the date at which the plant could affect the operations of existing plants depends on the channel for any agglomeration economies. If the agglomeration economies are a consequence of supplier relationships, then they could occur as soon as the plant is announced. For example, the new plant's management might visit existing plants and provide suggestions on operations. Alternatively, the agglomeration spillovers may be driven by the labor market and therefore may depend on sharing labor. In this case, agglomeration economies may not be evident until the plant is operating. Based on these data and conceptual issues, there is not clear guidance on when the new plant could affect other plants. Rather than take an unsupportable stand, we emphasize results using the earlier of the year of the publication of the magazine article and the year that the new plant appears in the SSEL. We also report separate results based on using these two years as the opening date. The basic findings are robust to these alternatives.

## **B. Summary Statistics**

Table 1 presents summary statistics on the sample of plant location decisions that forms the basis of the analysis. As discussed in the previous subsection, there are 47 manufacturing MDP openings that we can match to plant level data. There are plants in the same 2-digit SIC industry in both winning and losing counties in the 8 years preceding the opening for just 16 of these openings.

The table reveals some other facts about the plant openings.<sup>23</sup> We refer to the winner and accompanying loser(s) associated with each plant opening announcement as a "case." There are two or more losers in 16 of the cases, so there are a total of 73 losing counties along with 47 winning counties. Some counties appear multiple times in the sample (as either a winner or loser), and the average county in the sample appears a total of 1.09 times. The difference between the year of the MDP article's publication and the year the plant appears in the SSEL is roughly spread evenly across the categories -2 to -1 years, 0 years, and 1 to 3 years. For clarity, positive differences refer to cases where the article appears after the plant is identified in the SSEL. The date of the plant openings ranges from the early 1980s through the early 1990s.

The remainder of Table 1 provides summary statistics on the new MDP plants five years after their assigned opening date to provide a sense of their magnitude. These MDPs are quite large: they are more than twice the size of the average incumbent plant and account for roughly nine percent of the average county's total output one year prior to its opening.

---

<sup>23</sup> A number of the statistics in Table 1 are reported in broad categories to comply with the Census Bureau's confidentiality restrictions and to avoid disclosing the identities of any individual plants.

Table 2 provides summary statistics on the measures of industry linkages and further descriptions of these variables. In all cases, the proximity between industries is increasing in the value of the variable. For ease of interpretation in the subsequent regressions, these variables are normalized to have a mean of zero and a standard deviation of one.

Table 3 presents the means of county-level and plant-level variables across counties. These means are reported for winners, losers, and the entire United States in columns (1), (2), and (3), respectively.<sup>24</sup> In the winner and loser columns, the plant-level variables are calculated among the incumbent plants present in the *ASM* in the 8 years preceding the assigned opening date and the assigned opening date. All entries in the entire United States column are weighted across years to produce statistics for the year of the average MDP opening in our sample. Further, the plant characteristics are only calculated among plants that appear in the *ASM* for at least 9 consecutive years. Column (4) presents the t-statistics from a test that the entries in (1) and (2) are equal, while Column (5) repeats this for a test of equality between columns (1) and (3). Columns (6) through (10) repeat this exercise among the cases where there are plants within the same 2-digit SIC industry as the MDP plant. In these columns, the plant characteristics are calculated among the plants in the same 2-digit industry.

This exercise provides an opportunity to assess the validity of the research design, as measured by pre-existing observable county and plant characteristics. To the extent that these observable characteristics are balanced among winning and losing counties, this should lend credibility to the analysis. The comparison between winner counties and the rest of the United States provides an opportunity to assess the validity of the type of analysis that would be undertaken in the absence of a quasi-experiment.

The top panel reports on county-level characteristics measured in the year before the assigned plant opening and the percentage change between 7 years and 1 year before the opening. It is evident that compared to the rest of the country, winning counties have higher incomes, population and population growth, labor force participation rates and growth, and a higher share of labor in manufacturing. Among the 8 variables in this panel, 6 of the 8 differences would be judged to be statistically significant at conventional levels. These differences are substantially mitigated when the winners are compared to losers, and this is reflected in the fact that 3 of the 8 variables are statistically different at the 5% level but none are at the 1% level. Notably, the raw differences between winners and losers within the subset of cases where there are plants in the same 2 digit SIC industry are generally smaller, and none of them would be judged to be statistically significant.

---

<sup>24</sup> The losing county entries in column (2) are weighted in the following manner. Losing counties are weighted by the inverse of their number in that case. Losing plants are weighted by the inverse of their number per-county, multiplied by the inverse of the number of losing counties in their case. The result is that each county (and each plant within each county) is given equal weight within the case and then all cases are given equal weight.

The second panel reports on the number of sample plants and provides information on some of their characteristics. In light of our sample selection criteria, the number of plants is of special interest. On average, there are 18.8 plants in the winner counties and 25.6 in the loser ones (and just 8.0 in the United States). The covariates are well balanced between plants in winning and losing counties; in fact, there are no statistically significant differences either among all plants or among the plants within the same 2-digit industry.<sup>25</sup>

Overall, Table 3 has demonstrated that the MDP winner-loser research design balances many (although not all) observable county-level and plant-level covariates. In the subsequent analysis, we demonstrate that trends in TFP were similar in winning and losing counties prior to the opening of the MDP, which lends further credibility to this design. Of course, this exercise does not guarantee that unobserved variables are balanced across winner and loser counties or their plants. The next section outlines our full econometric model and highlights the exact assumptions necessary for consistent estimation.

#### IV. Econometric Model

Building on the model in section I, we start by assuming that incumbent plants use the following Cobb-Douglas technology:

$$(6) \quad Y_{pijt} = A_{pijt} L_{pijt}^{\beta_1} K_{pijt}^{\beta_2} K_{pijt}^{\beta_3} M_{pijt}^{\beta_4}$$

where  $p$  references plant,  $i$  industry,  $j$  case, and  $t$  year;  $Y_{ijct}$  is the total value of shipments;  $A_{pijt}$  is TFP; and we allow total labor hours of production ( $L_{pijt}$ ), building capital stock ( $K_{pijt}^B$ ), machinery and equipment capital stock ( $K_{pijt}^E$ ), and the dollar value of materials ( $M_{pijt}$ ) to have separate impacts on output. In practice, the two capital stock variables are calculated with the permanent inventory method that uses earlier years of the data on book values and subsequent investment.<sup>26</sup>

Recall that equation (1) in Section I allows for agglomeration spillovers by assuming that TFP is a function of the number of firms that are active in a county:  $A_{pijt} = A(N_{pijt})$ . Here we also allow for some additional heterogeneity in  $A_{pijt}$ . In particular, we generalize equation (1) by allowing for permanent

<sup>25</sup> Roughly 20% of the winners were in the Rust Belt, compared to roughly 25% of the losers (where the Rust Belt is defined as MI, IN, OH, PA, NJ, IL, WI, NY). Roughly 65% of the winners were in the South, compared to roughly 45% of the losers.

<sup>26</sup> For the first date available, plants' historical capital stock book values are deflated to constant dollars using BEA data by 2-digit industry. In all periods, plants' investment is deflated to the same constant dollars using Federal Reserve data by 3-digit industry. Changes in the capital stock are constructed by depreciating the initial deflated capital stock using Federal Reserve depreciation rates and adding deflated investment. In each year, productive capital stock is defined as the average over the beginning and ending values, plus the deflated level of capital rentals. The analysis is performed separately for building capital and machinery capital. This procedure is described further by Becker et al. (2005), Chiang (2004), and Davis et al. (1996), from whose files we gratefully obtained deflators.

differences in TFP across plants ( $\alpha_p$ ), cases ( $\lambda_j$ ), industry-specific time-varying shocks to TFP ( $\mu_{it}$ ), and a stochastic error term ( $\varepsilon_{pijt}$ ):

$$\ln(A_{pijt}) = \alpha_p + \mu_{it} + \lambda_j + \varepsilon_{pijt} + A(N_{pijt}).$$

The goal is to estimate the causal effect of winning a plant on incumbent plants' TFP. To do so, we need to impose some structure on  $A(N_{pijt})$ . In particular, we use a specification that allows for the new plant in winning counties to affect both the level of TFP as well as its growth over time:

$$(7) \quad \begin{aligned} \ln(A_{pijt}) = & \delta 1(\text{Winner})_{pj} + \psi \text{trend}_{jt} + \Omega (\text{trend} * 1(\text{Winner}))_{pjt} \\ & + \kappa 1(\tau \geq 0)_{jt} + \gamma (\text{trend} * 1(\tau \geq 0))_{jt} \\ & + \theta_1 (1(\text{Winner}) * 1(\tau \geq 0))_{pjt} + \theta_2 (\text{trend} * 1(\text{Winner}) * 1(\tau \geq 0))_{pjt} \\ & + \alpha_p + \mu_{it} + \lambda_j + \varepsilon_{pijt} \end{aligned}$$

where  $1(\text{Winner})$  is a dummy equal to 1 if plant  $p$  is located in a winner county; and  $\tau$  denotes year, but it is normalized so that for each case the assigned year of the plant opening is  $\tau = 0$ . The variable  $\text{trend}_{jt}$  is a simple time trend.

Combining equations (6) and (7) and taking logs, we obtain the regression equation that forms the basis of our empirical analysis:

$$(8) \quad \begin{aligned} \ln(Y_{pijt}) = & \beta_1 \ln(L_{pijt}) + \beta_2 \ln(K^B_{pijt}) + \beta_3 \ln(K^E_{pijt}) + \beta_4 \ln(M_{pijt}) \\ & + \delta 1(\text{Winner})_{pj} + \psi \text{trend}_{jt} + \Omega (\text{trend} * 1(\text{Winner}))_{pjt} + \kappa 1(\tau \geq 0)_{jt} + \gamma (\text{trend} * 1(\tau \geq 0))_{jt} \\ & + \theta_1 (1(\text{Winner}) * 1(\tau \geq 0))_{pjt} + \theta_2 (\text{trend} * 1(\text{Winner}) * 1(\tau \geq 0))_{pjt} \\ & + \alpha_p + \mu_{it} + \lambda_j + \varepsilon_{pijt}. \end{aligned}$$

Equation (8) is an augmented Cobb-Douglas production function that allows labor, building capital, machinery capital, and materials to have differential impacts on output. The paper's focus is the estimation of the impact of the new plant on incumbent plants' TFP, so the parameters of interest are  $\theta_1$  and  $\theta_2$ , which are the spillover effects. The former tests for a mean shift in TFP among incumbent plants in the winning county after the opening of the MDP, while the latter allows for a trend break in TFP among the same plants.

In practice, we estimate two variants of Equation (8). In some specifications, we fit a parsimonious model that simply tests for a mean shift. In this model, any productivity effect is assumed to occur immediately and to remain constant over time. Specifically, we make the restrictions that  $\psi = \Omega = \gamma = \theta_2 = 0$ , which assumes that differential trends are not relevant here. This specification is essentially a difference in differences estimator and we refer to it as Model 1. Formally, after adjustment for the inputs,  $1(\text{Winner})_p$ , and  $1(\tau \geq 0)_{jt}$ , the consistency of  $\theta_1$  in this model requires the assumption that  $E[(1(\text{Winner}) * 1(\tau \geq 0))_{pjt} \varepsilon_{pijt} | \alpha_p, \mu_{it}, \lambda_j] = 0$ .

In other specifications, we estimate the model without imposing such restrictions on the trends. In

other words, we estimate the entire Equation (8). We label this Model 2. While Model 1 only allows for a mean shift in productivity, this specification allows both for a mean shift and a trend break in productivity. In other words, Model 2 allows us to investigate whether any productivity effect occurs immediately and whether the impact evolves over time. This specification is demanding of the data, because our sample is only balanced through  $\tau = 5$  so there are only 6 years per case to estimate  $\theta_1$  and  $\theta_2$ .<sup>27</sup>

Equation (8) allows for unobserved determinants of TFP that are unrelated to the plant opening but could be confounded with the spillover effects if not properly accounted for. It includes a differential intercept for all observations from winning counties,  $\delta$ . It also allows for a common time trend,  $\psi$ . The parameter  $\Omega$  allows the time trend to differ for winning counties prior to the plant opening. This will serve as an important way to assess the validity of this research design. Finally,  $\kappa$  and  $\gamma$  capture the level change and trend break in TFP common to plants in winning and losing counties after the MDP opening (i.e., when  $\tau \geq 0$ ).

In addition, all our models include three sets of fixed effects. First, they include separate fixed effects for each plant,  $\alpha_p$ , so the comparisons are within a plant. Second,  $\mu_{it}$  represents the parameters associated with a vector of 2-digit SIC industry by year fixed effects to account for industry-specific shocks to TFP. Third, the  $\lambda_j$ 's are separate fixed effects for each case that ensure that the impact of the MDP's opening is identified from comparisons within a winner-loser pair; they are a way to retain the intuitive appeal of pairwise differencing in a regression framework.

A few further estimation details bear noting. First, unobserved demand shocks are likely to affect input utilization, and this raises the possibility that the estimated  $\beta$ 's are inconsistent (see, e.g., Griliches and Mairesse 1995). This has been a topic of considerable research and we are unaware of a bullet-proof solution. We implement the standard fixes including modeling the inputs with alternative functional forms (e.g., the translog), using cost shares at the plant and industry-level rather than estimating the  $\beta$ 's, and controlling for flexible functions of investment, capital and materials (Syverson 2004; van Biesebroeck 2004; Olley and Pakes 1996; Levinsohn and Petrin 2003). Additionally, we experiment with adding fixed effects for region by year or region by industry by year, and allowing the effect of inputs to differ by industry or by winner- and post-MDP status. The basic results are unchanged by these alterations in the specification. We also note that unobserved demand shocks are only a concern for the consistent estimation of the parameters of interest (i.e.,  $\theta_1$  and  $\theta_2$ ) if they systematically affect incumbent plants in winning counties in the years after the MDP's opening, after adjustment for the rich set of covariates in equation (8).

---

<sup>27</sup> This specification allows for spillovers to affect the level of TFP and to grow over time as in Glaeser and Mare (2001).

Second, in some cases this equation is fit on a sample of plants from the entire country, but in most specifications the sample is limited to plants from winning and losing counties in the *ASM* for every year from  $\tau = -8$  through  $\tau = 0$ . When data from the entire country is used, the sample is limited to plants that are in the *ASM* for at least 14 consecutive years. The smaller sample of plants from the winning and losing counties allows for the impact of the inputs and the industry shocks to differ in the winning-losing county sample from the rest of the country. Finally, for most of the analysis, we further restrict the sample to observations in the years between  $\tau = -7$  and  $\tau = 5$ . Due to the dates of the MDP openings, this is the longest period for which we have data from all cases.<sup>28</sup>

Third, we probe the validity and robustness of our estimates with a number of alternative specifications. For example, we investigate whether changes in the price of output, capital utilization, public investment, or attrition influence the estimates.

Fourth, all of the reported standard errors are clustered at the county level to account for the correlation in outcomes among plants in the same county.<sup>29</sup>

Fifth, we focus on weighted versions of equation (8). Specifically, the specifications are weighted by the square root of the total value of shipments in  $\tau = -8$  to account for heteroskedasticity associated with differences in plant size. This weighting also means that the results measure the change in productivity for the average dollar of output, which in our view is more meaningful than the impact of the MDP on the average plant.

The analysis will also explore two additional issues. It will report on the fitting of versions of equation (8) that interact the spillover variables with measures of economic distance between the MDP and the incumbent plant. These specifications assess whether the magnitude of the estimated spillovers varies with economic distance. Finally, the paper will assess whether the MDP's opening affects local skill adjusted wages and the entry and exit decisions of plants in the MDP's county.

## V. Results

This section is divided into four subsections. The first reports baseline estimates of the effect of the opening of a new Million Dollar Plant on the productivity of incumbent plants in the same county through the estimation of equation (8). The second subsection discusses the validity of our design and explores the robustness of our estimates to a variety of different specifications. The third subsection explores potential channels for the agglomeration effects by testing whether the estimated spillovers vary as a function of economic distance. The final subsection discusses the implications of our estimates for

---

<sup>28</sup> Data from all cases is also available for  $\tau = -8$ , but output in this period is used to weight the regressions.

<sup>29</sup> We experimented with clustering the standard errors at the 2-digit SIC by county level, but this occasionally produced variance-covariance matrices that weren't positive definite. In instances where they were positive definite, these standard errors were similar to those from clustering at the county level.

the profits of local firms.

### A. Baseline Estimates

Columns (1) and (2) of Table 4 report estimated parameters and their standard errors from a version of equation (8). Specifically, the natural log of output is regressed on the natural log of inputs, year by 2-digit SIC industry fixed effects, plant fixed effects, and the event time indicators in a sample that is restricted to the years  $\tau = -7$  through  $\tau = 5$ . The parameters associated with the event time indicators report mean TFP in winning and losing counties, respectively, in each event year relative to the year before the MDP opened (i.e., we have subtracted off the  $\tau = -1$  parameter estimates for the winners (losers) from the winner (loser) estimates for each event year). Column (3) reports the difference between the estimated TFP levels within each year.

The top panel of Figure 1 separately plots the mean TFP levels for winner and loser counties (taken from columns (1) and (2) of Table 4) against  $\tau$ . The bottom panel of Figure 1 plots the difference in the estimated winner and loser coefficients against  $\tau$ . Thus, it is a graphical version of column (3) of Table 4.

Two important findings are apparent in these figures. First, the trends in TFP among incumbent plants were very similar in the winning and losing counties in the years before the MDP opening. In fact, a statistical test fails to reject that the trends were equal. This finding supports the validity of our identifying assumption that incumbent plants in losing counties provide a valid counterfactual for incumbents in winning counties.

Second, there is a sharp upward break in the difference in TFP between the winning and losing counties beginning with the year that the plant opened. The top graph reveals that this relative improvement is due to the continued decline in TFP in losing counties and a flattening out of the TFP trend in winning counties. The figures also serve to underscore the importance of the availability of losing counties as a counterfactual. For example, a naïve comparison of mean TFP in winning counties before and after the opening would lead to the conclusion that the opening had a small negative impact on incumbents' TFP. Overall, these graphs reveal much of the paper's primary finding. This relative increase in TFP among incumbent plants in winning counties will be confirmed throughout the battery of tests in the remainder of the paper.

Before proceeding, it is worth noting that the TFP of incumbent plants was declining in both sets of counties in advance of the MDP plant opening. This finding may appear surprising, because productivity increases over time in the overall economy. However, this downward trend in TFP among large manufacturing plants appears to be a general phenomenon and is not specific to plants in winning and losing counties. For example, we estimated augmented Cobb-Douglas production functions where



the constant dollar total value of shipments is the dependent variable and the covariates are the capital stock, labor, materials (also in constant dollars), 2-digit industry fixed effects, and plant fixed effects. The equations are weighted by the total value of shipments and the sample is restricted to plants in the *ASM* for at least 14 consecutive years in the period 1973-1998. Thus, the approach and sample selection rule are similar in spirit to the ones used in the MDP analyses.

The average 6 year change in TFP calculated over the years preceding the MDP openings was a statistically significant decline of 4.7%.<sup>30</sup> This decline is similar to the decline among incumbents in winning and losing counties in Figure 1. To the best of our knowledge, this finding of declining TFP among large manufacturing plants has not been noted previously.<sup>31</sup>

Now turning to the statistical models, the first four columns of Table 5 present the results from fitting different versions of equation (8). Models 1 and 2 are in Panels A and B, respectively. Panel A reports the estimated mean shift parameter,  $\theta_1$ , and its standard error (in parentheses) in the “Mean Shift” row. Panel B reports the estimated impact of the MDP on incumbent plants’ TFP evaluated at  $\tau = 5$  in the “Effect after 5 years” row, which is determined by  $\theta_1$  and  $\theta_2$  that are also both reported.<sup>32</sup> The row “Pre-Trend” contains the coefficient measuring the difference in the pre-existing trends between plants in the winning and losing counties. In all of these specifications, the estimated impact of the MDP’s opening is determined during the period where  $-7 \leq \tau \leq 5$ , as the sample is balanced during these years.

In columns (1) and (2), the sample includes all manufacturing plants in the *ASM* that report data for at least 14 consecutive years, excluding all plants owned by the MDP firm. In column (3), the sample is restricted to include only plants in counties that won or lost a MDP. This restriction means that the impact of the inputs and the industry-year fixed effects are estimated solely from plants in these counties. Incumbent plants are now required to be in the data only for  $-8 \leq \tau \leq 0$  (not also for 14 consecutive years, though this does not change the results). Finally in column (4), the sample is restricted further to include only plant-year observations within the period of interest (where  $\tau$  ranges from -7 through 5). This forces the input parameters and industry-year fixed effects to be estimated solely on plant by year observations that identify the spillover parameters. This sample is used throughout the remainder of the paper. Estimation details are noted at the bottom of the table and apply to both Models 1 and 2.

The entries in Table 5 confirm the visual impression from Figure 1 that the opening of the MDP

---

<sup>30</sup> The 6 year average is a weighted averaged calculated over the 6 year periods before each of the MDP openings where the weights are the number of plant openings associated with each 6 year period. For example, if there are 2 plant openings in 1987 and 1 in 1988, then the average change between 1980 and 1986 receives twice the weight as the average change calculated between 1981 and 1987.

<sup>31</sup> Since we are looking at large plants that have been active for a large number of years, we speculate that this decline may have to do with aging. Additionally, many of the years preceding the MDP openings were in the late 1970s and early 1980s, which was a period of poor economic performance. Foster, Haltiwanger, and Krizan (2000) have documented that within plant productivity growth is positively correlated with the economic cycle.

<sup>32</sup> This is calculated as  $\theta_1 + 6\theta_2$ , because we allow the MDP to affect outcomes from  $\tau = 0$  through  $\tau = 5$ .

is associated with a substantial increase in TFP among incumbent plants in winning counties. Specifically, Model 1 implies an increase in TFP of roughly 4.8%. As the figure highlighted, however, the impact on TFP appeared to be increasing over time so Model 2 seems more appropriate. This model's results suggest that the MDP's opening is associated with an approximately 12% increase in TFP five years later. The estimates from both models would be judged to be statistically different from zero by conventional criteria and are unaffected by the changes in the specifications. Furthermore, the entries in the "Pre-trend" row demonstrate that the null hypothesis of equal trends in TFP among incumbents in winning and losing counties cannot be rejected.

The numbers in square brackets in column 4 measure the average size of the spillover from a MDP opening in millions of 2006\$. This figure is calculated by multiplying the estimated impact by the mean value of incumbent plants' total shipments in winning counties in  $\tau = -1$ . This calculation indicates that the increase in TFP from a MDP was associated with an increase in total output of about \$170 million per year in Model 1. The Model 2 estimate is even larger, suggesting an increase in output of roughly \$429 million in year  $\tau = 5$ . These numbers are large, with the Model 2 effect at  $\tau = 5$  nearly the average level of MDP output.

Column (5) presents the results from a "naïve" estimator that is based on using plant openings without an explicit counterfactual. Specifically, a set of 47 plant openings were randomly chosen from the *ASM* in the same years and industries as the MDP openings. The remainder of the sample includes all manufacturing plants in the *ASM*, not also owned by the randomly chosen firm, that report data for at least 14 consecutive years. With these data, we fit a regression of the natural log of output on the natural log of inputs, year by 2-digit SIC fixed effects, and plant fixed effects. In Model 1, two additional dummy variables are included for whether the plant is in a winning county 7 to 1 years before the randomly chosen opening or 0 to 5 years after. The reported mean shift is the difference in these two coefficients (i.e., the average change in TFP following the opening). In Model 2, the same two dummy variables are included along with pre- and post-trend variables. The shift in level and trend are reported, along with the pre-trend and the total effect evaluated after 5 years.

This naïve "first-difference" style estimator indicates that the opening of a new plant is associated with a -6% to -8% effect on incumbent plants' TFP, depending on the model. If the estimates from the MDP research design are correct, then this naïve approach understates the extent of spillovers by 13% (Model 1) to 18% (Model 2). Interestingly, the parameter on the "pre-trend" indicates that the TFP of the incumbent plants was on a downward trend in advance of the openings in the counties that attracted these new plants. This is similar to what is observed in our MDP sample of winners. Overall, the primary message is that the absence of a credible research design can lead to misleading inferences in this setting.

It is natural to wonder about the degree of heterogeneity in the treatment effects from the 47

separate case studies that underlie the estimates presented thus far. Figure 2 explores this heterogeneity by plotting case-specific estimates of parameter  $\theta_1$  in Model 1 and their 95% confidence interval. Specifically, the Figure plots results from a version of Model 1 that interacts the variable  $(1(\text{Winner}) * 1(\tau \geq 0))$  with indicators for each of the cases. There are 45 estimates of  $\theta_1$ , one for each case. (Results from two cases were omitted to comply with the Census Bureau's confidentiality rules). The figure reveals that there is heterogeneity in the estimated impacts on TFP of incumbent plants. 27 of the 45 estimates are positive. 13 of the positive estimates would be judged to be statistically different from zero at the 5% level, while the comparable figure for the negative estimates is 9.

We assessed whether the estimated spillover effects are related to characteristics of the MDP. Specifically, we regressed the estimates against three measures of the MDP's size, whether the MDP is owned by a foreign company, and whether it is an auto company. When these multiple measures were included jointly, none were significantly related to the estimated effect of the MDP's opening.<sup>33</sup>

Ultimately, TFP is a residual and residual labeling must be done cautiously. As an alternative way to examine the impact of the MDP on incumbents' productivity, we have estimated directly the impacts of a MDP opening on output (unadjusted for inputs) and inputs. The intent is to contrast the changes in outputs and inputs to shed light on whether productivity increased without imposing the structural assumptions of the production function. Put another way, are the incumbents producing more with less after the MDP's opening? Appendix Table 1 reports on estimates of the impact of a MDP opening on incumbents' output and usage of inputs. The estimates are from the Model 1 and Model 2 versions of equation (8) with the key difference that these equations do not include the inputs as covariates. Again, we use the Model 2 results to estimate the impact of the opening 5 years afterwards.

Column (1) reports that the MDP opening is associated with an 8-12% increase in output. Columns (2) through (5) report the results for the four inputs. It is striking that the change in all of the inputs is roughly equal to or less than the increase in output. The Model 2 results are especially noteworthy, because the 8% increase in output is accompanied by no increase in either form of capital. Overall, the results in Appendix Table 1 indicate that, after the MDP's opening, incumbent plants produced more with less; that is, they suggest that these plants became more productive, and this is

---

<sup>33</sup> Separate regressions of the case specific effects on the MDP's total output or the MDP's total labor force generated statistically significant negative coefficients. This result is consistent with the possibility that when the MDP is very large incumbents are left to hire labor and other inputs that are inferior in unobserved ways. On the other hand, we failed to find any significant differences when separately testing whether the productivity effect varied by the ratio of the MDP's output to county-wide manufacturing output, whether the MDP is owned by a foreign company, or whether the MDP is an auto company.

consistent with the TFP increases uncovered in Table 5.<sup>34</sup>

## B. Threats to Validity

Estimates in Table 5 appear to be consistent with significant agglomeration spillovers generated by MDP openings. Although the comparisons in Table 3 and the similarity of the pre-existing trends in TFP in winning and losing counties support the validity of the research design, it is, of course, possible that there is a form of unobserved heterogeneity that accounts for the higher levels of TFP in winning counties after the MDP's opening. Consequently, this subsection investigates several possible alternative interpretations of the estimated spillover effects and explores the robustness of the estimates to a variety of different assumptions. Specifically, we investigate (i) the role of functional form assumptions and the possible presence of unobserved industry and regional shocks; (ii) the endogeneity of inputs; (iii) changes in the price of output; (iv) the possible role of public investment; (v) changes in capital utilization; and (vi) attrition.

**(i) Functional Form, Industry Shocks and Regional Shocks.** Table 6 reports on a series of specification checks. For convenience, column (1) reports the results from the preferred specification in column (4) specification of Table 5. These estimates are intended to serve as a basis of comparison for the estimates in the remainder of the table.

We begin by generalizing our assumption on technology. Estimates in Table 5 assume a Cobb-Douglas technology. In column (2) of Table 6, the inputs are modeled with the translog functional form. Column (3) is based on a Cobb-Douglas technology but allows the effect of each production input to differ at the 2-digit SIC level. This model accounts for possible differences in technology across industries, as well as for possible differences in the quality of inputs used by different industries. For example, it is possible that even if technology was similar across different manufacturers, some industries use more skilled labor than others. Column (4) allows the effect of the inputs to differ in winning/losing counties and before/after the MDP opening.

Columns (5) and (6) add census division by year fixed effects and census division by year by 2-digit industry fixed effects. These specifications aim to purge the spillover effects of unobserved region-wide shocks or region by industry shocks to productivity that might be correlated with the probability of winning a MDP.

Taken together, the results in columns (2) through (6) of Table 6 are striking. The estimates appear to be insensitive to the specific functional form of the production function. None of the

---

<sup>34</sup> Column (6) of Appendix Table 1 presents evidence on changes in the capital/labor ratio. The model suggests that firms should substitute away from labor and toward capital. The estimated change in the capital/labor ratio is poorly determined, making definitive conclusions unwarranted, but the point estimate is not supportive of this prediction.

specifications contradict the findings from the baseline specification in Table 5. Although many of the estimates are smaller than the baseline ones, the magnitude of the decline is modest. For example, they are all within one standard error of the baseline estimate in both Models 1 and 2. Overall, these results fail to undermine the conclusion from Table 5 that the opening of a MDP leads to a substantial increase in TFP among incumbent plants and this is consistent with theories of spillovers.

**(ii) Endogeneity of Inputs.** An important conceptual concern is that capital and labor inputs should be treated as endogenous, because the same forces that determine output also determine a firm's optimal choice of inputs (Griliches and Mairesse 1995). Unlike the usual estimation of production functions, our aim is the consistent estimation of the spillover parameters,  $\theta_1$  and  $\theta_2$ , so the endogeneity of capital and labor is only relevant to the extent that it results in biased estimates of these parameters. This subsection employs the productivity literature's state of the art techniques to control for the endogeneity of capital and labor to assess this issue's relevance in this paper's setting.

We do this in two ways. First, in columns (7) and (8) of Table 6, we calculate TFP for each plant by fixing the parameters on the inputs at the relevant input's share of total costs (van Biesebroeck 2004; Syverson 2004; Foster, Haltiwanger, and Syverson 2007). This method may mitigate any bias in the estimation of the parameters on the inputs associated with unobserved demand shocks. In these two columns, the cost shares are calculated at the plant level and the 3-digit SIC industry level over the full sample, respectively. The estimated spillover effects are largely insensitive to this restriction.

Second, Table 7 presents estimates based on the widely-used methodologies proposed by Olley and Pakes (1996) and Levinsohn and Petrin (2003). These methods are based on the result that, under certain conditions, adjustment for investment or intermediate inputs (e.g., materials) will remove the correlation between input levels and unobserved shocks to output. For example, the column (2) specification adds 4<sup>th</sup> degree polynomial functions of log building investment and log machinery/equipment investment to the baseline specification.<sup>35</sup> The column (3) specification adds 4<sup>th</sup> degree polynomials in the two types of log capital stocks to the column (2) equation. The column (4) specification is even richer as it adds all the "own" interactions between polynomials in current investment and capital (i.e., the building investment polynomial is interacted with the building capital stock polynomial, but it is not also interacted with the machinery/equipment polynomials for stocks or investment) to column (3). Column (5) adds a 4<sup>th</sup> degree function of log materials to the baseline specification. In the column (6) specification, a 4<sup>th</sup> degree polynomial in materials is fully interacted with 4<sup>th</sup> degree polynomials in building capital and machinery capital. Column (7) includes fourth-degree polynomials in log materials and log investment and log capital stock for both types of capital (not

---

<sup>35</sup> The exact measure used is  $\log(1+\text{investment})$ , so zero values are not dropped. The results are very similar when including polynomial functions of the level of investment and a dummy variable for values equal to zero.

interacted). Finally, column (8) includes the controls from the columns (4) and (6) specifications.

The estimated spillover effects in Table 7 are generally consistent with the findings from the baseline specification. This finding holds in both Models 1 and 2. Overall, this exercise fails to suggest that the possible endogeneity of labor and capital is the source of the estimated productivity spillovers.

**(iii) Changes in the Price of Output.** Another concern is that the theoretically correct dependent variable is the *quantity* of output. However, due to the data limitations faced by virtually all of the rest of the productivity literature, the dependent variable in our models is the *value* of output or price multiplied by quantity. Consequently, it is possible that the estimated spillover effect reflects higher prices, instead of higher productivity.

We do not expect this to be a major factor in our context. First, the sample is comprised of manufacturing establishments that generally produce nationally traded goods. Therefore, it is likely that in many cases the price of output is set at the national level, and has little to do with what happens in the county where the goods are produced. In the extreme case of a perfectly competitive industry that produces a nationally traded good, there should be no effect on prices.

Second, we tested whether the size of the estimated productivity effect is larger in industries that are more regional or more concentrated. The idea is that if price increases are possible, then they should be larger in industries that are more local and/or in industries that are less competitive. Consider for example, the case of an industry that sells mainly at the local level (e.g., cement). The opening of a large new plant may ultimately increase incumbents' demand by raising local income (even though the initial effect on demand should be negative). If the industry is not very competitive, the increase in demand may ultimately lead to price increases for the incumbents' output.

To implement the test, we estimated a Model 1 version of equation (8) that interacts  $1(\text{Winner})_{pj}$ ,  $1(\tau \geq 0)_{jt}$ , and  $(1(\text{Winner}) * 1(\tau \geq 0))_{pjt}$  with incumbents' industry-specific measure of average distance traveled by output between production and consumption. We also conducted a similar exercise where these same variables are interacted with a measure of the incumbent's industry concentration.<sup>36</sup> These specifications fail to produce evidence that the estimated spillover effects are larger in more local industries or more concentrated industries; in fact, there is some evidence for larger effects on incumbent plants that ship their products further. Our conclusion is that price increases do not appear to be the source of the estimated spillover effects.

**(iv) Public Investment.** State and local governments frequently offer substantial subsidies to new manufacturing plants to locate within their jurisdictions. These incentives can include tax breaks,

---

<sup>36</sup> The information on distance is from Weiss (1972). Distance varies between 52 and 1337 miles, with a mean of 498. Examples of regional industries are: hydraulic cement, iron and steel products, metal scrap and waste tailings, ice cream and related frozen desserts, and prefabricated wooden buildings. The information on industry concentration is from the Bureau of Census ("Concentration Ratios", 2002).

worker training funds, the construction of roads, and other infrastructure investments. It is possible that some of the public investment in infrastructure benefits firms other than the beneficiary of the incentive package. For example, the construction of a new road intended for a MDP plant may also benefit the productivity of some of the incumbent firms. If the productivity gains we have documented are due to public investment, then it is inappropriate to interpret them as evidence of spillovers.

To investigate this possibility, we estimated the effect of MDP openings on government total capital expenditures and government construction expenditures with data from the *Annual Survey of Governments*. In models similar to equation (8), we find that the opening of a MDP plant is associated with statistically insignificant increases in capital and construction expenditures. In fact, in most specifications the estimated impact of a MDP opening is negative and statistically insignificant. Even in the specifications that produce positive insignificant estimates, there is no plausible rate of return that could generate a meaningful portion of the productivity gains in winning counties. Based on these measures of public investment, it seems reasonable to conclude that public investment cannot explain the paper's results.

**(v) Changes in Capital Utilization.** Another potential threat to validity is that incumbent plants may respond to the MDPs by increasing the intensity of their capital usage. This could happen if depressed counties where the existing capital stock had been used below capacity win the MDPs and increase production simply by operating their capital stock closer to capacity. As an indirect test of this possibility, we estimated whether the MDP's opening affected the ratio of the dollar value of energy usage (which is increasing in the use of the capital stock) to the capital stock. Column 7 of Appendix Table 1 reports small and insignificant changes in this measure. Thus, we conclude that greater capacity utilization is unlikely to be the source of the findings of productivity spillovers.

**(vi) Attrition of Sample Plants.** Differential attrition in the sample of incumbent plants in winning and losing counties could contribute to the measured differential in productivity trends among survivors after the MDP's opening. This attrition could result from plants shutting down operations or from plants continuing operations but dropping out of the group of plants that are surveyed with certainty as part of the ASM.<sup>37</sup>

The available evidence suggests that differential attrition is unlikely to explain the finding of spillovers in winning counties. First, in the baseline sample (i.e., the one used to produce the results in column (4) of Table 5 and in Tables 6 and 7), 72% of the winning county plants operating in the year of the MDP opening were still in the sample at its end (i.e.,  $\tau = 5$ ). The analogous figure in losing counties is 68%. The slightly larger attrition rate in losing counties is consistent with the paper's primary result.

---

<sup>37</sup> Recall, establishments are sampled with certainty if they are part of a company with manufacturing shipments exceeding \$500 million or their total employment was at least 250.

Specifically, one seemingly reasonable interpretation of this result is that the MDP's opening allowed some winning county plants to remain open that would have otherwise closed. Thus to the extent that a MDP opening keeps weaker plants operating, the above analysis will underestimate the overall TFP increase. Second, the estimation of equation (8) on the sample of plants that is present for all years from -7 to +5 yields results that are qualitatively similar to those from the full sample. Third, the null hypothesis of equal trends in TFP among attriting plants in winning and losing counties prior to the MDP's opening cannot be rejected; the TFP trend in winning counties minus the TFP trend in losing counties was -0.0052 (0.0080).<sup>38</sup>

### C. Estimates of Spillovers by Economic Distance

What can explain the productivity gains uncovered above? Section I A discussed some possible mechanisms that may be responsible for agglomeration spillovers. Tables 8 and 9 attempt to shed some light on the possible mechanisms by investigating how the measured spillover effect varies as a function of economic distance.

**By Industry.** Table 8 shows separate estimates from the baseline model for samples of incumbent plants in the MDP's 2-digit industry and all other industries. In general, one might expect that the effects of spillovers decline with economic distance (equation 5'). Although looking within-industry does not shed direct light on which channel is the source of the spillovers, it seems reasonable to presume that spillovers would be larger within an industry.

In examining the 2-digit SIC MDP industry results, it is important to recall that just 16 of the 47 cases have plants in the MDP's 2-digit industry. We also note that there can be substantial heterogeneity in technologies and labor forces among the industries within a 2-digit SIC industry. However, this research design and the available data do not permit an examination at finer industry definitions.

Column 1 of Table 8 repeats the all industries estimates from column (4) of Table 5 and is intended to serve as a basis of comparison. Columns (2) and (3) report on estimates from the baseline specification for incumbent plants in the MDP's 2-digit industry and all other industries, respectively. The entries in these columns are from the same regression. Just as in Table 5, the numbers in square brackets convert the parameter estimates into millions of 2006\$.

---

<sup>38</sup> In addition to the specification checks described in this section, we also tested whether the results are sensitive to the choice of the date of the MDP's opening. The estimated spillovers are virtually unchanged when we use the year that the plant is first observed in SSEL as the MDP's opening date. When the year of the MDP article in *Site Selection* is used as the plant's opening date, the Model 1 results are nearly identical to those in the Table 5 column (4) specification and roughly 5% in the Model 2 specification. When the estimating equation is unweighted, the evidence in favor of spillover effects is weaker indicating that the spillovers are concentrated among the larger plants in the sample. As discussed above, our view is that the economically relevant concept of spillover is the change in productivity for the average dollar of output, rather than the average plant.



The impacts are substantially larger in the own 2-digit industry. For example, the estimated increase in TFP for plants in the same 2-digit industry is a statistically significant 17% in Model 1 and a poorly determined 33% at  $\tau = 5$  in Model 2. In contrast, the estimates for plants in other industries are a statistically insignificant 3.3% in Model 1 and marginally significant 8.9% in Model 2. These basic findings in the own 2-digit and other industries are robust to the different specifications in tables 6 and 7.<sup>39</sup>

Figures 3 and 4 provide 2-digit MDP industry and other industry analogues to Figure 1. Importantly, there is not evidence of differential trends in the years before the MDP's opening and statistical tests confirm this visual impression. The 2-digit MDP industry estimates are noisy due to the small sample size, which was also evident in the statistical results. Just as in Figure 1, the estimated impact reflects the continuation of a downward trend in TFP in losing counties and a cessation of the downward trend in winning counties.

**By Direct Measure of Economic Proximity.** Having found that the spillover is larger for incumbent plants in the MDP industry, we investigate the role of economic proximity more directly by using several explicit measures of economic proximity that capture worker flows, technological proximity, and input-output flows. To ease the interpretation, the economic proximity or linkage variables are standardized to have a mean of zero and standard deviation of one. In all cases, a positive value indicates a “closer” relationship between the industries.

Specifically, we estimate the following equation:

$$(8') \quad \ln(Y_{pijt}) = \beta_1 \ln(L_{pijt}) + \beta_2 \ln(K_{pijt}^B) + \beta_3 \ln(K_{pijt}^E) + \beta_4 \ln(M_{pijt}) \\ + \delta 1(\text{Winner})_{pj} + \kappa 1(\tau \geq 0)_{jt} + \theta_1 (1(\text{Winner}) * 1(\tau \geq 0))_{pj} \\ + \pi_1 1(\text{Winner})_{pj} * \text{Proximity}_{ij} + \pi_2 (1(\tau \geq 0))_{jt} * \text{Proximity}_{ij} \\ + \pi_3 (1(\text{Winner}) * 1(\tau \geq 0))_{pj} * \text{Proximity}_{ij} + \alpha_p + \mu_{it} + \lambda_j + \varepsilon_{pijt}.$$

where  $\text{Proximity}_{ij}$  is a measure of economic proximity between the incumbent plant industry and the MDP industry. This equation is simply an augmented version of Model 1 that adds interactions of the industry linkage variables with  $1(\text{Winner})_{pj}$ ,  $1(\tau \geq 0)_{jt}$ , and  $(1(\text{Winner}) * 1(\tau \geq 0))_{pj}$ . The coefficient of interest is  $\pi_3$ , which is the coefficient on the triple interaction between the dummy for winner, the dummy for “after,” and the measure of proximity. This coefficient assesses whether “closer” industries benefit more from the MDP's opening. A positive coefficient means that the estimated productivity spillover is larger after the MDP opening for incumbents that are geographically and economically close to the new plant,

---

<sup>39</sup> Within the same 2-digit SIC, 71% of incumbents in winning counties and 69% of incumbents in losing counties were still in the sample 5 years after the opening. Additionally, attriting plants within the same 2-digit SIC were also on statistically indistinguishable trends prior to the MDP opening. Thus, differential attrition seems unlikely to explain the 2-digit results.

relative to incumbents that are geographically close but economically distant from the new plant (relative to the same comparison among incumbents in loser counties). A zero coefficient means that the estimated productivity spillover is the same for all the incumbents in a county, regardless of their economic proximity to the new plant.

Table 9 reports estimates of  $\pi_3$ . The first 6 columns include the interactions in one at a time. For example, column (1) suggests that a one standard deviation increase in the *CPS* Worker Transitions variable between incumbent plants' industry and the MDP's industry is associated with a 7% increase in the spillover. This finding is consistent with the theory that spillovers occur through the flow of workers across firms. One possibility is that new workers share ideas on how to organize production or information on new technologies that they learned with their previous employer. This measure tends to be especially high within 2-digit industries, so this finding was foreshadowed by the own 2-digit results in Table 8.

The measures of intellectual or technological linkages indicate meaningful increases in the spillover. The precise mechanism by which these ideas are shared is unclear, although both the flow of workers across firms and the mythical exchange of ideas over beers between workers from different firms are possibilities. Notably, there is more variation in these measures within 2-digit industries than in the *CPS* labor transitions measure.

Columns (5) and (6) provide little support for the flow of goods and services in determining the magnitude of spillovers. Thus, the data fail to support the types of stories where an auto manufacturer encourages (or even forces) its suppliers to adopt more efficient production techniques. Recall, all plants owned by the MDP's firm are dropped from the analysis, so this finding does not rule out this channel within firms. The finding on the importance of labor flows is consistent with the results in Ellison, Glaeser and Kerr (2007) and Dumais, Ellison, and Glaeser (2002), while the finding on input and output flows stands in contrast with these papers' findings.

In the column (7) specification, we include all the interactions simultaneously. The labor flow, the citation pattern, and the technology input interactions all remain positive but now would be judged to be statistically insignificant. The interactions with proximity to customers and suppliers are now both negative.

Overall, this analysis provides some support for the notion that spillovers occur between firms that share workers and between firms that use similar technologies. In terms of Section IC, this evidence is consistent with intellectual externalities, to the extent that they are embodied in workers who move from firm to firm, and to the extent that they occur among firms that use technologies that are reasonably similar. Table 9 seems less consistent with the hypothesis that agglomeration occurs because of proximity to customers and suppliers. We caution against definitive conclusions, because the utilized

measures are all imperfect proxies for the potential channels. Further, the possibility of better matches between workers and firms could not be directly tested with these data.

#### **D. Entry and Labor Costs as Indirect Tests of Spillovers**

The paper has uncovered economically sizable productivity gains for incumbent establishments following the opening of the new MDP. For example, a MDP plant opening is associated with a 12% increase in TFP five years later. This effect is even larger for incumbent plants that are in the same 2-digit industry of the new plant and for plants that tend to share workers and technologies. In the presence of positive spillovers, the model has two empirical predictions which this subsection tests.

First, if the spillovers are of a sufficient magnitude (i.e., they are larger than the increase in costs in the short run), the MDP's county should experience entry by new firms (relative to the losing counties). Table 10 tests this prediction. The entries in Panel 1 come from regressions that use data from the Census of Manufactures, which is conducted every five years. The dependent variables are the log of the number of establishments (column 1) and the log of total manufacturing output (column 2) in the county, respectively. In both columns, all plants owned by the MDP's firm are excluded from the dependent variable. The sample is comprised of observations from winning and losing counties only. The covariates include a full set of county fixed effects, year fixed effects, case fixed effects, and an indicator for whether the observation is from after the MDP's opening. The parameter of interest is associated with the interaction of indicators for an observation from a winning county and the post-opening indicator, so it is a difference in differences estimator of the impact of the MDP's opening.<sup>40</sup>

Column (1) reports that the number of manufacturing plants increased by roughly 12.5% in winning counties after the MDP plant's opening. A limitation of this measure is that it assumes that all plants are of an equal size. The total value of output is economically more meaningful, because it treats an increase in output at an existing plant and a new plant equally. As column (2) highlights, the opening of a MDP plant is associated with a roughly 14.5% increase in total output in the manufacturing sector although this is not estimated precisely.

Overall, these results are consistent with the TFP results of substantial spillovers in that it appears that the MDP attracted new economic activity to the winning counties (relative to losing ones) in the manufacturing sector.<sup>41</sup> Presumably, this new activity located in the winning counties to gain access to the spillovers.

---

<sup>40</sup> Because data is available every 5 years, depending on the Census year relative to the MDP opening, the sample years are 1 – 5 years before the MDP opening and 4 – 8 years after the MDP opening. Thus, each MDP opening is associated with one earlier date and one later date. Models are weighted by the number of plants in the county in years -6 to -10 and column 4 is weighted by the county's total manufacturing output in years -6 to -10.

<sup>41</sup> It is possible that the MDP's spillovers extended beyond manufacturing. In this case, it might be reasonable to expect increased entry in other sectors too.

The second prediction is that if the spillovers are positive, the prices of local inputs will increase as firms compete for these factors of production. The most important locally supplied input for manufacturing plants is labor. Column (3) in Panel 2 of Table 10 reports the results from regressions of the log wage using data from the 1970, 1980, 1990, and 2000 Censuses of Population from the winning and losing counties.<sup>42</sup> These data are preferable to the measure of labor costs reported in the Census of Manufacturers (i.e., the aggregate wage bill for production and non-production workers), which does not provide information on the quality of the labor force (e.g., education and experience). Specifically, we estimate models for  $\ln(\text{wage})$  and control for dummies for interactions of worker age and year, age-squared and year, education and year, sex and race and Hispanic and citizen, and case fixed effects. We also include indicators for whether the observation is from a winning county, occurs after the MDP's opening, and the interaction of these two indicators.<sup>43</sup> This interaction is the focus of the regression and is an adjusted difference in differences estimator of the impact of the MDP's opening on wages. This equation is analogous to the Model 1 version of equation (8) that was used to analyze TFP.

The estimate indicates that after adjusting for observable heterogeneity, wages increase by 2.7% in winning counties after the MDP's opening. This effect appears quantitatively sizable and is marginally statistically significant. The multiplication of the estimated 2.7% wage increase by the average labor earnings in winning counties implies that the quality-adjusted annual wage bill for employers in all industries increased by roughly \$151 million after the MDP's opening. This finding is consistent with positive spillovers and an upward sloping labor supply curve, perhaps due to imperfect mobility of labor (as in Section I).

It is possible to use the estimated increase in wages to make some back of the envelope calculations of the MDP's impact on incumbent plants' profits. Recall, the Model 1 result in Table 5 indicated an increase in TFP of approximately 4.8% (we focus on Model 1 because it is impossible to estimate a version of Model 2 with the decennial population Census data). If we assume that workers are homogenous or that high and low skill workers are perfectly substitutable in production, then the labor market-wide increase in wages applies throughout the manufacturing sector. In our sample, labor accounts for roughly 23% of total costs, so the estimated 2.7% increase in skill adjusted wages implies that manufacturers' costs increased by approximately 0.62%. The increased production costs due to

---

<sup>42</sup> The sample is limited to individuals who worked last year, worked more than 26 weeks, usually work more than 20 hours per week, are not in school, are at work, and who work for wages in the private sector. One important limitation of the Census data is that they lack exact county identifiers for counties with populations below 100,000. Instead, it is possible to identify PUMAs in the Census, which in rural areas can include several counties. This introduces significant measurement error, which is partly responsible for the imprecision of the estimate.

<sup>43</sup> The pre-period is defined as the most recent census before the MDP opening. The post-period is defined as the most recent census 3 or more years after the MDP opening. Thus, the sample years are 1 – 10 years before the MDP opening and 3 – 12 years after the MDP opening.

higher wages are therefore 13% of the gain in TFP.

These calculations demonstrate that the gains in TFP do not translate directly to profits due to the higher costs of local inputs. Since the prices and quality of other inputs are not observable, it is not possible to determine the total increase in production costs. Further, the observations on wages occurs only a few years after the MDP's opening and the impact on wages may increase more as production expands to gain access to the spillover. For these reasons, this back of the envelope calculation should be interpreted as a lower bound of the increase in input costs. In the long run, an equilibrium requires that the total impact on profits is zero.

## **VI. Summary and Interpretation**

This paper makes three principal contributions. This section summarizes them and places them in some context.

The first and most robust result is that the successful attraction of a "Million Dollar Plant" is on average associated with increased TFP for incumbent plants. This finding is robust to a battery of specification tests. In the preferred Model 2, the estimates suggest that incumbents' TFP in winning counties was about 12% higher. This translates into an additional \$430 million in annual output five years after the MDP's opening. This is an economically large number and it is natural to wonder whether it is "too large" to be plausible.

There are several related issues worth noting when considering this possibility. First, there may be an unobserved factor correlated with the MDP's opening that can explain these results and would invalidate the identifying assumption. The likelihood of this possibility is diminished by the similarity of the pre-trends in TFP documented in Figure 1 and the balancing of many of the ex-ante observable characteristics of winning and losing counties and their incumbent plants. Nevertheless, this possibility cannot be dismissed as would be the case in a randomized experiment.

Second, it is possible that the estimated impacts on TFP are influenced by changes in the quality of the workforce employed by incumbent plants. The sign of this change is *a priori* unclear. On the one hand, the MDP may attract higher quality workers to the county, allowing the incumbents to upgrade the quality of their workforce. On the other hand, the MDP could hire away the incumbents' best workers. Regardless of which force prevails, this could affect the estimated impact on TFP because labor is measured as total hours by production workers (recall that education or other measures of skill are not included in the *ASM*).

Third, the external validity of the results is unknown. In particular, the 47 MDP plants in the sample differ from the average manufacturing plant in many respects. Perhaps most importantly, these plants generated bidding from local governments, which is a first indication that there may be an ex-ante

expectation of significant spillovers. Further, these plants are substantially larger than the typical manufacturing plant. The point is that these results are unlikely to generalize to the opening of more typical plants.

Fourth, the estimated impact on TFP is unlikely to be a structural estimate of the MDP's impact on the TFP of incumbents. As Table 10 indicated, the MDP's opening appears to lead to other plants opening in the same county. Thus, the estimated spillover may reflect spillovers from the MDP and these new plants. Consequently, it is likely more appropriate to interpret the TFP effect as the reduced form effect of the MDP's opening and everything that accompanies it rather than the impact of the MDP alone.

The second contribution is a theoretical one that is supported by the data. In particular, the model underscored that increases in TFP do not translate directly into higher profits available for other manufacturing plants. The model predicts that the increases in TFP will be accompanied by increases in local input prices and that these increases are necessary for an equilibrium. The finding of higher prices for quality-adjusted labor in Table 10 is consistent with this prediction. Further, the increased levels of economic activity also documented in Table 10 reflect the increased demand to locate in the winning county that leads to higher local prices and the new equilibrium.

The third contribution is that the paper has shed some light on the channels that underlie the estimated spillovers. Specifically, our tentative conclusion is that the spillovers are larger between firms that share workers and use similar technologies. This is consistent with intellectual externalities, to the extent that they are embodied in workers who move from firm to firm and occur among firms that use technologies that are reasonably similar. Additionally, it is consistent with higher rates of TFP due to improved efficiencies of worker-firm matches.

Finally, these results may have some surprising policy implications. A standard critique of local governments providing subsidies to new plants to locate within their jurisdictions is that it may be rational for the locality but it is welfare decreasing for the nation. Although the economics of this argument are not always transparent, we believe that it generally refers to cases of disequilibrium where local factors of production (e.g., labor or buildings) are unemployed. In this case, tax competitions may be beneficial locally but suboptimal nationally (at least in the absence of significant worker moving costs).

In contrast, the finding of spillovers is the basis of an economic justification for local subsidies, even from a national perspective. Specifically, the MDP plant cannot capture these spillovers on its own and consequently might choose a location where its costs are low but the spillovers are minimal. However, the socially efficient outcome is for these plants to locate where the sum of their profits and the spillovers are greatest. In this setting, national welfare is maximized when payments are made to plants that produce the spillovers so that they internalize this externality in making their location decision.

In thinking about the policy implications, it is important to bear in mind that the estimated 12%

gain in TFP is an average effect. As Figure 2 demonstrated, there is substantial variability. For example, the estimated impact is negative in 40% of the cases. The point is that risk averse local governments may be unwilling to provide tax incentives with this distribution of outcomes.

## **VII. Conclusion**

Overall, this paper has documented that there are substantial spillovers flowing from large new plants to incumbent plants. Further, these spillovers are larger between plants that share labor pools and use similar technologies. Thus, this paper has provided evidence consistent with the idea that firms agglomerate in certain localities, at least in part, because they are more productive for being close to other firms.

There are several implications for future research. First, the paper has demonstrated the value of quasi-experiments that plausibly avoid the confounding of spillovers with differences in the determinants of TFP across locations. Second, the paper highlights that tests for the presence of spillovers can be conducted by directly measuring TFP. These tests can serve as an important complement to measurement of rates of coagglomeration that may reflect spillovers, cost shifters, or natural advantages. In this spirit, it is important to determine whether impacts on TFP are evident outside the manufacturing sector. Third, the heterogeneity in the estimated spillovers documented in Figure 2 and the results on the mechanism in Table 9 underscore that there is still much to learn about the structural source of these spillovers. This is an especially fruitful area for future research.

## References

- Acemoglu, Daron. 1996. "A Microfoundation for Social Increasing returns in Human Capital Accumulation." *Quarterly Journal of Economics*, 11 (August): 779-804.
- Audretsch, David and Maryann Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86 (June): 630-640.
- Audretsch, David and Maryann Feldman. 2004. "Knowledge Spillovers and the Geography of Innovation." In *Handbook of Urban and Regional Economics* 4, edited by J. Vernon Henderson and Jacques-François Thisse. Amsterdam: North-Holland.
- Becker, Randy, John Haltiwanger, Ron Jarmin, Shawn Klimek, and Daniel Wilson. 2005. "Micro and Macro Data Integration: The Case of Capital." *Bureau of the Census: Center for Economic Studies Working Paper 05-02*.
- Bloom, Nick, Mark Schankerman, and John Van Reenen. 2007. "Identifying Technology Spillovers and Product Market Rivalry." National Bureau of Economic Research Working Paper No. 13060.
- Branstetter, Lee G. 2001. "Are Knowledge Spillovers International or Intranational in Scope? Microeconomic evidence from US and Japan." *Journal of International Economics* 53 (February): 53-79.
- Card, David, Kevin F. Hallock, and Enrico Moretti. 2007. "The Geography of Giving: The Effect of Corporate Headquarters on Local Charities." University of California, Berkeley Mimeograph.
- Chiang, Hyowook. 2004. "Learning By Doing, Worker Turnover, and Productivity Dynamics." Econometric Society 2004 Far Eastern Meeting Paper No. 593.
- Davis, James and J. Vernon Henderson. 2004. "The Agglomeration of Headquarters." *Bureau of the Census: Center for Economic Studies Working Paper 04-02*.
- Davis, Steven J., John C. Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction*. Cambridge, MA: MIT Press.
- Davis, Donald and David E. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review* 92 (December).
- Dumais, Guy, Ellison, Glenn, and Glaeser, Edward L. 2002. "Geographic Concentration as a Dynamic Process." *Review of Economics and Statistics* 84 (May): 193-204.
- Duranton, Giles and Diego Puga. 2004 "Micro-Foundations of Urban Agglomeration Economies." In *Handbook of Urban and Regional Economics* 4, edited by J. Vernon Henderson and Jacques-François Thisse. Amsterdam: North-Holland.
- Ellison, Glenn and Edward L. Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* 105 (October): 889-927.
- Ellison, Glenn and Edward L. Glaeser. 1999. "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review* 89 (May): 311-316.
- Ellison, Glenn, Edward L. Glaeser, and William Kerr. 2007. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." National Bureau of Economic Research Working Paper No. 13068.
- Foster, Lucia, John C. Haltiwanger, and Chad Syverson. 2007. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" University of Chicago Mimeograph..
- Foster, Lucia, John C. Haltiwanger, and C.J. Krizan. 2000. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." Mimeograph.
- Glaeser, Edward L. 1999. "Learning in Cities." *Journal of Urban Economics* 46(September): 254-77.
- Glaeser, Edward L. 2001. "The Economics of Location-Based Tax Incentives." *Harvard Institute of Economic Research*, Discussion Paper Number 1932.
- Glaeser, Edward L. and Janet E. Kohlhase. 2003. "Cities, Regions and the Decline of Transport Costs." *Papers in Regional Science* 83 (January): 197-228.
- Glaeser, Edward L. and David C. Mare. 2001. "Cities and Skill." *Journal of Labor Economics* 19(2): 316-342.

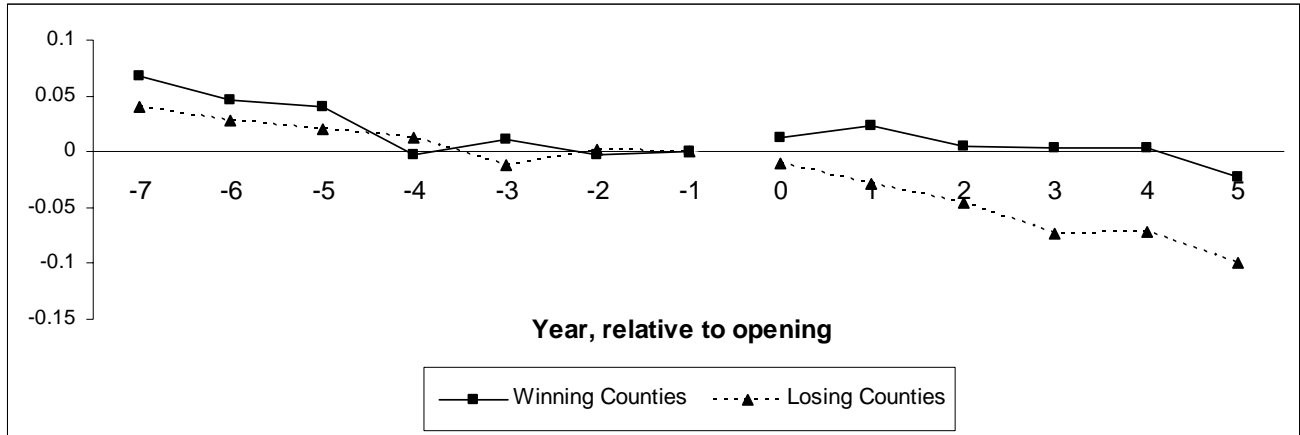


- Greenstone, Michael and Enrico Moretti. 2004. "Bidding for Industrial Plants: Does Winning a 'Million Dollar Plant' Increase Welfare?" National Bureau of Economic Research Working Paper No. 9844.
- Griliches, Zvi. 1958. "Research Cost and Social Returns: Hybrid Corn and Related Innovations." *Journal of Political Economy* 66 (October): 419-431.
- Griliches, Zvi and Jacques Mairesse. 1995. "Production Functions: the Search for Identification." National Bureau of Economic Research Working Paper No. 5067.
- Grossman, Gene M. and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press.
- Henderson, J. Vernon. 2001. "Urban Scale Economies." In *Handbook of Urban Studies*, edited by Ronan Paddison. London: SAGE Publications.
- Henderson, J. Vernon. 2003. "Marshall's Scale Economies." *Journal of Urban Economics* 55 (January): 1-28.
- Henderson, J. Vernon and Duncan Black. 1999. "A Theory of Urban Growth." *Journal of Political Economy*, 107 (April): 252-284
- Jaffe, Adam B. 1986. "Technological Opportunities and Spillovers of R&D: Evidence from Firm's Patents, Profit and Market Value." *American Economic Review* 76 (December): 984-1001.
- Jaffe, Adam B. and Manuel Trajtenberg. 2002. *Patents, Citations and Innovations: A Window on the Knowledge Economy*. Cambridge, MA: MIT Press.
- Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty. 2000. "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors." *American Economic Review* 90 (May): 215-218.
- Jaffe, Adam B., Manuel Trajtenberg and Rebecca M. Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citation." *Quarterly Journal of Economics* 108 (August): 577- 598.
- Jovanovic, Boyan and Rafael Rob. 1989. "The Growth and Diffusion of Knowledge." *Review of Economic Studies* 56 (October): 569-582.
- Krugman, Paul. 1991a. *Geography and Trade*. Cambridge, MA: MIT Press.
- Krugman, Paul. 1991b. "Increasing Returns and Economic Geography," *Journal of Political Economy* 99 (June): 483-499.
- Levinsohn, James and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies* 70 (April): 317-342.
- Lucas, Robert E. Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22 (July): 3-42.
- Marshall, Alfred. 1890. *Principles of Economics*. New York: Macmillan and Co.
- McGahan, Anita W. and Brian S. Silverman. 2001. "How Does Innovative Activity Change as Industries Mature?" *International Journal of Industrial Organization* 19 (July): 1141-1160.
- Moretti, Enrico. 2004a. "Estimating the external return to higher education: Evidence from cross-sectional and longitudinal data." *Journal of Econometrics* 120 (July-August): 175-212.
- Moretti, Enrico. 2004b. "Workers' Education, Spillovers and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review* 94 (June): 656-690.
- Moretti, Enrico. 2004c. "Human Capital Externalities in Cities." In *Handbook of Urban and Regional Economics* 4, edited by J. Vernon Henderson and Jacques-François Thisse. Amsterdam: North-Holland.
- Mowery, David C. and Arvids A. Ziedonis. 2001. "The Geographic Reach of Market and Non-market Channels of Technology Transfer: Comparing Citations and Licenses of University Patents." National Bureau of Economic Research Working Paper No. 8568.
- Olley, Steven G. and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64 (November): 1263-98.
- Ottaviano, Gianmarco and Jacques-François Thisse. 2004. "Agglomeration and Economic Geography." In *Handbook of Urban and Regional Economics* 4, edited by J. Vernon Henderson and Jacques-François Thisse. Amsterdam: North-Holland.
- Petrongolo, Barbara and Christopher A. Pissarides. 2005. "Scale Effects in Markets with Search." *The*

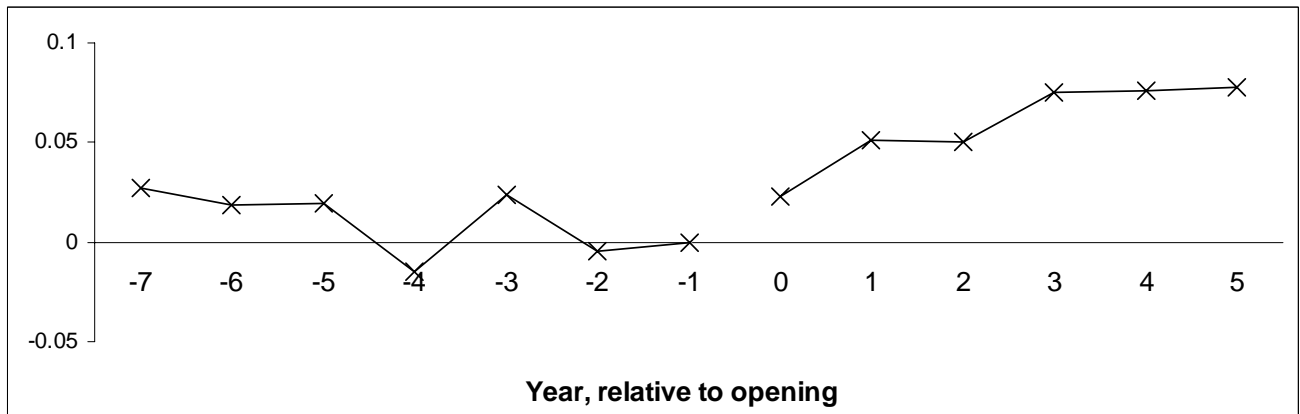
- Economic Journal*, 116 (January): 21-44.
- Rauch, James E. 1993. "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities." *Journal of Urban Economics* 34 (November): 380-400.
- Roback, Jennifer. 1982. "Wages, Rents and the Quality of Life." *Journal of Political Economy* 90 (December): 1257-1278.
- Roback, Jennifer. 1988. "Wages, rents and amenities: Differences among workers and regions." *Economic Inquiry* 26 (January): 23-41.
- Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (April): 41-55.
- Rosenthal, Stuart S. and William C. Strange. 2001. "The Determinants of Agglomeration." *Journal of Urban Economics* 50 (September): 191-229.
- Rosenthal, Stuart S. and William C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration." In *Handbook of Urban and Regional Economics* 4, edited by J. Vernon Henderson and Jacques-François Thisse. Amsterdam: North-Holland.
- Saxenian, Anna Lee. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.
- Syverson, Chad. 2004a. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy* 6 (December): 1181-1222.
- Syverson, Chad. 2004b. "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics* 2 (May): 534-550.
- van Biesebroeck, Johannes. 2004. "Robustness of Productivity Estimates." National Bureau of Economic Research Working Paper No. 10303.
- Venable, Tim. 1992. "BMW Drives into South Carolina with \$300 Million Auto Plant." *Site Selection* August: 630-631.
- Weiss, Leonard W. 1972. "The Geographic Size of Markets in Manufacturing." *The Review of Economics and Statistics* 54 (August): 245-257.
- Zucker, Lynne G., Michael R. Darby, and Marilyn B. Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review* 88 (March): 290-306.

Figure 1: The Effect of a “Million Dollar Plant” Opening on TFP of All Manufacturing Plants in Winner and Loser Counties.

All Industries: Winners Vs. Losers

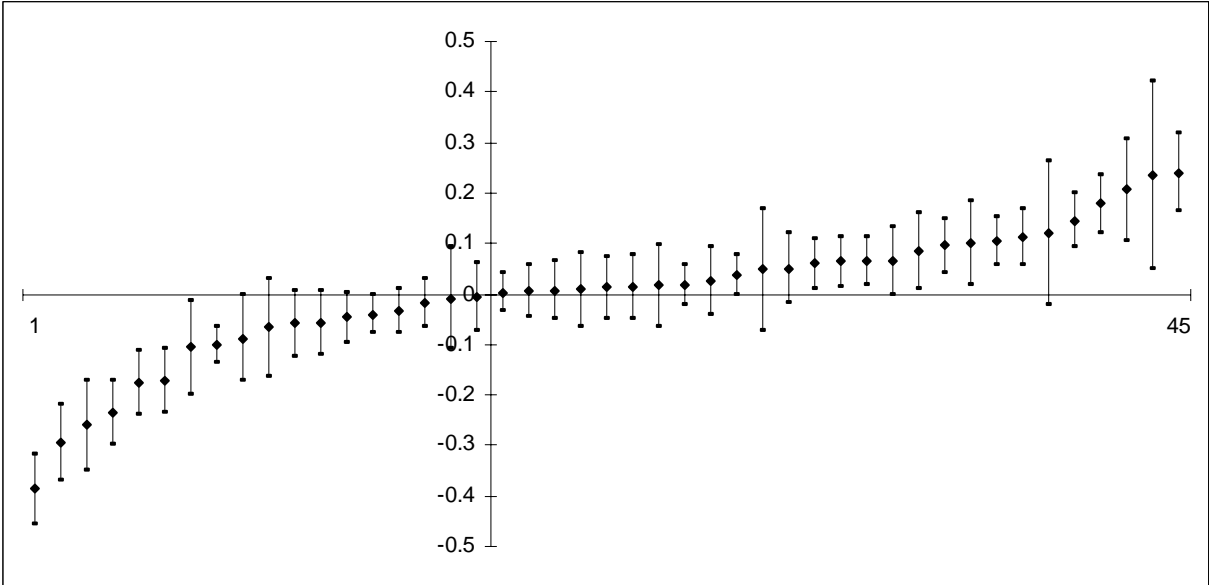


Difference: Winners – Losers



Notes: These figures accompany Table 4.

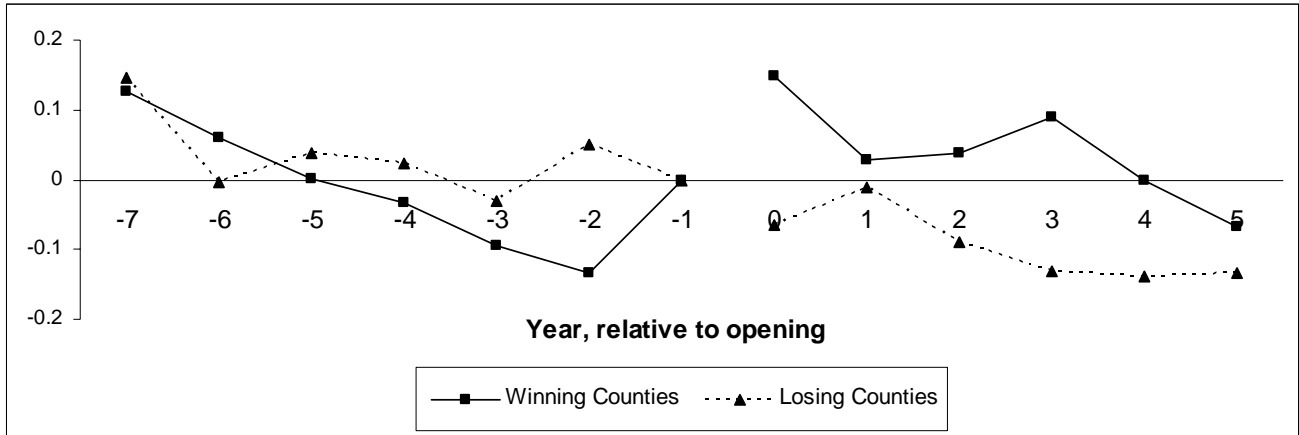
Figure 2. Distribution of Case-Specific Mean Shift Effects from the Opening of a “Million Dollar Plant”



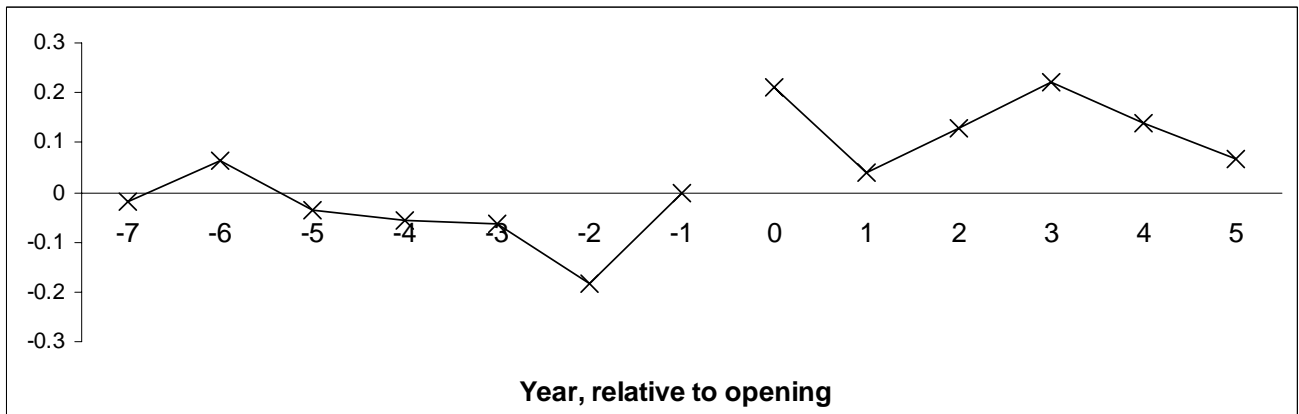
Notes: The figure reports results from a version of Model 1 that estimates the parameter  $\theta_1$  for each of the 47 MDP cases. The figure reports only 45 estimates because two cases were dropped for Census confidentiality reasons.

Figure 3: The Effect of a “Million Dollar Plant” Opening on TFP of Manufacturing Plants in the MDP’s 2-Digit Industry in Winner and Loser Counties.

2-digit MDP Industry: Winners Vs. Losers



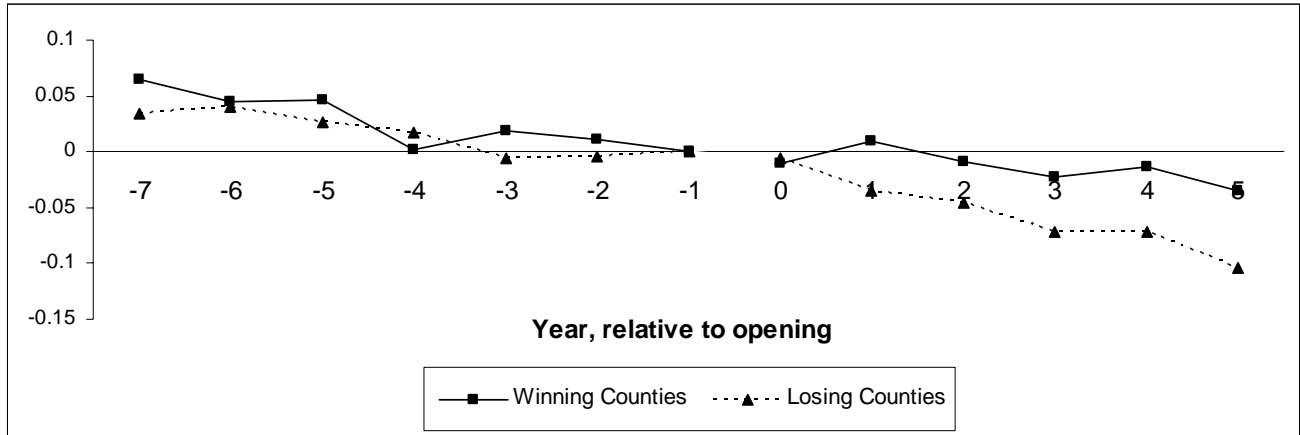
Difference (Winners – Losers)



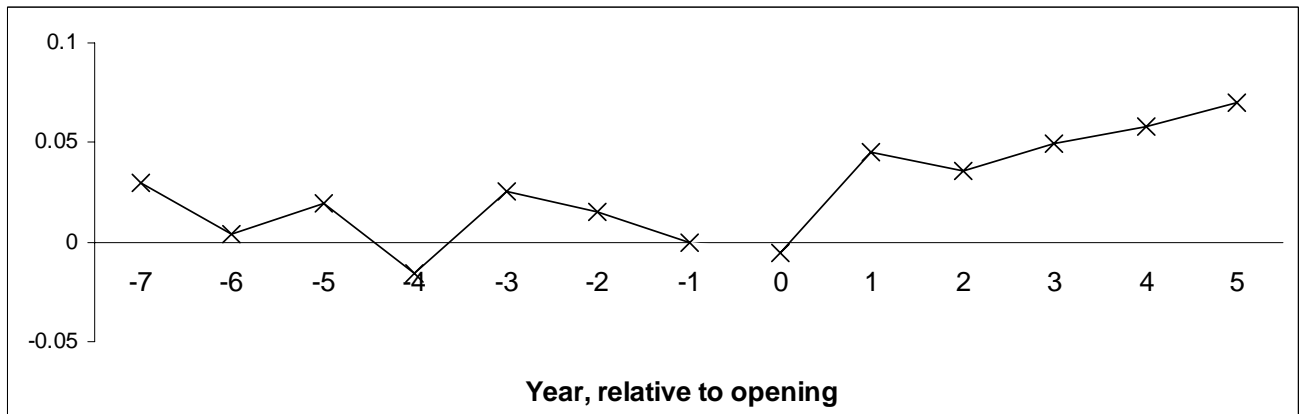
Notes: These figures accompany Table 8, Column 2 (MDP’s 2-digit Industry).

Figure 4: The Effect of a “Million Dollar Plant” Opening on TFP of Manufacturing Plants in All Industries, Except the MDP’s 2-Digit Industry in Winner and Loser Counties.

Other Industries: Winners Vs. Losers



Difference (Winners – Losers)



Notes: These figures accompany Table 8, Column 3 (All 2-digit Industries, except the MDP’s 2-digit Industry).

**Table 1. The “Million Dollar Plant” Sample**

	(1)
Sample MDP Openings <sup>1</sup> :	
Across All Industries	47
Within Same 2-digit SIC	16
Across All Industries:	
Number of Loser Counties per Winner County:	
1	31
2+	16
Reported Year – Matched Year: <sup>2</sup>	
-2 to -1	20
0	15
1 to 3	12
Reported Year of MDP Location:	
1981 – 1985	11
1986 – 1989	18
1990 – 1993	18
MDP Characteristics, 5 years after opening: <sup>3</sup>	
Output (\$1000)	452801
	(901690)
Output, relative to county output 1 year prior	0.086
	(0.109)
Hours of Labor (1000)	2986
	(6789)

<sup>1</sup> Million Dollar Plant openings that were matched to the Census data and for which there were incumbent plants in both winning and losing counties that are observed in each of the eight years prior to the opening date (the opening date is defined as the earliest of the magazine reported year and the year observed in the SSEL.) This sample is then restricted to include matches for which there were incumbent plants in the Million Dollar Plant’s 2-digit SIC in both locations.

<sup>2</sup> Only a few of these differences are 3. Census confidentiality rules prevent being more specific.

<sup>3</sup> Of the original 47 cases, these statistics represent 28 cases. A few very large outlier plants were dropped so that the mean would be more representative of the entire distribution (those dropped had output greater than half of their county’s previous output and sometimes much more). Of the remaining cases: most SSEL matches were found in the *ASM* or *CM* but not exactly 5 years after the opening date; a couple of SSEL matches in the 2xxx-3xxx SICs were never found in the *ASM* or *CM*; and a couple of SSEL matches not found were in the 4xxx SICs. The MDP characteristics are similar for cases identifying the effect within same 2-digit SIC. Standard deviations are reported in parentheses. All monetary amounts are in 2006 US dollars.

**Table 2. Summary Statistics for Measures of Industry Linkages**

Measure of Industry Linkage	Description	Mean			Standard Deviation
		All Plants	Only 1 <sup>st</sup> Quartile	Only 4 <sup>th</sup> Quartile	
Labor Market Pooling:					
<i>CPS</i> Worker Transitions	Proportion of workers leaving a job in this industry that move to the MDP industry (15 months later)	0.119	0.002	0.317	0.249
Intellectual or Technology Spillovers:					
Citation pattern	Percentage of manufactured industry patents that cite patents manufactured in MDP industry	0.022	0.001	0.057	0.033
Technology Input	R&D flows from MDP industry, as a percentage of all private sector technological expenditures	0.022	0.000	0.106	0.084
Technology Output	R&D flows to MDP industry, as a percentage of all original research expenditures	0.011	0.000	0.042	0.035
Proximity to Customers and Suppliers:					
Manufacturing Input	Industry inputs from MDP industry, as a percentage of its manufacturing inputs	0.017	0.000	0.075	0.061
Manufacturing Output	Industry output used by MDP industry, as a percentage of its output to manufacturers	0.042	0.000	0.163	0.139

Notes: *CPS* Worker Transitions was calculated from the frequency of worker industry movements in the rotating *CPS* survey groups. This variation is by Census Industry codes, matched to 2-digit SIC. The last 5 measures of cross-industry relationships were provided by Ellison, Glaeser, and Kerr (NBER Working Paper 13068). These measures are defined in a 3-digit SIC by 3-digit SIC matrix, though much of the variation is at the 2-digit level. In all cases, more positive values indicate a closer relationship between industries. Column 1 reports the mean value of the measure for all incumbent plants matched to their respective MDP. Column 2 reports the mean for the lowest 25% and column 3 reports the mean for the highest 25%. Column 4 reports the standard deviation across all observations. The sample of plants is all incumbent plants, as described for Table 1, for which each industry linkage measure is available for the incumbent plant and its associated MDP. These statistics are calculated when weighting by the incumbent plant's total value of shipments eight years prior to the MDP opening.



**Table 3. County & Plant Characteristics by Winner Status, One Year Prior to a “Million Dollar Plant” Opening**

	All Plants					Within Same Industry (2-digit SIC)				
	Winning Counties (1)	Losing Counties (2)	All US Counties (3)	(1) – (2) t-stat (4)	(1) – (3) t-stat (5)	Winning Counties (6)	Losing Counties (7)	All US Counties (8)	(6) – (7) t-stat (9)	(6) – (8) t-stat (10)
# of Counties	47	73				16	19			
County Characteristics:										
Total Per-capita Earnings (\$)	17,418	20,628	11,259	<b>-2.05</b>	<b>5.79</b>	20,230	20,528	11,378	-0.11	<b>4.62</b>
% Change, over last 6 years	0.074	0.096	0.037	-0.81	1.67	0.076	0.089	0.057	-0.28	0.57
Population	322,745	447,876	82,381	-1.61	<b>4.33</b>	357,955	504,342	83,430	-1.17	<b>3.26</b>
% Change, over last 6 years	0.102	0.051	0.036	<b>2.06</b>	<b>3.22</b>	0.070	0.032	0.031	1.18	1.63
Employment-Population ratio	0.535	0.579	0.461	-1.41	<b>3.49</b>	0.602	0.569	0.467	0.64	<b>3.63</b>
Change, over last 6 years	0.041	0.047	0.023	-0.68	<b>2.54</b>	0.045	0.038	0.028	0.39	1.57
Manufacturing Labor Share	0.314	0.251	0.252	<b>2.35</b>	<b>3.12</b>	0.296	0.227	0.251	1.60	1.17
Change, over last 6 years	-0.014	-0.031	-0.008	1.52	-0.64	-0.030	-0.040	-0.007	0.87	<b>-3.17</b>
# of Sample Plants	18.8	25.6	7.98	-1.35	<b>3.02</b>	2.75	3.92	2.38	-1.14	0.70
Plant Characteristics:										
Output (\$1000)	190,039	181,454	123,187	0.25	<b>2.14</b>	217,950	178,958	132,571	0.41	1.25
% Change, over last 6 years	0.082	0.082	0.118	0.01	-0.97	-0.061	0.177	0.182	-1.23	<b>-3.38</b>
Hours of labor (1000s)	1,508	1,168	877	1.52	<b>2.43</b>	1,738	1,198	1,050	0.92	1.33
% Change, over last 6 years	0.122	0.081	0.115	0.81	0.14	0.160	0.023	0.144	0.85	0.13

Notes: For each case to be weighted equally, counties are weighted by the inverse of their number per-case. Similarly, plants are weighted by the inverse of their number per-county multiplied by the inverse of the number of counties per-case. The sample includes all plants reporting data in the *ASM* for each year between the MDP opening and eight years prior. Excluded are all plants owned by the firm opening a MDP. Also excluded are all plants from two uncommon 2-digit SIC values so that subsequently estimated clustered variance matrices would always be positive definite. The sample of all United States counties excludes winning counties and counties with no manufacturing plant reporting data in the *ASM* for nine consecutive years. These other United States counties are given equal weight within years and are weighted across years to represent the years of MDP openings. Reported t-statistics are calculated from standard errors clustered at the county level. All monetary amounts are in 2006 US dollars.

**Table 4. Incumbent Plant Productivity, Relative to the Year of a MDP Opening**

Event Year	In Winning Counties (1)	In Losing Counties (2)	Difference (1) – (2) (3)
$\tau = -7$	0.067 (0.058)	0.040 (0.053)	0.027 (0.032)
$\tau = -6$	0.047 (0.044)	0.028 (0.046)	0.018 (0.023)
$\tau = -5$	0.041 (0.036)	0.021 (0.040)	0.020 (0.025)
$\tau = -4$	-0.003 (0.030)	0.012 (0.030)	-0.015 (0.024)
$\tau = -3$	0.011 (0.022)	-0.013 (0.022)	0.024 (0.021)
$\tau = -2$	-0.003 (0.027)	0.001 (0.011)	-0.005 (0.028)
$\tau = -1$	0	0	0
$\tau = 0$	0.013 (0.018)	-0.010 (0.011)	0.023 (0.019)
$\tau = 1$	0.023 (0.026)	-0.028 (0.024)	0.051* (0.023)
$\tau = 2$	0.004 (0.036)	-0.046 (0.046)	0.050 (0.033)
$\tau = 3$	0.003 (0.047)	-0.073 (0.057)	0.076+ (0.043)
$\tau = 4$	0.004 (0.053)	-0.072 (0.062)	0.076* (0.033)
$\tau = 5$	-0.023 (0.069)	-0.100 (0.067)	0.077* (0.035)
R-squared	0.9861		
Observations	28732		

Notes: Standard errors are clustered at the county level. Columns 1 and 2 report coefficients from the same regression: the natural log of output is regressed on the natural log of inputs (all worker hours, building capital, machinery capital, materials), year x 2-digit SIC fixed effects, plant fixed effects, case fixed effects, and the reported dummy variables for whether the plant is in a winning or losing county in each year relative to the MDP opening. When a plant is a winner or loser more than once, it receives a dummy variable for each incident. Plant-year observations are weighted by the plant's total value of shipments eight years prior to the MDP opening. Data on plants in all cases is only available 8 years prior to the MDP opening and 5 years after. Capital stocks were calculated using the permanent inventory method from early book values and subsequent investment. The sample of incumbent plants is the same as in columns 1 – 2 of Table 3. \*\* denotes significance at 1% level, \* denotes significance at 5% level, + denotes significance at 10% level.

**Table 5. The Effect of the Opening of a MDP Plant on the Productivity of Incumbent Plants**

	<u>All Counties</u>		<u>MDP Counties</u>		<u>All Counties</u>
	MDP Winners – MDP Losers		MDP Winners – MDP Losers		Random Winners
	(1)	(2)	(3)	(4)	(5)
Model 1:					
Mean Shift	0.0442+	0.0435+	0.0524*	0.0477*	-0.0824**
	(0.0233)	(0.0235)	(0.0225)	(0.0231)	(0.0177)
				[\$170m]	
R-squared	0.9811	0.9812	0.9812	0.9860	0.9828
Observations (plant x year)	418064	418064	50842	28732	426853
Model 2:					
Effect after 5 years	0.1301*	0.1324*	0.1355**	0.1203*	-0.0559+
	(0.0533)	(0.0529)	(0.0477)	(0.0517)	(0.0299)
				[\$429m]	
Level Change	0.0277	0.0251	0.0255	0.0290	-0.0197
	(0.0241)	(0.0221)	(0.0186)	(0.0210)	(0.0312)
Trend Break	0.0171+	0.0179*	0.0183*	0.0152+	-0.0060
	(0.0091)	(0.0088)	(0.0078)	(0.0079)	(0.0072)
Pre-trend	-0.0057	-0.0058	-0.0048	-0.0044	-0.0057**
	(0.0046)	(0.0046)	(0.0046)	(0.0044)	(0.0029)
R-squared	0.9811	0.9812	0.9813	0.9861	0.9828
Observations (plant x year)	418064	418064	50842	28732	426853
Plant & Ind-Year FEs	YES	YES	YES	YES	YES
Case FEs	NO	YES	YES	YES	N/A
Years Included	All	All	All	$-7 \leq \tau \leq 5$	All

Notes: The table reports results from the fitting of several versions of equation (8). Specifically, entries are from a regression of the natural log of output on the natural log of inputs, year x 2-digit SIC fixed effects, plant fixed effects, and case fixed effects. In Model 1, two additional dummy variables are included for whether the plant is in a winning county 7 to 1 years before the MDP opening or 0 to 5 years after. The reported mean shift indicates the difference in these two coefficients, i.e., the average change in TFP following the opening. In Model 2, the same two dummy variables are included along with pre- and post-trend variables. The shift in level and trend are reported, along with the pre-trend and the total effect evaluated after 5 years. In columns (1), (2), and (5), the sample is composed of all manufacturing plants in the *ASM* that report data for 14 consecutive years, excluding all plants owned by the MDP firm. In these models, additional control variables are included for the event years outside the range from  $\tau = -7$  through  $\tau = 5$  (i.e., -20 to -8 and 6 to 17). Column (2) adds the case fixed effects that equal 1 during the period that  $\tau$  ranges from -7 through 5. In columns (3) and (4), the sample is restricted to include only plants in counties that won or lost a MDP. This forces the industry-year fixed effects to be estimated solely from plants in these counties. Incumbent plants are now required to be in the data only when the MDP opens and all 8 years prior (not also for 14 consecutive years, though this does not change the results). For column (4), the sample is restricted further to include only plant-year observations within the period of interest (where  $\tau$  ranges from -7 to 5). This forces the industry-year fixed effects to be estimated solely on plant by year observations that identify the parameters of interest. In column (5), a set of 47 plant openings in the entire country were randomly chosen from the *ASM* in the same years and industries as the MDP openings. For all regressions, plant-year observations are weighted by the plant's total value of shipments eight years prior to the opening. Plants not in a winning or losing county are weighted by their total value of shipments in that year. All plants from two uncommon 2-digit SIC values were excluded so that estimated clustered variance-covariance matrices would always be positive definite. In brackets is the value in 2006 US\$ from the estimated increase in productivity: the percent increase is multiplied by the total value of output for the affected incumbent plants in the winning counties. Standard error clustered at the county level in parenthesis. \*\* denotes significance at 1% level, \* denotes significance at 5% level, + denotes significance at 10% level.

**Table 6 The Effect of the Opening of a MDP Plant on the Productivity of Incumbent Plants. Robustness to Different Specifications**

	Baseline specification (1)	Translog functional form (2)	Input - industry interactions (3)	Input- winner, input-post (4)	Region - Year FE (5)	Region - Year - Industry FE (6)	Fixed Input Cost Shares: plant level (7)	Fixed Input Cost Shares: SIC-3 level (8)
Model 1: Mean Shift	0.0477* (0.0231)	0.0471* (0.0226)	0.0406+ (0.0220)	0.0571* (0.0245)	0.0442+ (0.0230)	0.0369+ (0.0215)	0.0364 (0.0228)	0.0325 (0.0241)
Model 2: After 5 years	0.1203* (0.0517)	0.1053+ (0.0535)	0.0977* (0.0487)	0.1177* (0.0538)	0.1176* (0.0520)	0.0879* (0.0442)	0.0971 (0.0656)	0.0938+ (0.0538)

Notes: The table reports results from the fitting of several versions of equation (8). Column 1 reports estimates from the same specification as in column 4 of Table 5. Column 2 uses a translog functional form. Column 3 allows the effect of each input to differ at the 2-digit SIC level. Column 4 allows for the effect of inputs to differ in winning/losing counties and before/after the MDP opening. Column 5 controls for region (9 census divisions) by year fixed effects. Column 6 includes census division by year by industry fixed effects. Column 7 reports estimates when fixing the coefficient on each input to be its average cost share for each plant over the sample period. Per-period capital costs were calculated from capital rental rates using BLS data. Column 8 fixes the coefficient to be the average across all plants in each 3-digit SIC. Standard error clustered at the county level in parenthesis. \*\* denotes significance at 1% level, \* denotes significance at 5% level, + denotes significance at 10% level.

**Table 7. The Effect of the Opening of a MDP Plant on the Productivity of Incumbent Plants. Olley-Pakes (1996) and Levinsohn-Petrin (2003) Specifications.**

	Baseline specification	Investment	Investment, Capital	Investment- Capital Interactions	Materials	Materials- Capital Interactions	Investment, Materials, Capital	Materials- Capital and Investment- Capital Interactions
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model 1: Mean Shift	0.0477* (0.0231)	0.0460* (0.0230)	0.0431+ (0.0226)	0.0399+ (0.0213)	0.0451+ (0.0230)	0.0399+ (0.0216)	0.0410+ (0.0222)	0.0397* (0.0199)
Model 2: After 5 years	0.1203* (0.0517)	0.1149* (0.0529)	0.0971* (0.0490)	0.0989* (0.0482)	0.1153* (0.0526)	0.1004* (0.0487)	0.0919+ (0.0493)	0.1004* (0.0487)

Notes: The table reports results from the fitting of several versions of equation (8). Column (1) reports estimates from the same specification as in column 4 of Table 5. To this baseline specification, the column (2) specification adds 4<sup>th</sup> degree polynomial functions of log building investment and log machinery/equipment investment. The column (3) specification adds 4<sup>th</sup> degree polynomials in the two different capital stocks to the polynomials in investment in the column (2) equation. The column (4) specification adds all the “own” interactions between polynomials in current investment and capital (i.e., the building investment polynomial is interacted with the building capital stock polynomial, but is not also interacted with the machinery/equipment polynomials for stocks or investment). Column (5) adds a 4<sup>th</sup> degree function of log materials to the baseline specification. In the column (6) specification, a 4<sup>th</sup> degree polynomial in materials is fully interacted with 4<sup>th</sup> degree polynomials in building capital and machinery capital. Column (7) includes fourth-degree polynomials in log materials and in log investment, and log capital stock for both types of capital (not interacted). Finally, column (8) includes the controls from the columns (4) and (6) specifications. Standard error clustered at the county level in parenthesis. \*\* denotes significance at 1% level, \* denotes significance at 5% level, + denotes significance at 10% level.

**Table 8 The Effect of the Opening of a MDP Plant on the Productivity of Incumbent Plants, For Incumbent Plants in the MDP's 2-digit Industry and All Other Industries**

	All Industries	MDP's 2-digit Industry	All Other 2-Digit Industries
	(1)	(2)	(3)
Model 1:			
Mean Shift	0.0477* (0.0231) [\$170m]	0.1700* (0.0743) [\$102m]	0.0326 (0.0253) [\$104m]
R-squared	0.9860		0.9861
Observations	28732		28732
Model 2:			
Effect after 5 years	0.1203* (0.0517) [\$429m]	0.3289 (0.2684) [\$197m]	0.0889+ (0.0504) [\$283m]
Level Change	0.0290 (0.0210)	0.2814** (0.0895)	0.0004 (0.0171)
Trend Break	0.0152+ (0.0079)	0.0079 (0.0344)	0.0147+ (0.0081)
Pre-trend	-0.0044 (0.0044)	-0.0174 (0.0265)	-0.0026 (0.0036)
R-squared	0.9861		0.9862
Observations	28732		28732

Notes: The table reports results from the fitting of several versions of equation (8). For comparison, Column 1 reports the same results as column 4 of Table 5. Columns 2 and 3 report estimates from the same regression, which fully interacts the winner/loser and pre/post variables with indicators for whether the incumbent plant is in the same 2-digit industry as the MDP or a different industry. Standard error clustered at the county level in parenthesis. The numbers in brackets are the value (2006 US\$) from the estimated increase in productivity: the percent increase is multiplied by the total value of output for the affected incumbent plants in the winning counties. \*\* denotes significance at 1% level, \* denotes significance at 5% level, + denotes significance at 10% level.

**Table 9. How the Effect of the Opening of a MDP Plant on the Productivity of Incumbent Plants Varies with Measures of Economic Distance**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CPS Worker Transitions	0.0701** (0.0237)						0.0374 (0.0260)
Citation pattern		0.0545** (0.0192)					0.0256 (0.0208)
Technology Input			0.0320+ (0.0173)				0.0501 (0.0421)
Technology Output				0.0596** (0.0216)			0.0004 (0.0434)
Manufacturing Input					0.0060 (0.0123)		-0.0473 (0.0289)
Manufacturing Output						0.0150 (0.0196)	-0.0145 (0.0230)
R-squared	0.9852	0.9852	0.9851	0.9852	0.9851	0.9852	0.9853
Observations	23397	23397	23397	23397	23397	23397	23397

Notes: The table reports results from the fitting of several versions of equation (8). Building on the Model 1 specification in column 4 of Table 5, each column adds the reported interaction terms between winner/loser and pre/post status with the indicated measures of how an incumbent plant's industry is linked to its associated MDP's industry. For assigning this linkage measure, the incumbent plant's industry is held fixed at its industry the year prior to the MDP opening. Whenever a plant is a winner or loser more than once, it receives an additive dummy variable and interaction term for each occurrence. These industry linkage measures are defined and described in Table 2 and here the measures are normalized to have a mean of zero and a standard deviation of one. The sample of plants is that in column 4 of Table 5, but it is restricted to plants that have industry linkage data for each measure. Standard errors clustered at the county level are reported in parentheses. \*\* denotes significance at 1% level, \* denotes significance at 5% level, + denotes significance at 10% level.

**Table 10. The Effect of the Opening of a MDP Plant on Wages and Number of Plants in the County**

	Panel 1		Panel 2
	(Census of Manufacturers)		(Census of Population)
	Dep. Var.: Log(Plants)	Dep. Var.: Log(Total Output)	Dep. Var.: log(Wage)
	(1)	(2)	(3)
Difference-in-difference	0.1255* (0.0550)	0.1454 (0.0900)	0.0268+ (0.0139)
R-squared	0.9984	0.9931	0.3623
Observations	209	209	1057999

Notes: The table reports results from the fitting of three regressions. In Panel 1, the dependent variables are the log of number of establishments and the log of total manufacturing output in the county, respectively. Controls include case effects, county, and year fixed effects. Entries are the county-level diff in diff estimates for winning a MDP, based on data from the Census of Manufacturers. Because data is available every 5 years, depending on the Census year relative to the MDP opening, the sample years are 1 – 5 years before the MDP opening and 4 – 8 years after the MDP opening. Thus, each MDP opening is associated with one earlier date and one later date. The column (1) model is weighted by the number of plants in the county in years -6 to -10 and the column (2) model is weighted by the county's total manufacturing output in years -6 to -10. In Panel 2, the dependent variable is log wage and controls include dummies for age\*year, age-squared\*year, education\*year, sex\*race\*Hispanic\*citizen, and case fixed effects. The entry is the county-level diff in diff estimate for winning a MDP. The pre-period is defined as the most recent census before the MDP opening. The post-period is defined as the most recent census 3 or more years after the MDP opening. Thus, the sample years are 1 – 10 years before the MDP opening and 3 – 12 years after the MDP opening. The sample is limited to individuals who worked last year, worked more than 26 weeks, usually work more than 20 hours per week, are not in school, are at work, and who work for wages in the private sector. The number of observations reported refers to unique individuals – some IPUMS county groups include more than one FIPS, so all individuals in a county group were matched to each potential FIPS. The same individual may then appear in more than one FIPS and observations are weighted to give each unique individual the same weight (i.e., an individual appearing twice receives a weight of ½).

Standard errors clustered at the county level are reported in parentheses. \*\* denotes significance at 1% level, \* denotes significance at 5% level, + denotes significance at 10% level.



**Appendix Table 1. The Effect of the Opening of a MDP Plant on Inputs and Output (Unadjusted for Inputs) of Incumbent Plants**

	Output (1)	Worker Hours (2)	Machinery Capital (3)	Building Capital (4)	Materials (5)	Capital / Labor (6)	Energy / Capital (7)
Model 1: Mean Shift	0.1200** (0.0354)	0.0789* (0.0357)	0.0401 (0.0348)	0.1327+ (0.0691)	0.0911** (0.0302)	-0.0251 (0.0418)	0.0008 (0.0372)
Model 2: After 5 years	0.0826+ (0.0478)	0.0562 (0.0469)	-0.0089 (0.0300)	-0.0077 (0.0375)	0.0509 (0.0541)	-0.1310 (0.1228)	0.0585 (0.0493)

Notes: Standard errors clustered at the county level are reported in parentheses. All outcome variables are run in logs. In columns (6) and (7), “Capital” is the combined stock of building and machinery capital. In column (7), “energy” is the combined cost of electricity and fuels: 163 plant-year observations are excluded that have missing or zero values for energy. See the text for more details. \*\* denotes significance at 1% level, \* denotes significance at 5% level, + denotes significance at 10% level.