Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings

Michael Greenstone

Massachusetts Institute of Technology

Richard Hornbeck

Harvard University

Enrico Moretti

University of California, Berkeley

We quantify agglomeration spillovers by comparing changes in total factor productivity (TFP) among incumbent plants in "winning" counties that attracted a large manufacturing plant and "losing" counties that were the new plant's runner-up choice. Winning and losing counties have similar trends in TFP prior to the new plant opening. Five years after the opening, incumbent plants' TFP is 12 percent higher in winning counties. This productivity spillover is larger for plants sharing similar labor and technology pools with the new plant. Consistent with spatial equilibrium models, labor costs increase in winning

counties, indicating that profits ultimately increase less than productivity.

I.   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 potential sources of agglomeration advantages include cheaper and faster supply of intermediate goods and services, proximity to workers or consumers, better quality of worker-firm matches in thicker labor markets, lower risk of unemployment for workers and lower risk of unfilled vacancies for firms following idiosyncratic shocks, and knowledge spillovers.[1]

The possibility of documenting productivity advantages through agglomeration 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 such as New York, San Francisco, or London that are 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 an 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 plants generates agglomeration externalities. In the absence of positive externalities, it is difficult to justify the use of taxpayer money for subsidies based on economic efficiency grounds. The optimal

---

[1] See Duranton and Puga (2004), Glaeser and Gottlieb (2009), and Moretti (forthcoming) for recent surveys.

magnitude of incentives depends on the magnitude of agglomeration spillovers, if they exist.[2]

The existence and exact magnitude of agglomeration spillovers are considered open questions by many, despite their enormous theoretical and practical relevance.[3] This paper has three objectives. First, we test for and quantify agglomeration spillovers in manufacturing by estimating how the productivity of incumbent plants changes when a large plant opens in their county. We estimate augmented Cobb-Douglas production functions that allow the TFP of incumbent plants to depend on the presence of the new plant, using plant-level data from the Annual Survey of Manufactures. Second, we shed light on the possible mechanisms by investigating whether the magnitude of the spillovers depends on economic linkages between the incumbent plant and the new plant. We consider different measures of linkages, including input and output flows, measures of labor flows between firms, and technological linkages. Third, we measure the extent to which the productivity gains generated by the spillover are reflected in higher local factor prices.

Because the new plant's 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 and that are difficult to measure, such as local transportation infrastructure, current and future costs of factors of production, quality of the workforce, presence of intermediate input suppliers, and any other local cost shifter.

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 the absence of the plant opening. These rankings come from the corporate real estate journal

[2] We discuss in more detail the policy implications of local subsidies in Greenstone and Moretti (2004). See also Glaeser (2001), Card, Hallock, and Moretti (2007), and Glaeser and Gottlieb (2008).

[3] To date, there are two primary approaches in testing for spillovers. The first tests for an unequal geographic distribution of firms. These "dartboard" style tests reveal that firms are spread unevenly and that co-agglomeration rates are higher between industries that are economically similar (Ellison, Glaeser, and Kerr, forthcoming). 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 (see, e.g., Henderson 2003). 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 advantages, or other cost shifters. A causal estimate of the magnitude of spillovers requires a solution to this problem of identification.

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 two or three 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 counties form a valid counterfactual for the incumbents in the winning counties, after conditioning on differences in preexisting 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 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 the 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. This finding 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 percent relative increase in incumbent plants' TFP. 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), with inputs held constant. A 12 percent increase in TFP is equivalent to moving a county from the 10th percentile of the county-level TFP distribution to the 27th percentile; alternatively, it is equivalent to a 0.6-standard-deviation increase in the distribution of county TFP. We interpret this finding as evidence of large productivity spillovers generated by increased agglomeration.

Notably, the estimated productivity gains experienced by incumbent plants in winning counties are highly heterogeneous. The average county-level TFP increase is very large in some instances, small in some other cases, and even negative for a nonnegligible number of counties.

Having found evidence in favor of the existence of agglomeration spillovers, we then turn to the question of what might explain these spillovers. We follow Moretti (2004*c*) and Ellison et al. (forthcoming)

and investigate how the magnitude of the spillovers depends on measures of economic proximity 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.[4]

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 transition 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 support for the notion that spillovers occur between firms that share workers and use similar technologies.

To 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 to entry of firms that are interested in gaining access to the spillover. The original plant opening and subsequent new entry lead to competition for inputs, so incumbent firms face higher prices for labor, land, and other local inputs. In the model, firms produce nationally traded goods and cannot raise output prices in response to 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 higher input prices.

Consistent with these predictions, we find increases in quality-adjusted labor costs following MDP openings. These higher wages are consistent with the documented increase in economic activity in the winning counties and with a local labor supply curve that is upward sloping (at least in the medium run). We also find positive net entry in winning counties, which the model predicts will occur if there are sufficiently large positive spillovers to generate an overall increase in profitability.

The findings in this paper are related to two earlier studies that use a similar approach to identify productivity spillovers at the local level: Henderson (2003) documents agglomeration spillovers for the machinery and high-tech industries, and Moretti (2004c) estimates productivity

---

[4] We are deeply indebted to Glenn Ellison, Edward Glaeser, and William Kerr for providing their data for five of these measures of economic distance.

spillovers generated by increased concentration of human capital in a location. Consistent with the findings in this paper, the findings in Moretti's study point to the existence of productivity spillovers that are economically nontrivial, vary significantly depending on economic distance, and are largely offset by increased labor costs.

Our findings have two sets of important implications. First, our findings have implications for local economic development policies. The magnitude and form of agglomeration spillovers are crucial to understanding the economic rationale for location-based policies and their welfare consequences. In a world with significant agglomeration spillovers, government intervention may be efficient from the point of view of a locality, although not always from the point of view of aggregate welfare. We discuss how our results inform the debate on local economic development policies.

Second, our findings have implications for understanding industrial clusters. Urban economists have long noted that economic activity is spatially concentrated by industry. This industrial concentration appears to be a pervasive feature of the geographical distribution of economic activity in most counties, and it appears to be fairly stable over time. Because the increase in labor costs that we find is countywide whereas the productivity spillovers decline in economic distance, incumbent firms that are economically further away may become less profitable. In the long run, this process may result in increased agglomeration of similar plants in each MDP location. The interaction between spillovers and input costs may therefore help explain the existence and persistence of industrial clusters.

The remainder of the paper is organized as follows. Section II presents a simple model. Section III discusses the identification strategy. Section IV introduces the data sources. Section V presents the econometric model. Sections VI and VII describe the empirical results. Section VIII interprets the results and discusses implications for policy. Section IX concludes.

II.   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 briefly reviewing theories of agglomeration.[5] We then present a simple theoretical framework that guides the subsequent empirical exercise and aids in interpreting the results.

[5] See Glaeser and Gottlieb (2009) and Moretti (forthcoming) for comprehensive surveys of this literature.

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, if there are search frictions and jobs and workers are heterogeneous, then 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.[6] Second, large labor markets may provide insurance against idiosyncratic shocks, either on the firm side or on the worker side (Krugman 1991*a*). If firms experience idiosyncratic and unpredictable demand shocks that lead to layoffs and moving/hiring is costly for workers/firms, then thicker labor markets will reduce the probability that a worker is unemployed and a firm has unfilled vacancies.[7]

These two hypotheses have different implications for the relationship between the 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.[8]

However, if the only effect of thickness in the 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. In contrast to the case of improved matching described above, the production function does not change: for the same

[6] For a related point in a different context, see Petrongolo and Pissarides (2005).

[7] 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.

[8] 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 TFP will be higher regardless.

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. 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.[9] This change in relative factor prices will change the relative use of labor and capital.

2. A second reason why the concentration of economic activity may be beneficial has to do with transportation costs (Krugman 1991*a*, 1991*b*; 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 are likely 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 the 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 (see, e.g., Marshall 1890; Lucas 1988; Jovanovic and Rob 1989; Grossman and Helpman 1991; Saxenian 1994; Glaeser 1999; Moretti 2004*a*, 2004*b*, 2004*c*). 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, Trajtenberg, and Henderson 1993). Saxenian (1994) argues that geographic proximity of high-tech firms in Silicon Valley is associated with a more ef-

[9] 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 because of lower risk of unfilled vacancies (generated by an increase in labor demand in denser areas).

ficient flow of new ideas and ultimately causes faster innovation.[10] 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, this form of agglomeration would lead to 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. However, if density of economic activity results only 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 the 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 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, there should not be differences in productivity between dense areas and less dense areas, although there would be differences in wages that reflect the compensating differential.

5. Finally, spatial concentration of some industries may be explained by the presence of *natural advantages* or *productive amenities.* 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 of suitable weather and land. For some manufacturing productions, the presence of a harbor may be important. Natural advantages imply that firms located in areas with a high concentration of similar firms are more productive, but of course this correlation is unrelated to 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.

---

[10] 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, Darby, and Brewer 1998). Moretti (2004c) finds stronger human capital spillovers between pairs of firms in the same city that are economically or technologically closer.

B.   A Simple Model

We begin by considering the case in which incumbent firms are homogeneous 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.

1.   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 one. Incumbent firms choose their amount of labor, L, capital, K, and land, T, to maximize the following expression:

$$\max_{L,K,T} f(A,\ L,\ K,\ T) - wL - rK - qT,$$

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:

$$A = A(N), \tag{1}$$

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 in which A increases in N: $\partial A/\partial N > 0$. If instead $\partial A/\partial N = 0$, 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.

As in Moretti (forthcoming), we attribute the upward-sloping labor supply curve to the existence of preferences for location. We assume that workers' indirect utility depends on wages, cost of housing, and idiosyncratic preferences for location and that in equilibrium marginal

workers are indifferent across locations. 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 in the original county. When a new plant opens in county c, wages there start rising, and some workers find it optimal to move to county c. The number of workers who move, and therefore the slope of the labor supply function, depend on the importance of preferences for location (see Moretti [forthcoming] for details). 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 is 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, Π^*, as

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

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:

$$\frac{d\Pi^*}{dN} = \left(\frac{\partial f}{\partial A} \times \frac{\partial A}{\partial N}\right) \tag{2}$$
$$+ \frac{\partial w}{\partial N}\left\{\left[\frac{\partial L^*}{\partial w}\left(\frac{\partial f}{\partial L} - w\right) - L^*\right] + \left[\frac{\partial K^*}{\partial w}\left(\frac{\partial f}{\partial K} - r\right)\right] + \left[\frac{\partial T^*}{\partial w}\left(\frac{\partial f}{\partial T} - q\right)\right]\right\}$$
$$+ \frac{\partial q}{\partial N}\left\{\left[\frac{\partial L^*}{\partial q}\left(\frac{\partial f}{\partial L} - w\right)\right] + \left[\frac{\partial K^*}{\partial q}\left(\frac{\partial f}{\partial K} - r\right)\right] + \left[\frac{\partial T^*}{\partial q}\left(\frac{\partial f}{\partial T} - q\right) - T^*\right]\right\}.$$

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

$$\frac{d\Pi^*}{dN} = \left(\frac{\partial f}{\partial A} \times \frac{\partial A}{\partial N}\right) - \left(\frac{\partial w}{\partial N}L^* + \frac{\partial q}{\partial N}T^*\right). \tag{3}$$

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, $(\partial f/\partial A \times \partial A/\partial N)$. This effect is unambiguously positive because it allows an incumbent firm to produce more output using the same amount of inputs. Formally, $\partial f/\partial A > 0$ by assumption, and if there are positive spillovers, $\partial A/\partial N > 0$.

The second term, $-[(\partial w/\partial N)L^* + (\partial q/\partial N)T^*]$, 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 $\partial w/\partial N > 0$ and $\partial q/\partial N > 0$, whereas 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. This point is illustrated in a similar context in Moretti (2004c).

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 utilization, it mechanically raises production costs. Second, it leads the firm to reoptimize 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: $\partial K^*/\partial N \geq 0$.

By contrast, the effect on the use of labor and land is ambiguous. On one hand, the productivity of all factors increases. On the other hand, the price of labor and land increases. 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 an MDP opening that causes positive spillovers. We initially consider the case in which 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 entry and could cause some existing firms to exit.

The alternative case is that $d\Pi^*/dN > 0$, which occurs when the magnitude of the spillover due to the MDP opening exceeds the increase

in factor prices due to the MDP's demand for local inputs. In the short run, profits will be positive for new entrants. These positive profits will disappear over time as the price of local factors, such as 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 as a result of a less than infinite elasticity of local labor supply, as noted above. These adjustments make marginal 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 not possible in our empirical context 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.[11]

---

[11] This model focuses on the case in which the productivity benefits of the agglomeration spillovers are distributed equally across all factors. What happens when agglomeration spillovers are factor biased? Assume, e.g., that agglomeration spillovers raise the productivity of labor but not the productivity of capital. As 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 θ is a productivity shifter. We define factor-biased agglomeration spillover as the case in which the productivity shifter θ depends positively on the density of the economic activity in the county: $\theta = \theta(N)$ and $\partial\theta/\partial N > 0$. If $\partial A/\partial N = 0$ 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

$$\frac{d\Pi^*}{dN} = \left(\frac{\partial f}{\partial H} \times \frac{\partial \theta}{\partial N}\right) H^* - \left(\frac{\partial w}{\partial N} H^* + \frac{\partial q}{\partial N} T^*\right).$$

The effect on profits can be decomposed into two parts. The first term represents the increased productivity of labor. It is the product of the sensitivity of output to labor ($\partial f/\partial H > 0$) times the magnitude of the agglomeration spillover ($\partial\theta/\theta N > 0$ by definition) times the number of workers. The second term is the same as in eq. (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: $\partial L^*/\partial N \le 0$. Land is equally productive, but its price increases, so its equilibrium use declines: $\partial T^*/\partial N \le 0$. 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.

2.   Heterogeneous Incumbents

What happens if the population of incumbent firms is nonhomogeneous? Consider the case in which 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:

$$\frac{d\Pi_H^*}{dN_H} = \left(\frac{\partial f_H}{\partial A_H} \times \frac{\partial A_H}{\partial N_H}\right) - \left(\frac{\partial w_H}{\partial N_H} L_H^* + \frac{\partial q}{\partial N_H} T^*\right) \tag{4}$$

and

$$\frac{d\Pi_L^*}{dN_H} = \left(\frac{\partial f_L}{\partial A_L} \times \frac{\partial A_L}{\partial N_H}\right) - \left(\frac{\partial w_L}{\partial N_H} L_L^* + \frac{\partial q}{\partial N_H} T^*\right). \tag{5}$$

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:

$$\frac{\partial f_H}{\partial A_H} \times \frac{\partial A_H}{\partial N_H} > \frac{\partial f_L}{\partial A_L} \times \frac{\partial A_L}{\partial N_H}. \tag{5'}$$

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:

$$\frac{\partial w_H}{\partial N_H} > \frac{\partial w_L}{\partial N_H}. \tag{5''}$$

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

   This model of heterogeneous incumbents has two main implications. First, it may be reasonable to expect larger spillovers on firms that are economically "closer" 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 both larger 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); and
4. the price of locally supplied factors of production will increase.

We test for changes in the price of quality-adjusted labor, which is arguably the most important local factor for manufacturing plants.

III.   Plant Location Decisions and Research Design

In testing the four empirical predictions outlined above, the main econometric challenge is that firms do not choose their location randomly. Firms maximize profits and choose to locate where their expectation of the present discounted value 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, and so forth. These factors are frequently unobserved, and, problematically, they are likely to be correlated with the TFP of existing plants.

Therefore, a naive 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 Bavarian Motor Works (BMW) picked the location for one of its plants.[12] 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 it had narrowed the list of potential candidates to 20 U.S. counties. Six months later, BMW announced that the two finalists in the competition were Greenville-Spartanburg, South Carolina, and Omaha, Nebraska. In

---

[12] This plant is in Greenstone and Moretti's (2004) set of 82 MDP plants. Owing to Census confidentiality restrictions, we cannot report whether this plant is part of this paper's analysis.

1992, BMW announced that it would site the plant in Greenville-Spartanburg and that it 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; numerous global firms in the area, including 58 German companies; a 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 U.S. 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 encourages 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's presence. According to local officials, the facility's ex ante expected 5-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. Thus, the subsidy is likely to be a function of the expected gains from agglomeration for the county.[13]

This possibility is relevant for this paper's identification strategy because the magnitude of the spillover from a particular plant depends

[13] The fact that business organizations such as chambers of commerce support these incentive plans (which 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 welfare-decreasing for the county.

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 because of 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 in which 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 "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").[14] As the BMW case study indicates, the winner and losers are usually chosen from an initial sample of "semifinalist" sites that in many cases number more than 100. The losers are counties that have survived a long selection process but narrowly lost the competition.

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 pairs of winning and losing counties belonging to the same 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 fails to hold, we presume that this pairwise approach is more reliable than using re-

---

[14] 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 four cases, among the original 82, we were able to identify the losing counties. Comprehensive data on the subsidy offered by winning and losing counties are unavailable in the *Site Selection* articles.

gression adjustment to compare the TFP of incumbent plants in counties with new plants to the other 3,000 U.S. counties or to using a matching procedure based on observable variables.

IV. 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 chooses to locate (the winning county) and one or two runner-up counties (the losing counties). The articles tend to focus on large manufacturing plants that are the target of local government subsidies. An important limitation of these articles is that the magnitude of subsidy offered by winning counties is often unobserved and the subsidy offered by losing counties is almost always unobserved. In addition, when there is more than one losing county, there is no indication of the plants' relative preferences among the losing counties.

We identified the MDPs in the Standard Statistical Establishment List (SSEL), which is the Census Bureau's "most complete, current, and consistent data for U.S. business establishment,"[15] and matched the plants to the Annual Survey of Manufactures (ASM) and the Census of Manufactures (CM) from 1973–98.[16] Of the 82 MDP openings in all industries used in Greenstone and Moretti (2004), we identified 47 usable MDP openings in the manufacturing data. In order to qualify as a usable MDP manufacturing opening, we imposed the following criteria: (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 present for each of the previous 8 years. Among the 35 MDP openings that did not qualify, we identified 10 openings in the retail and wholesale trade sectors whose effects we examine in robustness specifications.

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 stocks, materials, total value of ship-

[15] 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.

[16] 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.

ments, and firm identifiers. The four-digit Standard Industrial Classification (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. Our main analysis uses a sample of 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 all plants owned by firms that own an MDP. 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 preexisting 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 percent of countywide manufacturing shipments in the last CM before the MDP opening.

In addition to 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 on the basis of some measure of economic distance. We focus on 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 three-digit industry and the fraction of each industry's outputs sold to manufacturers that are purchased by each three-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 two-digit industry that move to firms in each two-digit industry. Third, to measure technological proximity, we use data on the fraction of patents manufactured in a three-digit industry that cite patents manufactured in each three-digit industry. We also use data on the amount of R&D expenditure in a three-digit industry that is used in other three-digit industries.[17]

___

[17] 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 plant but at other times are written when the location decision is made or when 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

TABLE 1
THE MILLION DOLLAR PLANT SAMPLE

	(1)
Sample MDP openings:[a]	
Across all industries	47
Within same two-digit SIC	16
Across all industries:	
Number of loser counties per winner county:	
1	31
2+	16
Reported year − matched year:[b]	
−2 to −1	20
0	15
1 to 3	12
Reported year of MDP location:	
1981–85	11
1986–89	18
1990–93	18
MDP characteristics, 5 years after opening:[c]	
Output ($1,000s)	452,801
	(901,690)
Output, relative to county output 1 year prior	.086
	(.109)
Hours of labor (1,000s)	2,986
	(6,789)

[a] 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 8 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 MDP's two-digit SIC in both locations.

[b] Only a few of these differences are 3. Census confidentiality rules prevent our being more specific.

[c] 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 two-digit SIC. Standard deviations are reported in parentheses. All monetary amounts are in 2006 U.S. dollars.

B. Summary Statistics

Table 1 presents summary statistics on the sample of plant location decisions that form the basis of the analysis. As discussed in the previous subsection, there are 47 manufacturing MDP openings that we can

---

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 agglomeration spillovers. If the agglomeration spillovers 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 spillovers may not be evident until the plant is operating. On the basis of these data and conceptual issues, there is not clear guidance on when the new plant could affect other plants. To be conservative and allow for each possibility, we emphasize results using the earliest of (1) the publication year of the magazine article and (2) the year that the matched MDP appears in the SSEL.

match to plant-level data. There are plants in the same two-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.[18] We refer to the winner and accompanying loser(s) associated with each plant opening 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 winner and/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 publication and the year the plant appears in the SSEL is roughly spread evenly across the categories −2 to −1 years, 0 years, and 1–3 years. For clarity, positive differences refer to cases in which the article appears after the plant is identified in the SSEL. The dates of the plant openings range from the early 1980s to the early 1990s.

The remainder of table 1 provides summary statistics on the MDPs 5 years after their assigned opening date. These MDPs are quite large: they are more than twice the size of the average incumbent plant and account for roughly 9 percent of the average county's total output 1 year prior to their opening.

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.[19] 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 calculated among plants that appear in the ASM only for at least 9 consecutive years. Column 4 presents the *t*-statistics from a test that the entries in columns 1 and 2 are equal, and

---

[18] 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.

[19] The losing county entries in col. 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.

TABLE 2
SUMMARY STATISTICS FOR MEASURES OF INDUSTRY LINKAGES

MEASURE OF INDUSTRY LINKAGE	DESCRIPTION	All Plants	MEAN Only 1st Quartile	Only 4th Quartile	STANDARD DEVIATION
Labor market pooling:					
CPS worker transitions	Proportion of workers leaving a job in this industry that move to the MDP industry (15 months later)	.119	.002	.317	.249
Intellectual or technology spillovers:					
Citation pattern	Percentage of manufactured industry patents that cite patents manufactured in MDP industry	.022	.001	.057	.033
Technology input	R&D flows from MDP industry, as a percentage of all private-sector technological expenditures	.022	.000	.106	.084
Technology output	R&D flows to MDP industry, as a percentage of all original research expenditures	.011	.000	.042	.035
Proximity to customers and suppliers:					
Manufacturing input	Industry inputs from MDP industry, as a percentage of its manufacturing inputs	.017	.000	.075	.061
Manufacturing output	Industry output used by MDP industry, as a percentage of its output to manufacturers	.042	.000	.163	.139

NOTE.—The variable "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 two-digit SIC values. The five other measures of cross-industry relationships were provided by Ellison et al. (forthcoming). These measures are defined in a three-digit SIC by three-digit SIC matrix, though much of the variation is at the two-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 percent, and col. 3 reports the mean for the highest 25 percent. 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 8 years prior to the MDP opening.

TABLE 3

County and Plant Characteristics by Winner Status, 1 Year Prior to a Million Dollar Plant Opening

	All Plants					Within Same Industry (Two-Digit SIC)				
	Winning Counties (1)	Losing Counties (2)	All U.S. Counties (3)	t-Statistic (Col. 1 − Col. 2) (4)	t-Statistic (Col. 1 − Col. 3) (5)	Winning Counties (6)	Losing Counties (7)	All U.S. Counties (8)	t-Statistic (Col. 6 − Col. 7) (9)	t-Statistic (Col. 6 − Col. 8) (10)
					A. County Characteristics					
No. of counties	47	73				16	19			
Total per capita earnings ($)	17,418	20,628	11,259	**−2.05**	**5.79**	20,230	20,528	11,378	−.11	**4.62**
% change, over last 6 years	.074	.096	.037	−.81	1.67	.076	.089	.057	−.28	.57
Population	322,745	447,876	82,381	−1.61	**4.33**	357,955	504,342	83,430	−1.17	**3.26**
% change, over last 6 years	.102	.051	.036	**2.06**	**3.22**	.070	.032	.031	1.18	1.63
Employment-population ratio	.535	.579	.461	−1.41	**3.49**	.602	.569	.467	.64	**3.63**
Change, over last 6 years	.041	.047	.023	−.68	**2.54**	.045	.038	.028	.39	1.57
Manufacturing labor share	.314	.251	.252	**2.35**	**3.12**	.296	.227	.251	1.60	1.17
Change, over last 6 years	−.014	−.031	−.008	1.52	−.64	−.030	−.040	−.007	.87	**−3.17**
					B. Plant Characteristics					
No. of sample plants	18.8	25.6	7.98			2.75	3.92	2.38	−1.14	.70
Output ($1,000s)	190,039	181,454	123,187	−1.35	**3.02**	217,950	178,958	132,571	.41	1.25
% change, over last 6 years	.082	.082	.118	.25	**2.14**	−.061	.177	.182	−1.23	**−3.38**
Hours of labor (1,000s)	1,508	1,168	877	.01	−.97	1,738	1,198	1,050	.92	1.33
% change, over last 6 years	.122	.081	.115	1.52	**2.43**	.160	.023	.144	.85	.13

Note.—For each case to be weighted equally, counties are weighted by the inverse of the number of counties 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 8 years prior. Excluded are all plants owned by the firm opening an MDP. Also excluded are all plants from two uncommon two-digit SIC values so that subsequently estimated clustered variance matrices would always be positive definite. The sample of all U.S. counties excludes winning counties and counties with no manufacturing plant reporting data in the ASM for 9 consecutive years. These other U.S. 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. t-statistics greater than 2 are reported in bold. All monetary amounts are in 2006 U.S. dollars.

column 5 repeats this for a test of equality between columns 1 and 3. Columns 6–10 repeat this exercise among the cases in which there are plants within the same two-digit SIC industry as the MDP. In these columns, the plant characteristics are calculated among the plants in the same two-digit industry.

This exercise provides an opportunity to assess the validity of the research design, as measured by preexisting 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.

Panel A reports 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. 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 eight variables in this panel, six of the eight differences are statistically significant at conventional levels. These differences are substantially mitigated when the winners are compared to losers: three of the eight variables are statistically different at the 5 percent level, and none are at the 1 percent level. Notably, the raw differences between winners and losers within the subset of cases in which there are plants in the same two-digit SIC industry are generally smaller, and none are statistically significant.

Panel B 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 counties (and just 8.0 in the average U.S. county). 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 plants within the same two-digit industry.[20]

Overall, table 3 shows that the MDP winner-loser research design balances many (although not all) observable county-level and plant-level covariates. Of course, this exercise does not guarantee that unobserved variables are balanced across winner and loser counties or their plants. In the subsequent analysis, we find that trends in TFP were similar in winning and losing counties prior to the MDP opening, which lends

---

[20] Roughly 20 percent of the winners were in the Rust Belt, compared to roughly 25 percent of the losers (the Rust Belt is defined as Michigan, Indiana, Ohio, Pennsylvania, New Jersey, Illinois, Wisconsin, and New York). Roughly 65 percent of the winners were in the South, compared to roughly 45 percent of the losers.

further credibility to this design. The next section outlines our full econometric model and highlights the exact assumptions necessary for consistent estimation.

V.   Econometric Model

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

$$Y_{pijt} = A_{pijt} L_{pijt}^{\beta_1} K_{pijt}^{B\beta_2} K_{pijt}^{E\beta_3} M_{pijt}^{\beta_4}, \tag{6}$$

where p references plant, i industry, j case, and t year; Y_{pijt} is the total value of shipments minus changes in inventories; 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 data on book values and deflated values of subsequent investment.[21]

Recall that equation (1) in Section II 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 differences in TFP across plants α_p, cases λ_p, industry-specific time-varying shocks to TFP μ_{it}, and a stochastic error term ε_{pijt}:

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

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 and its growth over time:

---

[21] For the first date available, plants' historical capital stock book values are deflated to constant dollars using Bureau of Economic Analysis data by two-digit industry. In all periods, plants' investment is deflated to the same constant dollars using Federal Reserve data by three-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 Davis, Haltiwanger, and Schuh (1996), Chiang (2004), and Becker et al. (2005), from whose files we gratefully obtained deflators.

$$\ln (A_{pijt}) = \delta 1(\text{Winner})_{pj} + \psi \text{Trend}_{jt} + \Omega[\text{Trend}_{jt} \times 1(\text{Winner})_{pj}]$$
$$+ \kappa 1(\tau \geq 0)_{jt} + \gamma[\text{Trend}_{jt} \times 1(\tau \geq 0)_{jt}]$$
$$+ \theta_1[1(\text{Winner})_{pj} \times 1(\tau \geq 0)_{jt}]$$
$$+ \theta_2[\text{Trend}_{jt} \times 1(\text{Winner})_{pj} \times 1(\tau \geq 0)_{jt}]$$
$$+ \alpha_p + \mu_{it} + \lambda_p + \varepsilon_{pijt}, \tag{7}$$

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

$$\ln (Y_{pijt}) = \beta_1 \ln (L_{pijt}) + \beta_2 \ln (K_{pijt}^B) + \beta_3 \ln (K_{pijt}^E) + \beta_4 \ln (M_{pijt})$$
$$+ \delta 1(\text{Winner})_{pj} + \psi \text{Trend}_{jt} + \Omega[\text{Trend}_{jt} \times 1(\text{Winner})_{pj}]$$
$$+ \kappa 1(\tau \geq 0)_{jt} + \gamma[\text{Trend}_{jt} \times 1(\tau \geq 0)_{jt}]$$
$$+ \theta_1[1(\text{Winner})_{pj} \times 1(\tau \geq 0)_{jt}]$$
$$+ \theta_2[\text{Trend}_{jt} \times 1(\text{Winner})_{pj} \times 1(\tau \geq 0)_{jt}]$$
$$+ \alpha_p + \mu_{it} + \lambda_p + \varepsilon_{pijt}. \tag{8}$$

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 spillover effects of the new plant on incumbent plants' TFP, so the parameters of interest are θ_1 and θ_2. The former tests for a mean shift in TFP among incumbent plants in the winning county after the opening of the MDP, and the latter tests 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 more 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 rules out differential trends. This specification is essentially a difference-in-difference estimator, and we refer to it as model 1. Formally, after adjustment for the inputs, $1(\text{Winner})_{pj}$, and $1(\tau \geq 0)_{jt}$, the consistency of θ_1 in this model requires the assumption that

$$E[[1(\text{Winner})_{pj} \times 1(\tau \geq 0)_{jt}]\varepsilon_{pijt}|\alpha_p, \mu_{it}, \lambda_p] = 0.$$

In other specifications, we estimate the entirety of equation (8) without imposing such restrictions on the trends and label this model 2. This specification allows for both a mean shift and a trend break in productivity. In theory, model 2 allows us to investigate whether any productivity effect occurs immediately and whether the impact evolves over time. In practice, disentangling these effects is demanding of the data because our sample is balanced only through $\tau = 5$ and there are only 6 years per case to estimate θ_1 and θ_2. The other main practical difference between model 1 and model 2 is that the latter allows for differential pre-trends in incumbent plants' TFP.

The other terms in equation (8) control for unobserved determinants of TFP that might otherwise be confounded with the spillover effects of the MDP opening. These terms control for TFP differences in winning counties (δ), a time trend in winning and losing counties (ψ), a change in winning and losing counties after the MDP opening (κ), a trend break in winning and losing counties after the MDP opening (γ), and a differential time trend in winning counties prior to the MDP opening (Ω). This differential pre-trend in winning counties (Ω) will serve as an important way to assess the validity of this research design. The specification also includes three sets of fixed effects: plant fixed effects (α_p), so the comparisons are within a plant; two-digit SIC industry by year fixed effects (μ_{it}) to account for industry-specific TFP shocks; and separate fixed effects for each case (λ_p) to ensure that the impact of the MDP opening is identified from comparisons within a winner-loser pair. These case fixed effects recreate in a regression framework the intuitive appeal of pairwise differencing within cases, averaging this effect across all cases.

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 β's are inconsistent (see, e.g., Griliches and Mairesse 1995). This has been a topic of considerable research, and we are unaware of a complete solution. In a variety of robustness specifications, we implement the standard fixes, including modeling the inputs with alternative functional forms (e.g., the translog); fixing the β's equal to their cost shares at the plant and industry levels; controlling for flexible functions of investment, capital, materials, and labor; and instrumenting for current inputs with lagged changes in inputs (Olley and Pakes 1996; Blundell and Bond 1998; Levinsohn and Petrin 2003; Syverson 2004a, 2004b; van Biesebroeck 2004; Ackerberg, Caves, and Frazer 2006). 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 a concern for the consistent

estimation of our main parameters of interest (θ_1 and θ_2) only if they systematically affect incumbent plants in winning counties in the years after the MDP opening, controlling for the rich set of covariates in equation (8).

Second, in some cases this equation is estimated 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 $\tau = 0$.[22] This smaller sample of plants from only winning and losing counties allows for the impact of the inputs and the industry shocks to differ in these counties from those for the rest of the country. For most of the analysis, we further restrict the sample to observations in the years between $\tau = -7$ and $\tau = 5$. Because of the dates of the MDP openings, this is the longest period for which we have data from all cases.[23]

Third, we probe the validity and robustness of our estimates with a number of supplementary specifications. For example, we investigate how the estimates may be influenced by unobserved changes in plant inputs, attrition of sample plants, mismeasurement of TFP, and changes in prices of incumbents' output. A complementary analysis of plants' factor input demand provides corroborating evidence for TFP increases, without many of the biases associated with estimating plant-level TFP.

Fourth, all the reported standard errors are clustered at the county level to account for the correlation in outcomes among plants in the same county, both within periods and over time.

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.[24]

VI.   Results

This section is divided into three subsections. Subsection A reports baseline estimates of the effect of the opening of a new Million Dollar Plant

[22] When data from the entire country are used, the sample is limited to plants that are in the ASM for at least 14 consecutive years.

[23] Data from all cases are also available for $\tau = -8$, but shipments in this period are used to weight the regressions.

[24] Finally, we hired a graduate student at Princeton to review publicly disclosed and annotated versions of all STATA programs. This person was not associated with the authors or their institutions prior to serving as the program proofreader. To the best of his knowledge, the computer codes were correct. The authors remain fully responsible for any coding errors in the analysis.

TABLE 4
Incumbent Plant Productivity, Relative to the Year of
an MDP Opening

Event Year	In Winning Counties (1)	In Losing Counties (2)	Difference Col. 1 − Col. 2 (3)
$\tau = -7$	.067	.040	.027
	(.058)	(.053)	(.032)
$\tau = -6$	.047	.028	.018
	(.044)	(.046)	(.023)
$\tau = -5$	.041	.021	.020
	(.036)	(.040)	(.025)
$\tau = -4$	−.003	.012	−.015
	(.030)	(.030)	(.024)
$\tau = -3$	.011	−.013	.024
	(.022)	(.022)	(.021)
$\tau = -2$	−.003	.001	−.005
	(.027)	(.011)	(.028)
$\tau = -1$	0	0	0
$\tau = 0$	.013	−.010	.023
	(.018)	(.011)	(.019)
$\tau = 1$	.023	−.028	.051**
	(.026)	(.024)	(.023)
$\tau = 2$	.004	−.046	.050
	(.036)	(.046)	(.033)
$\tau = 3$	.003	−.073	.076*
	(.047)	(.057)	(.043)
$\tau = 4$	.004	−.072	.076**
	(.053)	(.062)	(.033)
$\tau = 5$	−.023	−.100	.077**
	(.069)	(.067)	(.035)
R^2	.9861		
Observations	28,732		

Note.—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 by two-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 8 years prior to the MDP opening. Data on plants in all cases are available only 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 cols. 1 and 2 of table 3.
  * Significant at the 10 percent level.
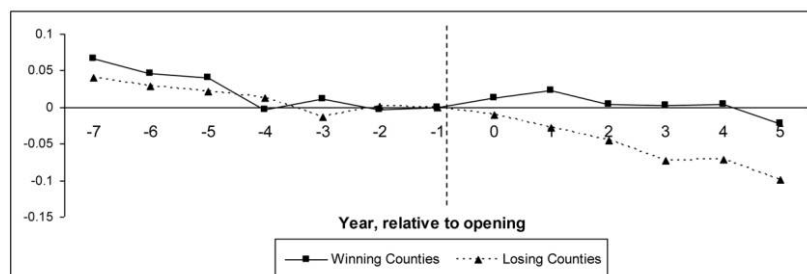 ** Significant at the 5 percent level.
  *** Significant at the 1 percent level.

on the productivity of incumbent plants in the same county through the estimation of equation (8). Subsection B explores potential channels for the agglomeration effects by testing whether the estimated spillovers vary as a function of economic distance. Subsection C explores the implications of the estimates for 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

All Industries: Winners vs. Losers
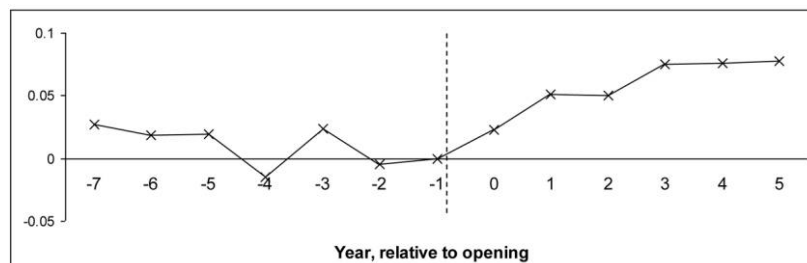

Difference: Winners – Losers

Fig. 1.—All incumbent plants' productivity in winning versus losing counties, relative to the year of an MDP opening. These figures accompany table 4.

log of output is regressed on the natural log of inputs, year by two-digit SIC industry fixed effects, plant fixed effects, case fixed effects, and the event time indicators in a sample that is restricted to the years $\tau = -7$ through $\tau = 5$. The reported coefficients on the event time indicators reflect yearly mean TFP in winning counties (col. 1) and losing counties (col. 2), relative to the year before the MDP opened. Column 3 reports the yearly difference between estimated mean TFP in winning and losing counties.

Figure 1 graphs the estimated coefficients from table 4. The top panel separately plots mean TFP in winning and losing counties (cols. 1 and 2 of table 4). The bottom panel plots the differences in the estimated winner and loser coefficients (col. 3 of table 4).

The figure has three important features. First, in the years before the MDP opening, TFP trends among incumbent plants were very similar in winning and losing counties. Indeed, 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, beginning in the year of the MDP opening, there is a sharp upward break in the difference in TFP between the winning and losing counties. The top panel shows that this relative improvement is mainly due to the continued TFP decline in losing counties and a flattening of the TFP trend in winning counties. This underscores the importance of the availability of losing counties as a counterfactual. For example, a naive comparison of TFP in winning counties before and after the MDP opening would suggest that it had a negligible 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 is confirmed throughout a variety of tests in the remainder of the paper. Third, TFP displays a negative trend. We discuss this feature in detail in Section VII.

For the statistical models, columns 1–4 of table 5 present results from fitting different versions of equation (8). For model 1, panel A reports the estimated mean shift parameter, θ_1, and its standard error (in parentheses) in the mean shift row. For model 2, panel B reports the estimated change in TFP evaluated at $\tau = 5$ in the effect after 5 years row, which is determined by the reported θ_1 (level change row) and θ_2 (trend break row).[25] The pre-trend row contains the coefficient measuring the difference in preexisting trends between plants in winning and losing counties. In all specifications, the estimated change after the MDP opening is determined during the period in which τ ranges from -7 through 5, since 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 an MDP. This restriction means that the input parameters and the industry by year fixed effects are estimated solely from plants in these counties. Incumbent plants are now required to be in the data only for $-8 \le \tau \le 0$ (not for 14 consecutive years, though this does not change the results). Finally, in column 4, the sample is restricted further to include only plant by year observations within the period of interest (where τ ranges from -7 through 5). This forces the input parameters and industry by year fixed effects to be estimated solely on plant by year observations that identify the spillover parameters. This sample is used throughout the remainder

---

[25] This is calculated as $\theta_1 + 6\theta_2$, because we allow the MDP to affect outcomes from $\tau = 0$ through $\tau = 5$.

TABLE 5
CHANGES IN INCUMBENT PLANT PRODUCTIVITY FOLLOWING AN MDP OPENING

	ALL COUNTIES: MDP WINNERS − MDP LOSERS		MDP COUNTIES: MDP WINNERS − MDP LOSERS		ALL COUNTIES: RANDOM WINNERS
	(1)	(2)	(3)	(4)	(5)
	A. Model 1				
Mean shift	.0442*	.0435*	.0524**	.0477**	− 0.0496***
	(.0233)	(.0235)	(.0225)	(.0231)	(.0174)
				[$170 m]	
R^2	.9811	.9812	.9812	.9860	∼0.98
Observations (plant by year)	418,064	418,064	50,842	28,732	∼400,000
	B. Model 2				
Effect after 5 years	.1301**	.1324**	.1355***	.1203**	−.0296
	(.0533)	(.0529)	(.0477)	(.0517)	(.0434)
				[$429 m]	
Level change	.0277	.0251	.0255	.0290	.0073
	(.0241)	(.0221)	(.0186)	(.0210)	(.0223)
Trend break	.0171*	.0179**	.0183**	.0152*	− 0.0062
	(.0091)	(.0088)	(.0078)	(.0079)	(.0063)
Pre-trend	−.0057	−.0058	−.0048	−.0044	−.0048
	(.0046)	(.0046)	(.0046)	(.0044)	(.0040)
R^2	.9811	.9812	.9813	.9861	∼.98
Observations (plant by year)	418,064	418,064	50,842	28,732	∼400,000
Plant and industry by year fixed effects	Yes	Yes	Yes	Yes	Yes
Case fixed effects	No	Yes	Yes	Yes	NA
Years included	All	All	All	$-7 \leq \tau \leq 5$	All

NOTE.—The table reports results from fitting several versions of eq. (8). Specifically, entries are from a regression of the natural log of output on the natural log of inputs, year by two-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 cols. 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 one during the period that τ ranges from −7 through 5. In cols. 3 and 4, the sample is restricted to include only plants in counties that won or lost an MDP. This forces the industry by year fixed effects to be estimated solely from plants in these counties. For col. 4, the sample is restricted further to include only plant by year observations within the period of interest (where τ ranges from −7 to 5). This forces the industry by year fixed effects to be estimated solely on plant by year observations that identify the parameters of interest. In col. 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 (this procedure was run 1,000 times, and reported are the means and standard deviations of those estimates). For all regressions, plant by year observations are weighted by the plant's total value of shipments 8 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 two-digit SIC values were excluded so that estimated clustered variance-covariance matrices would always be positive definite. In brackets is the value in 2006 U.S. dollars from the estimated increase in productivity: the percentage increase is multiplied by the total value of output for the affected incumbent plants in the winning counties. Reported in parentheses are standard errors clustered at the county level.

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

of the paper. Further 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 MDP opening is associated with a substantial relative increase in TFP among incumbent plants in winning counties. Specifically, model 1 implies an increase in TFP of roughly 4.8 percent. As the figure highlights, however, the impact on TFP appears to be increasing over time, so model 2 seems more appropriate. Results from model 2 suggest that the MDP opening is associated with an approximately 12 percent increase in TFP 5 years later. Estimates from both models are statistically different from zero by conventional criteria and are unaffected by the specification changes. Furthermore, 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.

In column 4, the numbers in brackets evaluate the average magnitude of TFP change in millions of 2006 dollars. These numbers are calculated by multiplying the estimated percentage change by the mean value of incumbent plants' total shipments in winning counties in $\tau = -1$. For model 1, this calculation indicates that the increase in TFP following an MDP opening was associated with an annual increase in total output of \$170 million. 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. Section VIII discusses the interpretation of this change and its magnitude.

Column 5 presents results from a "naive" estimator that is based on using plant openings without an explicit counterfactual. To begin, a set of 47 plant openings was 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 for 14 consecutive years and not also owned by firms that own the randomly chosen plants. With these data, we fit a regression of the natural log of output on the natural log of inputs, year by two-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. Finally, this procedure is implemented 1,000 times, and the reported parameters are the mean and standard deviation of those estimates.

This naive "first-difference" style estimator indicates that the opening
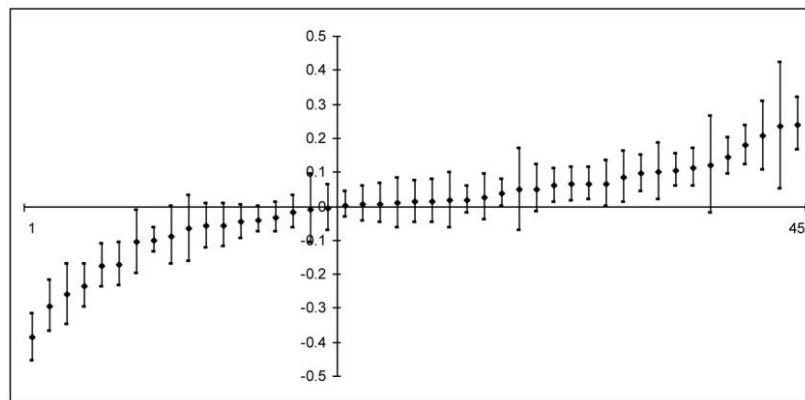
FIG. 2.—Distribution of case-specific mean shift estimates, following an MDP opening. The figure reports results from a version of model 1 that estimates the parameter θ_1 for each of the 47 MDP cases. The figure reports only 45 estimates because two cases were excluded for Census confidentiality reasons.

of a new plant is associated with a -3 percent to -5 percent change in incumbent plants' TFP, depending on the model. If the estimates from the MDP research design are correct, then this naive approach understates the extent of spillovers by 10 percent (model 1) to 15 percent (model 2). The estimated pre-trend indicates that the TFP of incumbent plants was on a downward trend in advance of the randomly selected new plant openings. This is similar to what is observed in our MDP sample of winners. Overall, the absence of a credible research design can lead to misleading inferences in this setting.

It is important to document 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 θ_1 in model 1 and their 95 percent confidence intervals. Specifically, the figure plots results from a version of model 1 that interacts the variable 1(Winner) $\times$ 1($\tau \geq 0$) with indicators for each of the cases. This specification yields 45 estimates of θ_1 since results from two cases were omitted to comply with the Census Bureau's confidentiality rules. Figure 2 reveals that there is substantial heterogeneity in the estimated impacts on TFP of incumbent plants. Twenty-seven of the 45 estimates are positive. Thirteen of the positive estimates and nine of the negative estimates are statistically different from zero at the 5 percent level. We explored whether this heterogeneity is related to the MDP characteristics, but the limited number of cases provides insufficient power to detect much with confidence. Specifically, we regressed the estimates against three measures of the MDP's size,

TABLE 6
CHANGES IN INCUMBENT PLANT OUTPUT AND INPUTS FOLLOWING AN MDP OPENING

	Output (1)	Worker Hours (2)	Machinery Capital (3)	Building Capital (4)	Materials (5)
Model 1: mean shift	.1200***	.0789**	.0401	.1327*	.0911***
	(.0354)	(.0357)	(.0348)	(.0691)	(.0302)
Model 2: after 5 years	.0826*	.0562	−.0089	−.0077	.0509
	(.0478)	(.0469)	(.0300)	(.0375)	(.0541)

NOTE.—The table reports results from fitting versions of eq. (8) for each of the indicated outcome variables (in logs). See the text for more details. Standard errors clustered at the county level are reported in parentheses.
    * Significant at the 10 percent level.
    ** Significant at the 5 percent level.
    *** Significant at the 1 percent level.

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.[26]

Ultimately, TFP is a residual, and residual labeling must be done cautiously. As an alternative way to examine the MDP impact, we estimate directly the changes in incumbent plant output (unadjusted for inputs) and inputs following an MDP opening. Contrasting changes in outputs and inputs can 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 opening? Factor input decisions also reflect firms' optimization decisions and do not share many of the same potential biases as changes in technology (e.g., output price effects).

Table 6 reports estimated changes in incumbent plant output and inputs following an MDP opening. These estimates are from the model 1 and model 2 versions of equation (8) but exclude the inputs as covariates. For model 1, output increases by 12 percent (col. 1) and inputs increase by 4–13 percent (cols. 2–5). For model 2, output increases by 8 percent and inputs increase less. Across all specifications, it is striking that the change in all of the inputs is roughly equal to or less than the increase in output. Overall, it appears that incumbent plants produced more with less after the MDP opening, which is consistent with the TFP

[26] 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. However, we failed to find any significant differences when separately testing whether the productivity effect varied by the ratio of the MDP's output to countywide manufacturing output, whether the MDP is owned by a foreign company, or whether the MDP is an auto company.

increases uncovered in table 5.[27] Furthermore, there is some evidence of increased input use, reflecting firms' optimization in the face of higher potential productivity.

B.   *Estimates of Spillovers by Economic Distance*

What mechanisms might explain the productivity gains estimated above? Section II.A discussed some mechanisms that may be responsible for agglomeration spillovers. This subsection attempts to shed some light on the possible mechanisms by investigating how the estimated spillover effect varies as a function of economic distance. A similar approach has been used by Moretti (2004*c*) and Ellison et al. (forthcoming).

1.   By Industry

Table 7 shows separate estimates from the baseline model for samples of incumbent plants in the MDP's two-digit industry and all other industries. In general, one might expect agglomeration spillovers to decline with economic distance (eq. [5′]). As a first pass, it is natural to explore whether spillovers are larger within an industry. While there can be substantial heterogeneity in technologies and labor forces among plants within a two-digit SIC industry, only 16 of the 47 cases have incumbent plants in the MDP's two-digit industry. Thus, the research design and available data do not permit a discrete analysis at finer industry definitions.[28]

  Column 1 of table 7 reports estimates for all industries from column 4 of table 5 as a basis of comparison. Columns 2 and 3 report estimates from the baseline specification for incumbent plants in the MDP's two-digit industry and all other industries, respectively. The entries in these columns are from the same regression. As in table 5, the numbers in

[27] The model suggests that firms should substitute away from labor and toward capital. The point estimates are not supportive of this prediction, though directly estimating changes in the capital/labor ratio gives imprecise estimates, making definitive conclusions unwarranted.

[28] In the spirit of work by Jaffe (1986) and Bloom, Shankerman, and Van Reenen (2007), we explore defining a continuous measure of technological overlap between industries. Lacking patent data, we define at the three-digit SIC industry level (1) the share of industrial output that is sold to each manufacturing industry and (2) the share of manufactured inputs that are received from each manufacturing industry. For each measure, we calculate the overlap between an incumbent firm's industry and the MDP industry by taking the product of those vectors. We then estimate eq. (9) below, interacting the MDP effect with each measure of industrial overlap. There is evidence of differential spillovers based on overlap defined with inputs, but not with outputs. We suspect that overlap in plant output consumed by only the manufacturing sector is a poor reflection of overall industrial overlap, and we are not confident that overlap in plant inputs is more persuasive.

TABLE 7
Changes in Incumbent Plant Productivity Following an MDP Opening for
Incumbent Plants in the MDP's Two-Digit Industry and All Other Industries

	All Industries (1)	MDP's Two-Digit Industry (2)	All Other Two-Digit Industries (3)
	A. Model 1		
Mean shift	.0477**	.1700**	.0326
	(.0231)	(.0743)	(.0253)
	[$170 m]	[$102 m]	[$104 m]
R^2	.9860	.9861	
Observations	28,732	28,732	
	B. Model 2		
Effect after 5 years	.1203**	.3289	.0889*
	(.0517)	(.2684)	(.0504)
	[$429 m]	[$197 m]	[$283 m]
Level change	.0290	.2814***	.0004
	(.0210)	(.0895)	(.0171)
Trend break	.0152*	.0079	.0147*
	(.0079)	(.0344)	(.0081)
Pre-trend	−.0044	−.0174	−.0026
	(.0044)	(.0265)	(.0036)
R^2	.9861	.9862	
Observations	28,732	28,732	

Note.—The table reports results from fitting versions of eq. (8). As a basis for comparison, col. 1 reports estimates from the baseline specification for incumbent plants in all industries (baseline estimates for incumbent plants in all industries, col. 4 of table 5). Columns 2 and 3 report estimates from a single regression, which fully interacts the winner/loser and pre/post variables with indicators for whether the incumbent plant is in the same two-digit industry as the MDP or a different industry. Reported in parentheses are standard errors clustered at the county level. The numbers in brackets are the value (2006 U.S. dollars) from the estimated increase in productivity: the percentage increase is multiplied by the total value of output for the affected incumbent plants in the winning counties.
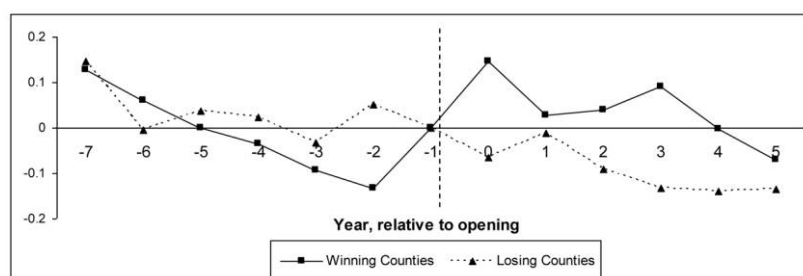 * Significant at the 10 percent level.
 ** Significant at the 5 percent level.
*** Significant at the 1 percent level.

brackets convert the estimated percentage changes into millions of 2006 dollars.

The estimated changes are substantially larger in the MDP's own two-digit industry. For example, the estimated increase in TFP for plants in the same two-digit industry is a statistically significant 17 percent in model 1 and a poorly determined 33 percent at $\tau = 5$ in model 2. In contrast, estimates for plants in other industries are a statistically insignificant 3.3 percent in model 1 and a marginally significant 8.9 percent in model 2.

Figures 3 and 4 graph annual changes in TFP, providing two-digit MDP industry and other industry analogues to figure 1. The two-digit MDP industry estimates are noisy because of the small sample size, which was also evident in the statistical results. Importantly, there is not any evidence of differential trends in the years before the MDP's opening, and statistical tests confirm this visual impression. As in figure 1, the

2-digit MDP Industry: Winners Vs. Losers
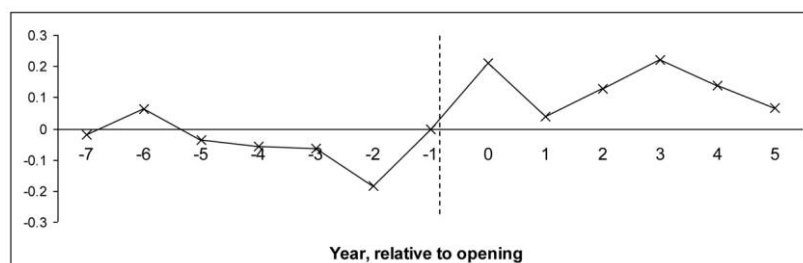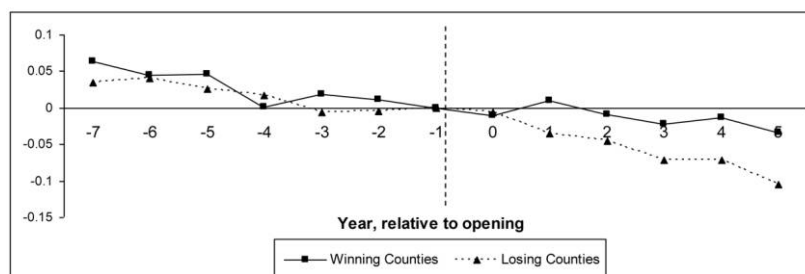

Difference (Winners – Losers)

Fɪɢ. 3.—Incumbent plants' productivity in the MDP's two-digit industry, winning versus losing counties, relative to the year of an MDP opening. These figures accompany table 7, column 2 (MDP's two-digit industry).

estimated impact reflects the continuation of a downward trend in TFP in losing counties and a cessation of the downward trend in winning counties.

To probe the role of economic distance further, we identified an additional 10 MDP openings that were in the retail and wholesale trade sectors. These plants are part of the original 82 MDP openings but are not included in the main sample of 47 manufacturing MDP openings. Estimating equation (8) for these 10 trade sector MDP openings, we find TFP changes of −2.2 percent (2.9 percent) in model 1 and 4.9 percent (6.5 percent) in model 2 (on a sample of 12,105 plant by year observations in 31 counties). It appears that the nonmanufacturing sector openings did not generate similar TFP increases, though the estimates in model 2 are too imprecise to reject their equality with the baseline estimates. These findings provide further evidence that the

Other Industries: Winners Vs. Losers

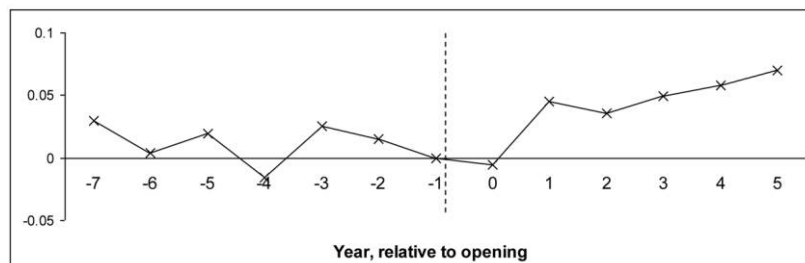**Difference (Winners – Losers)**

FIG. 4.—Incumbent plants' productivity in other industries (not the MDP's two-digit industry), winning versus losing counties, relative to the year of an MDP opening. These figures accompany table 7, column 3 (all two-digit industries, except the MDP's two-digit industry).

spillovers are concentrated among plants that are economically close to the new plant.[29]

2.   By Continuous Measures of Economic Distance

We now investigate the role of economic proximity more directly by using several measures of economic proximity that capture worker flows,

---

[29] These results may also provide a test of whether the estimated spillovers are due to increased competition for inputs causing plants to move closer to their production possibility frontier. Specifically, these new nonmanufacturing plants increase competition for land, labor, and other local inputs. The resulting increase in input prices may cause all plants (regardless of industry) to search for opportunities to increase productivity. In such a situation, all local plants would exhibit increased TFP. These results suggest that this mechanism does not explain this paper's primary findings.

technological proximity, and input-output flows (table 2). To ease the interpretation, these economic proximity variables are standardized to have a mean of zero and a standard deviation of one. In all cases, a positive value indicates a "closer" relationship between the industries.

Specifically, we estimate the following equation:

$$
\begin{aligned}
\ln(Y_{pijt}) = {} & \beta_1 \ln(L_{pijt}) + \beta_2 \ln(K^B_{pijt}) + \beta_3 \ln(K^E_{pijt}) + \beta_4 \ln(M_{pijt}) \\
& + \delta 1(\text{Winner})_{pj} + \kappa 1(\tau \geq 0)_{jt} + \theta_1[1(\text{Winner})_{pj} \times 1(\tau \geq 0)_{jt}] \\
& + \pi_1[1(\text{Winner})_{pj} \times \text{Proximity}_{ij}] + \pi_2[1(\tau \geq 0)_{jt} \times \text{Proximity}_{ij}] \\
& + \pi_3[1(\text{Winner})_{pj} \times 1(\tau \geq 0)_{jt} \times \text{Proximity}_{ij}] \\
& + \alpha_p + \mu_{it} + \lambda_p + \varepsilon_{pijt},
\end{aligned}
\tag{9}
$$

where 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 proximity variables with $1(\text{Winner})_{pj}$, $1(\tau \geq 0)$, and $[1(\text{Winner}) \times 1(\tau \geq 0)]_{pjt}$. The coefficient of interest is π_3, which is the coefficient on the triple interaction between the dummy for a winning county, the dummy for after the MDP opening, and the measure of proximity. This coefficient assesses whether plants in closer industries experience a greater increase in TFP after the MDP 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, 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 8 reports estimates of π_3 for six measures of economic proximity. Columns 1–6 include the proximity measures one at a time. For example, column 1 reports that a one-standard-deviation increase in the CPS worker transitions variable between the incumbent plants' industry and the MDP's industry is associated with a 7-percentage-point 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 two-digit industries, so this finding was foreshadowed by the results in table 7 based on the plant's own two-digit industry.

In columns 2, 3, and 4, the measures of intellectual or technological linkages indicate meaningful increases in the spillover. The precise

TABLE 8
CHANGES IN INCUMBENT PLANT PRODUCTIVITY FOLLOWING AN MDP OPENING, BY
MEASURES OF ECONOMIC DISTANCE BETWEEN THE MDP'S INDUSTRY AND INCUMBENT
PLANT'S INDUSTRY

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CPS worker transitions	.0701*** (.0237)						.0374 (.0260)
Citation pattern		.0545*** (.0192)					.0256 (.0208)
Technology input			.0320* (.0173)				.0501 (.0421)
Technology output				.0596*** (.0216)			.0004 (.0434)
Manufacturing input					.0060 (.0123)		−.0473 (.0289)
Manufacturing output						.0150 (.0196)	−.0145 (.0230)
R^2	.9852	.9852	.9851	.9852	.9851	.9852	.9853
Observations	23,397	23,397	23,397	23,397	23,397	23,397	23,397

NOTE.—The table reports results from fitting versions of eq. (9), which is modified from eq. (8). Building on the model 1 specification in col. 4 of table 5, each column adds 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 (a continuous version of results in table 7). 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 col. 4 of table 5, but it is restricted to plants that have industry linkage data for each measure. 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. Reported in parentheses are standard errors clustered at the county level.
  * Significant at the 10 percent level.
  ** Significant at the 5 percent level.
  *** Significant at the 1 percent level.

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 two-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 in which an auto manufacturer encourages (or even forces) its suppliers to adopt more efficient production techniques. Recall that 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 and technology flows is consistent with the results in Dumais, Ellison, and Glaeser (2002) and Ellison et al. (forthcoming), whereas the finding on input and output flows stands in contrast with these papers' findings.

In the column 7 specification, we include all the measures of eco-

nomic proximity simultaneously. The labor flow, the citation pattern, and the technology input interactions all remain positive but are statistically insignificant. The customer and supplier interactions are negative and statistically insignificant.

Overall, this analysis provides support for the notion that spillovers occur between firms that share workers and between firms that use similar technologies. In terms of Section II.C, 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. The estimates in table 8 seem 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.

C.   Firm Entry and Labor Costs as Indirect Tests of Spillovers

Baseline estimates found economically substantial productivity gains for incumbent establishments following the opening of the new MDP. In the presence of positive spillovers, the model makes two empirical predictions that are explored in this subsection: increased firm entry and increased local input costs.

First, if productivity spillovers are larger than short-run increases in the cost of local inputs, the MDP county should experience entry by new firms (relative to losing counties). Table 9 tests this prediction at the county level. The entries in panel A come from regressions that use data from the Census of Manufactures, which is conducted every 5 years. The dependent variables are the log of the number of establishments (col. 1) and the log of total manufacturing output (col. 2) in the county. The sample is restricted to winning and losing counties, and all plants owned by MDP firms are excluded from both dependent variables. The covariates include county fixed effects, year fixed effects, case fixed effects, and an indicator for whether the observation is from after the MDP opening. The parameter of interest is associated with the interaction of indicators for an observation from a winning county and from after the MDP opening, so it is a difference-in-difference estimator of the impact of the MDP opening.[30]

---

[30] Because data are 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 specification in col. 1 is weighted by the number of plants in the county in years -6 to -10, and the specification in col. 2 is weighted by the county's total manufacturing output in years -6 to -10.

TABLE 9
CHANGES IN COUNTIES' NUMBER OF PLANTS, TOTAL OUTPUT, AND SKILL-ADJUSTED
WAGES FOLLOWING AN MDP OPENING

	A. CENSUS OF MANUFACTURES		B. CENSUS OF POPULATION
	Dependent Variable: Log(Plants) (1)	Dependent Variable: Log(Total Output) (2)	Dependent Variable: Log(Wage) (3)
Difference-in-difference	.1255**	.1454	.0268*
	(.0550)	(.0900)	(.0139)
R^2	.9984	.9931	.3623
Observations	209	209	1,057,999

NOTE.—The table reports results from fitting three regressions. In panel A, the dependent variables are the log of number of establishments and the log of total manufacturing output in the county, based on data from the Census of Manufactures. Controls include county, year, and case fixed effects. Reported are the county-level difference-in-difference estimates for receiving an MDP opening. Because data are available every 5 years, depending on the census year relative to the MDP opening, the sample years are defined to be 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 col. 1 model is weighted by the number of plants in the county in years −6 to −10, and the col. 2 model is weighted by the county's total manufacturing output in years −6 to −10. In panel B, the dependent variable is log wage and controls include dummies for age by year, age squared by year, education by year, sex by race by Hispanic by citizen, and case fixed effects. Reported is the county-level difference-in-difference estimate for receiving an MDP opening. Because data are available every 10 years, the sample years are defined to be 1–10 years before the MDP opening and 3–12 years after the MDP opening. As in panel A, each MDP opening is associated with one earlier date and one later date. The sample is restricted to individuals who worked more than 26 weeks in the previous year, usually work more than 20 hours per week, are not in school, are at work, and work for wages in the private sector. The number of observations reported refers to unique individuals: some Integrated Public Use Microdata Series county groups include more than one Federal Information Processing Standard (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. Reported in parentheses are standard errors clustered at the county level.
  * Significant at the 10 percent level.
  ** Significant at the 5 percent level.
  *** Significant at the 1 percent level.

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

Overall, these results are consistent with estimated increases in TFP since it appears that the MDP attracted new economic activity to the winning counties (relative to losing counties) in the manufacturing sector. Presumably, these new manufacturing establishments decided to locate in the winning counties to gain access to the productivity advantages generated by the spillover effect.

The second theoretical prediction is that if 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. This prediction is tested using individual-level wage data for winning and losing counties from the 1970, 1980, 1990,

and 2000 Censuses of Population.[31] These data are preferable to the measure of labor costs reported in the Census of Manufactures (i.e., the aggregate wage bill for production and nonproduction workers), which does not provide information on the quality of the labor force (e.g., education and experience). Specifically, we estimate changes in log wages, controlling for dummies for interactions of worker age and year, age-squared and year, education and year, sex and race and Hispanic and U.S. citizenship, and case fixed effects. We also include indicators for whether the observation is from a winning county and occurs after the MDP opening and the interaction of these two indicators.[32] This interaction is the focus of the regression and is an adjusted difference-in-difference estimator of the impact of the MDP opening on wages. This equation is analogous to the model 1 version of equation (8).

Column 3 in panel B of table 9 reports that wages increase by 2.7 percent in winning counties after the MDP opening, after adjusting for observable individual heterogeneity. This effect appears quantitatively sizable and is marginally statistically significant. Multiplying the estimated 2.7 percent 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 opening. This finding is consistent with positive spillovers and an upward-sloping labor supply curve, as in the model in Section II. This finding is also consistent with that of Moretti (2004c), who finds significant productivity spillovers and increases in wages of similar magnitude.

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 that the model 1 result in table 5 indicated an increase in TFP of approximately 4.8 percent (we focus on model 1 because it is not possible to estimate a version of model 2 with the decennial population Census data). If we assume that workers are homogeneous or that high- and low-skill workers are perfectly substitutable in production, then the labor market–wide increase in wages applies

---

[31] 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 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 Public Use Microdata Areas 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.

[32] The preperiod is defined as the most recent census before the MDP opening. The postperiod 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.

throughout the manufacturing sector. In our sample, labor accounts for roughly 23 percent of total costs, so the estimated 2.7 percent increase in skill-adjusted wages implies that manufacturers' costs increased by approximately 0.62 percent. The increased production costs due to higher wages are therefore 13 percent 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, we expect the wage increase to be larger for plants and industries that experience TFP increases as plants enter or expand and compete for workers with the skills relevant for these sectors. 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.

VII. Validity and Robustness

Our main empirical finding in Section VI is that MDP openings are associated with a substantial average increase in TFP among incumbent plants in those counties, relative to incumbent plants in counties that narrowly missed receiving the new plants. The validity of this research design is supported by the similarity of pre-trends in TFP (fig. 1) and the balancing of many ex ante observable characteristics of winning and losing counties and their incumbent plants (table 3). Nevertheless, the possibility remains that the paper's identifying assumption is invalid and that incumbent plants in winning counties experienced unobserved positive productivity shocks coincident to the new plant's opening.

Consequently, this section explores the robustness of the estimates to various specifications and investigates several possible alternative interpretations of the estimated spillover effects. Specifically, this section analyzes (*A*) the role of functional form assumptions, unobserved industry and regional shocks, and weighting; (*B*) the general endogeneity of plant inputs; (*C*) unobserved changes in inputs; (*D*) attrition; (*E*) declining plant TFP and mismeasurement; and (*F*) changes in the price of plant output.

A. *Functional Form, Industry and Regional Shocks, and Weighting*

Table 10 reports estimates from a series of specification checks. As a basis for comparison, column 1 reports the results from the preferred specification in column 4 of table 5.

We begin by generalizing our assumption on plants' production technology. Estimates in table 5 assume a Cobb-Douglas technology. In col-

TABLE 10
CHANGES IN INCUMBENT PLANT PRODUCTIVITY FOLLOWING AN MDP OPENING, ROBUSTNESS TO DIFFERENT SPECIFICATIONS

	Baseline Specification (1)	Translog Functional Form (2)	Input-Industry Interactions (3)	Input-Winner, Input-Post (4)	Region-Year Fixed Effects (5)	Region-Year-Industry Fixed Effects (6)	Unweighted (7)
Model 1: mean shift	.0477**	.0471**	.0406*	.0571**	.0442*	.0369*	.0146
	(.0231)	(.0226)	(.0220)	(.0245)	(.0230)	(.0215)	(.0107)
Model 2: after 5 years	.1203**	.1053*	.0977**	.1177**	.1176**	.0879**	.0065
	(.0517)	(.0535)	(.0487)	(.0538)	(.0520)	(.0442)	(.0281)

NOTE.—The table reports results from fitting several versions of eq. (8). Column 1 reports estimates from the baseline specification (col. 4 of table 5). Column 2 uses a translog functional form for inputs. Column 3 allows the effect of each input to differ by two-digit SIC value. Column 4 allows the effect of inputs to differ in winning/losing counties and before/after the MDP opening. Column 5 includes region (nine census divisions) by year fixed effects. Column 6 includes region by year by industry fixed effects. Column 7 reports on an unweighted version of eq. (8). Reported in parentheses are standard errors clustered at the county level.

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

umn 2 of table 10, 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 two-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 were 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 two-digit industry fixed effects. These specifications aim to purge the spillover effects of unobserved regionwide shocks or region by industry shocks to productivity that might be correlated with the probability of winning an MDP (e.g., a declining Rust Belt).

Until this point, we have presented results based on specifications that weight observations by the square root of the plant's total value of shipments 8 years prior to the MDP opening. As discussed above, the resulting estimates measure the change in productivity per average dollar of output, which reflects the full economic impact of the plant. Nevertheless, column 7 reports the results from unweighted regressions that reveal the change in productivity for the average plant. For model 1, the estimated change is 1.46 percent (1.07 percent); for model 2, the estimated change is 0.65 percent (2.81 percent). These findings should be interpreted cautiously because the building capital coefficient becomes slightly negative in both models, which may be a sign of misspecification. If this concern is set aside, the results indicate that the spillovers are concentrated among the largest plants.[33] A promising avenue for future research is to explore why smaller plants fail to benefit from the new plant's presence.

Taken together, the results in table 10 are striking. The weighted estimates appear to be insensitive to the specific functional form of the production function. None of the 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 an MDP

---

[33] In unweighted regressions, the estimated effect among incumbent plants in the largest decile (8 years prior to the MDP opening) is 2.90 percent (3.12 percent) higher than the average effect of 1.16 percent (0.98 percent) for model 1 and 16.7 percent (11.0 percent) higher than the average effect of -1.16 percent (3.04 percent) for model 2.

leads to a substantial increase in TFP among incumbent plants, and this is consistent with theories of spillovers.[34]

B. General 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). In contrast to the usual estimation of production functions, our aim is the consistent estimation of the spillover parameters, θ_1 and θ_2, so the endogeneity of capital and labor is relevant only to the extent that it results in biased estimates of these parameters. This subsection employs the productivity literature's techniques to control for the endogeneity of capital and labor to assess this issue's relevance in this paper's setting.

We employ three main approaches, and the results are collected in Appendix table A1. First, in columns 2 and 3, we calculate TFP for each plant by fixing the parameters on the inputs at the relevant input's share of total costs (Syverson 2004a; van Biesebroeck 2004; Foster, Haltiwanger, and Syverson 2008). 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 three-digit SIC industry level over the full sample, respectively.

Second, columns 4–6 present estimates based on methodologies that build on work by Olley and Pakes (1996). 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. Column 4 controls for fourth-degree polynomial functions of log capital and log investment and the interaction of both functions (separately for both types of capital). Column 5 includes the same controls as column 4 but replaces investment with materials, an alternative proposed by Levinsohn and Petrin (2003). Building on column 5, column 6 includes interactions between log labor and log materials, since collinearity may complicate the estimation of the labor coefficient (Ackerberg et al. 2006).

Third, column 7 presents estimates that instrument for current input levels with lagged changes in inputs, a technique proposed by Blundell

---

[34] We also tested whether the results are sensitive to the choice of the date of the MDP opening. When we use the year that the plant is first observed in the SSEL as the MDP opening date, model 1 estimates a change of 5.23 percent (2.39 percent) and model 2 estimates a change of 11.2 percent (5.57 percent). When we use the year of the MDP article for the opening date, model 1 estimates a change of 4.58 percent (2.45 percent) and model 2 estimates a change of 4.88 percent (4.21 percent).

and Bond (1998). The increase in each input from $t = -2$ to $t = -1$ may predict input levels at $t = 0$ but may not be correlated with un-observed output shocks in $t = 0$. Indeed, the estimated first-stage results (not shown) have the expected sign, and column 7 reports the two-stage least-squares results. Of course, it is a strong assumption that this lagged change is not otherwise correlated with output, and we have used only the first lagged change due to potential weak instrument bias.

   Appendix table A1 also reports coefficients from the production func-tion as a way of assessing the effectiveness of the production function estimation. The typical endogeneity concern is that unobserved pro-ductivity shocks lead to changes in variable inputs (labor) but not fixed inputs (capital), so the estimated effect of capital is downward biased and is loaded onto labor. In the baseline specification, all inputs are positive and statistically significant, and the labor coefficient is an ex-pected 72 percent of the summed coefficients for labor and capital.[35] The overall production function has mild decreasing returns to scale, with a 1 percent increase in all inputs leading to a 0.86 percent increase in output.

   Overall, the estimated changes in TFP are consistent with the findings from the baseline specification. This exercise fails to suggest that the possible endogeneity of labor and capital is the source of the estimated productivity spillovers.

C.   Unobserved Changes in Inputs

The input measures in the ASM are not comprehensive of all inputs that affect plant output. Further, the available data may not adequately measure the degree of input usage or the quality of inputs. Conse-quently, it is possible that the estimated spillovers reflect changes in unobserved inputs, unobserved usage, and/or input quality. This sub-section explores these possibilities.

   State and local governments frequently offer substantial subsidies to new manufacturing plants to locate within their jurisdictions. These incentives can include tax breaks, worker training funds, the construc-tion of roads, and other infrastructure investments. It is possible that these investments benefit firms other than the MDP. For example, the construction of a new road intended for an MDP may also benefit the productivity of some of the incumbent firms (Chandra and Thompson 2000). If the productivity gains we have documented are due to public

---

[35] As a basis of comparison, the weighted cost shares are 70.9 percent for materials, 23.2 percent for labor, 3.98 percent for machinery capital, and 1.84 percent for building capital, among plants the year before the MDP opening.

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 an MDP is associated with statistically insignificant increases in capital and construction expenditures. In most specifications the estimated impact of an 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. On the basis of these measures of public investment, it seems reasonable to conclude that public investment cannot explain the paper's results.

Incumbent plants may respond to the MDP opening by increasing the intensity of their capital usage. If winning counties had been depressed and the capital stock was used below capacity, then incumbent plants might increase production simply by operating their capital stock closer to capacity. As an indirect test of this possibility, we estimated whether the MDP 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. In models identical to the version of equation (8) used in table 6, we find small and insignificant changes in this measure. This finding suggests that greater capital capacity utilization is unlikely to be the source of estimated productivity spillovers.

The results could also be influenced by unobserved changes in labor quality, though the direction of this bias is unclear. If the MDP poaches good workers from incumbent plants, the quality of the workforce in existing plants may decline. If the MDP receives bad workers from incumbent plants or the opening attracts higher-quality workers to the county, then incumbent plants may upgrade the quality of their workforce. Since the specifications control for the number of hours worked by production and nonproduction workers but not for their quality, this would lead to an underestimate or overestimate of the true TFP change for incumbent plants.

D.   *Attrition of Sample Plants*

If the MDP increases competition for inputs and raises local input prices, as suggested by the estimated changes in quality-adjusted wages, this might encourage plants with declining TFP to close. Indeed, for a variety of reasons, 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 opening.

This attrition could result either 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.

The available evidence suggests that differential attrition is unlikely to explain the finding of spillovers in winning counties. Similar numbers of winning and losing plants remained in the sample at its end: 72 percent in winning counties and 68 percent in losing counties (i.e., the number of plants at $\tau = 5$ as a fraction of the number of plants at $\tau = 0$). The slightly larger attrition rate in losing counties is consistent with the paper's primary result. Specifically, one seemingly reasonable interpretation of this result is that the MDP opening allowed some winning county plants to remain open that would have otherwise closed. Thus to the extent that an MDP opening keeps weakening plants operating, the baseline analysis will underestimate the overall TFP increase.[36]

Within aggregate attrition numbers, the MDP might change the mix of existing firms. If the MDP increases input prices for all plants and disproportionately increases productivity for plants in some industries, then we might expect to see increased agglomeration of those plants in each MDP location. This is an intriguing hypothesis but is difficult to test in this setting because such attrition may occur only in the long run (more than 5 years). Within the same two-digit SIC as the MDP, 71 percent of incumbents in winning counties and 69 percent in losing counties remained in the sample at $\tau = 5$.

E.   Declining Plant TFP and Mismeasurement

Table 4 and figure 1 show that incumbent plant TFP was declining prior to the MDP opening in both winning and losing counties and in losing counties after the MDP opening. This finding is striking since productivity generally increases over time in the overall economy. Here we explore whether it affects the interpretation of the results and conclude that it does not.

The decline in TFP is not necessarily inconsistent with rising TFP in the overall economy. Recall that the sample is restricted to a set of large and aging manufacturing plants that appear continuously in the ASM in the 8 years prior to the MDP opening. Further, the specifications include both plant and industry by year fixed effects. The estimated decline in TFP from this specification and sample will miss the process

---

[36] We cannot reject the null hypothesis of equal trends in TFP prior to the MDP opening among plants leaving the sample in winning and losing counties; the TFP trend in winning counties minus the TFP trend in losing counties was −0.0052 (0.0080). Further, the estimation of eq. (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.

of creative destruction in which less productive plants are replaced by more productive plants.[37]

The estimated downward trend appears to be a general phenomenon for similar samples of plants in all U.S. counties and is not limited to winning and losing counties. From our analysis of randomly selected manufacturing plant openings in the United States (table 5, col. 5), the TFP of incumbent plants declined annually by 0.5 percent prior to the opening. Further, we find declining TFP among similar plants in all U.S. counties over the 6-year periods following the MDP openings. Specifically, we created a sample of all U.S. plants that appear in the ASM for 14 straight years, deflated output and materials by the consumer price index, and regressed log output on log capital stocks (building and machinery, created using the same permanent inventory method), log labor hours, log materials, plant fixed effects, industry fixed effects, and year fixed effects and weighted the regression by plant output. The estimated year effects report average changes in TFP over each year, and we calculated average changes in TFP over 6-year periods that correspond to the periods following MDP openings.[38] Over these periods, we find that TFP declined in all U.S. counties by an average of 4.7 percent (with a standard error of 0.4 percent). In other words, the pre-MDP opening TFP changes among large and aging plants in our sample of winning and losing counties (and post-MDP opening changes in losing counties) are similar to TFP changes among large and aging plants throughout the United States.

An alternative explanation is that measured declines in TFP are a statistical artifact that reflects measurement error, particularly in the construction of capital stocks. Following standard practice in the existing literature, we construct capital stocks based on depreciating plants' past inputs and adding deflated investments in new capital.[39] This procedure uses standard National Bureau of Economic Research depreciation rates, but if these rates are too low for firms in our sample, then aging firms will begin to have more measured capital than they have in reality. Mechanically, this will make firms' TFP appear to decline in firm age. Because the regressions control for industry by year and plant fixed effects, TFP changes are estimated solely on aging plants. Similar biases would appear if firms' labor or materials became unobservably worse as plants aged.

---

[37] Moreover, we note that our sample overlaps the late 1970s and early 1980s, which was a period of poor economic performance and low productivity. Foster, Haltiwanger, and Krizan (2000) have documented that within-plant productivity growth is cyclical and is particularly low during downturns.

[38] For example, if there was one MDP opening in 1987, the period 1987–93 received a weight of 1/47.

[39] This is necessary in the later portions of the sample, when book values for current capital are no longer reported.

Such measurement problems are unlikely to affect our main results of interest since this bias need not affect the relative comparison of firm TFP in winning and losing counties. For it to affect our estimates, measurement of capital stocks (or other inputs) would need to be systematically biased in winning counties after the MDP opening. The previous robustness checks provide some reassurance on this issue: column 3 of table 10 allows input effects to vary by industry and column 4 allows input effects to vary after the MDP opening or in winning counties (but not the interaction of those two). Further, the specifications in Appendix table A1 would be affected differently by measurement error in capital stocks, but all show TFP increases in winning counties after the MDP opening.

F.   Changes in the Price of Plant Output

Another concern is that the theoretically correct dependent variable is the *quantity* of output. However, because of the data limitations faced by virtually all 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 output prices instead of higher productivity.

We do not expect this to be a major factor in our context. The sample comprises manufacturing establishments that generally produce goods traded outside the county.[40] In the extreme case of a perfectly competitive industry that produces a nationally traded good, output prices would not increase disproportionately in a county that experienced increased demand.[41]

To explore this possibility further, we examine whether the productivity change is larger in industries that are more regional or more concentrated. We estimate a model 1 version of equation (8) that interacts $1(\text{Winner})_{pj}$, $1(\tau \geq 0)_{jt}$, and $[1(\text{Winner})_{pj} \times 1(\tau \geq 0)_{jt}]$ with incumbents' industry-specific measure of average distance traveled by output between production and consumption. We also estimate this regression

[40] To give an indication of the tradability of goods produced by firms in a given industry, we use data by detailed industry code on the average distance traveled by a good between production and consumption (Weiss 1972). Across all sample plants, the 10th centile is 239 miles, the 25th centile is 355 miles, the median is 466 miles, the 75th centile is 602 miles, and the 90th centile is 722 miles. This suggests that most establishments in our sample produce goods that are widely traded outside the county. Across all industries in the Weiss data, distance varies between 52 and 1,337 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.

[41] Similarly, input (labor) spillovers may be less pronounced for incumbent plants that produce nationally traded goods (Black, McKinnish, and Sanders 2005).

with a measure of incumbents' industry concentration.[42] These specifications do not find that estimated changes are larger in more local or more concentrated industries; in fact, there is some evidence for larger effects on incumbent plants that ship their products further.

Earlier estimates found that the estimated spillovers were not larger for incumbent industries that tend to ship products to the MDP industry, a context in which output price effects might be largest. Further, the opening of a nonmanufacturing MDP might have similar effects on demand for incumbent plants' output, particularly since these nonmanufacturing MDPs were in the retail and wholesale trade sectors. However, these nonmanufacturing MDPs did not lead to similar estimated TFP increases. These exercises suggest that output price increases are not the source of the estimated spillover effects.

VIII. Discussion and Implications for Policy

A. Discussion

The preferred model 2 estimates suggest that incumbent plants' TFP increased by 12 percent following the opening of an MDP, whereas model 1 estimates find a 5 percent increase. The 12 percent TFP increase implies an additional $430 million in annual county manufacturing output 5 years after the MDP opening. In this section, we discuss how to interpret this large effect and what it implies for the spatial distribution of economic activity.

To put the magnitude of the estimated spillover effect in perspective, we calculate the fraction of overall variation in average manufacturing productivity explained by the MDP opening. There is a tremendous amount of cross-sectional variation in productivity across U.S. counties in the manufacturing sector. For example, the county at the 90th percentile of the TFP distribution has average TFP that is 56 percent higher than that of the county at the 10th percentile, indicating that plants located in counties at the top of the distribution are 56 percent more productive than similar plants located in counties at the bottom of the distribution, holding constant all production inputs.[43] A 12 percent increase in TFP is equivalent to moving from the 10th percentile of the county-level TFP distribution to the 27th percentile; alternatively, it is

[42] The information on distance is from Weiss (1972). The information on industry concentration is from the Bureau of Census ("Concentration Ratios," 2002).

[43] Specifically, these numbers are obtained using cross-sectional plant-level data from the 1987 Census of Manufactures (the midpoint of our sample period). We regress log output on log inputs (log building capital, log machinery capital, log materials, and log labor) and a full set of county fixed effects. We then look at the distribution of the county fixed effects, which represent the average TFP among all manufacturing firms in a given county.

equivalent to a 0.6-standard-deviation increase in the distribution of county TFP. Attracting an MDP is a major event for these counties, and we find this implied shift in the relative standing of counties large but not unrealistic.

Our estimates have interesting implications for the distribution of economic activity across locations. Along with substantial increases in TFP, increased firm entry and expansion suggest that profits increased, at least in the short run. However, the documented increase in TFP does not translate necessarily as a similarly large increase in profits for incumbent firms. Increased economic activity generated by the MDP leads to firms bidding up local factor prices such as labor and land. Difference-in-difference estimates found that wage rates increased by 2.7 percent, compared to TFP increases of 4.8 percent in the difference-in-difference model 1 (we are unable to estimate a trend break model for skill-adjusted wages because we have to rely on decennial Census data). Since this is a countywide increase in labor costs, incumbent firms that do not receive a spillover may become less profitable.

If labor costs increase equally for all incumbents and the productivity gains are larger for plants that are more similar to the new plant, there might be long-run agglomeration of similar plants in each MDP location. This is important because it helps explain the existence of industrial clusters, a pervasive feature of the spatial distribution of economic activity. It seems unlikely, however, that industries receiving larger spillovers would fully agglomerate around MDPs. While there is a great deal of documented cross-sectional agglomeration and co-agglomeration, we expect that local wages rise by more than 2.7 percent for the particular workers demanded by such industries. Indeed, despite experiencing substantially larger TFP increases, there is no evidence of incumbent plants in winning counties and the same two-digit SIC staying open in the sample more than winning incumbent plants in other industries.

In interpreting the magnitude of our estimates, three points need to be highlighted. First, it is inappropriate to interpret the estimated increase in TFP as the partial equilibrium impact of the MDP opening, holding constant everything else in the county's economy. Instead, it reflects the impact of the plant opening and all other associated changes. For example, other new plants opened in the county following the MDP opening and overall manufacturing output increased (table 10). Consequently, the TFP estimates should be interpreted as a general equilibrium reduced-form effect that combines both the direct impact of the MDP and the impacts of subsequent new plants and expanded output from incumbent plants.

Second, the effect of an MDP opening is not representative of the typical plant opening. The MDPs differ from the average manufacturing plant in several respects, most importantly size. MDPs are significantly

larger than the average new plant in the United States. Moreover, they are a selected sample. Unlike most manufacturing plants, the MDPs generated bidding from local governments, presumably because there was an ex ante expectation of substantial positive spillovers. If spillovers vary by industry, then it may be important to note that MDPs tend to be in the automotive, chemical, computer, and electronics industries (relative to the average manufacturing plant opening).

Third, the counties bidding for plants may be those that would particularly benefit from a new manufacturing plant opening. We do not expect that winners and losers were great counties that almost attracted special plants; rather, they were counties willing to provide tax subsidies for industrial stimulus. In considering potential locations, the MDPs might be attracted to a declining manufacturing sector and the expectation of lower future wages.

These estimates of agglomeration spillovers come from a selected set of plants and set of counties for which we expect large spillovers, which implies that our estimates are a likely upper bound (and perhaps substantially so). This is an issue of external validity rather than the consistency of our estimates. However, our estimates are representative of the benefits generated by large plants bid on by these local governments, which is a population of interest for public policy. From a research standpoint, finding spillovers from MDPs appears to be a necessary condition for agglomeration spillovers from a broader set of plants (rather than a sufficient condition) and a call for further research.

B. Implications for Local Economic Development Policies

The presence of significant agglomeration externalities implies that the attraction of a new plant to a locality generates external productivity benefits for existing firms. An important question is whether in this context publicly financed subsidies to attract new plants are efficiency-enhancing. From the point of view of an individual locality, the presence of significant agglomeration externalities indicates that providing subsidies can internalize externalities and may increase efficiency in some cases. However, from the aggregate point of view, the efficiency of policy depends on whether the benefits of attracting a new plant for the receiving county are homogeneous.

Consider the case in which agglomeration spillovers are homogeneous. This could happen, for example, if the functional form for agglomeration economies is linear and productivity spillovers do not depend on economic distance between existing plants and the new plant. Assuming that the new plant will locate somewhere in the United States irrespective of the provision of subsidies, providing subsidies for plants to locate in a particular city, county, or region of the country is socially

wasteful from a national perspective (Glaeser and Gottlieb 2009). Further, even from a local perspective, the bidding for plants is likely to be a zero-sum game in which all the benefits are bid away. This special case forms the intuition for the conventional wisdom that the provision of incentives for firms to locate in particular locations is wasteful.

However, our results indicate that productivity spillovers vary on the basis of the economic distance between the industry of the new plant and the industrial composition of plants located in the county in advance of its opening (see table 8). This heterogeneity is important because when the benefits of attracting a new plant are heterogeneous, the socially efficient outcome is for a plant to locate where the sum of its profits and the spillovers are greatest. The new plant cannot capture these spillovers on its own and consequently might choose a location where its profits are high but the spillovers are minimal. In this case, payments to plants that generate spillovers can increase national welfare because they can cause plants to internalize the externalities in making their location decision. Further, from the point of view of the local government, heterogeneous spillovers imply that local governments may not bid away all the benefits (Greenstone and Moretti 2004).

There are at least two issues of incidence that bear noting. First, the payments from localities to plants are a one-for-one exchange of land rents from the former to the latter if the supply of land is inelastic. If the supply of land is elastic, the increase in land rents is not necessarily one-for-one (see Moretti, forthcoming).

Second, figure 2 demonstrated that there is substantial variability in the spillovers, and this could affect the provision and/or magnitude of subsidies. For example, the estimated impact is negative in 40 percent of the cases. Consequently, risk-averse local governments may be unwilling to provide tax incentives with this distribution of outcomes or only willing to bid less than the average spillover.

IX.   Conclusions

This paper makes three main contributions. First, the estimates document substantial increases in TFP among incumbent plants following the opening of Million Dollar Plants. This is consistent with firms agglomerating in certain localities, at least in part because they are more productive from being close to other firms.

Second, the estimates shed light on the channels that underlie the estimated spillovers. Estimated spillovers are larger between plants that share labor pools and similar technologies. This is consistent with intellectual externalities, to the extent that they occur among firms that use similar technologies or are embodied in workers who move between firms. Additionally, this finding is consistent with higher rates of TFP

due to improved efficiencies of worker-firm matches. Clearly, our evidence on the mechanisms is not conclusive. Further research is needed to understand in more detail the sources of agglomeration economies.

Third, firms appear to pay higher costs in order to receive these productivity spillovers. Spatial equilibrium requires that increases in TFP are accompanied by increases in local input prices, so that firms are indifferent across locations. The finding of higher prices for quality-adjusted labor is consistent with this prediction. The increased levels of economic activity reflect increased demand to locate in the winning county, which leads to higher local prices and a new equilibrium.

This paper has demonstrated that tests for the presence of spillovers can be conducted by directly measuring TFP. These tests can serve as an important complement to the measurement of co-agglomeration rates 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. Further, the significant heterogeneity in estimated spillovers across cases and the variation in estimates across industries underscore that there is still much to learn about the structural source of these spillovers.

Appendix

TABLE A1

CHANGES IN INCUMBENT PLANT PRODUCTIVITY FOLLOWING AN MDP OPENING (and Production Function Coefficients for Plant Inputs), ROBUSTNESS TO SPECIFICATIONS ADJUSTING FOR THE ENDOGENEITY OF PLANT INPUTS

| | BASELINE SPECIFICATION (1) | FIXED INPUT COST SHARES | | INVESTMENT-CAPITAL INTERACTIONS (4) | MATERIALS-CAPITAL INTERACTIONS (5) | MATERIALS-CAPITAL AND MATERIALS-LABOR INTERACTIONS (6) | IV INPUTS, LAGGED CHANGE (7) |
		Plant Level (2)	SIC-3 Level (3)				
				A. Model 1			
Mean shift	.0477**	.0364	.0325	.0391*	.0399*	.0354*	.0613**
	(.0231)	(.0228)	(.0241)	(.0222)	(.0216)	(.0208)	(.0250)
Building capital	.0384**	...	...	...	...	...	.0205
	(.0152)						(.0731)
Equipment capital	.0771***	...	...	...	...	...	.0856**
	(.0218)						(.0364)
Labor	.2915***	...	...	.2660***	.2891***	...	.2197***
	(.0290)			(.0249)	(.0270)		(.0448)
Materials	.4504***	...	...	.4284***	...	...	.4359***
	(.0218)			(.0279)			(.0436)
Observations	28,732	28,732	28,732	18,856	28,732	28,732	23,992

594

	B. Model 2						
After 5 years	.1203**	.0971	.0938*	.0839	.1004**	.0869*	.1135
	(.0517)	(.0656)	(.0538)	(.0527)	(.0487)	(.0480)	(.0695)
Building capital	.0393**	. . .	. . .	. . .	. . .	. . .	.0215
	(.0169)						(.0727)
Equipment capital	.0790***	. . .	. . .	. . .	. . .	. . .	.0899**
	(.0149)						(.0371)
Labor	.2925***	. . .	. . .	.2668***	.2903***	. . .	.2217***
	(.0284)			(.0245)	(.0262)		(.0448)
Materials	.4510***	. . .	. . .	.4300***	. . .	. . .	.4348***
	(.0217)			(.0280)			(.0438)
Observations	28,732	28,732	28,732	18,856	28,732	28,732	23,992

NOTE.—The table reports results from fitting several versions of eq. (8). Column 1 reports estimates from the baseline specification (col. 4 of table 5) and the estimated coefficients on each plant production input (logs). Column 2 reports estimates when fixing the coefficient on each plant's inputs to be its average cost share over the sample period, where per-period capital costs were calculated from capital rental rates using Bureau of Labor Statistics data. Column 3 reports estimates when fixing the coefficient on each plant's inputs to be the average cost share for all plants in its three-digit SIC level. Column 4 controls for a fourth-degree polynomial function of log capital and log investment (separately for both types of capital; see Olley and Pakes 1996). Column 5 includes the same controls as col. 4 but replaces log investment with log materials (see Levinsohn and Petrin 2003). Column 6 adds interactions between log materials and log labor to the controls in col. 5 (see Ackerberg et al. 2006). Column 7 instruments for each input with the lagged change in input from 2 years prior to 1 year prior, dropping the first two years of data for each plant (see Blundell and Bond 1998). Reported in parentheses are standard errors clustered at the county level.
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

595

References

Acemoglu, Daron. 1996. "A Microfoundation for Social Increasing Returns in Human Capital Accumulation." *Q.J.E.* 11 (August): 779–804.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2006. "Structural Identification of Production Functions." Manuscript, Dept. Econ., Univ. California, Los Angeles.

Becker, Randy, John Haltiwanger, Ron Jarmin, Shawn Klimek, and Daniel Wilson. 2005. "Micro and Macro Data Integration: The Case of Capital." Working Paper no. 05-02, Center Econ. Studies, Bur. Census, Washington, DC.

Black, Dan, Terra McKinnish, and Seth Sanders. 2005. "The Economic Impact of the Coal Boom and Bust." *Econ. J.* 115 (April): 449–76.

Bloom, Nick, Mark Schankerman, and John Van Reenen. 2007. "Identifying Technology Spillovers and Product Market Rivalry." Working Paper no. 13060 (April), NBER, Cambridge, MA.

Blundell, Richard, and Stephen Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *J. Econometrics* 87:115–43.

Card, David, Kevin F. Hallock, and Enrico Moretti. 2007. "The Geography of Giving: The Effect of Corporate Headquarters on Local Charities." Manuscript, Dept. Econ. Univ. California, Berkeley.

Chandra, Amitabh, and Eric Thompson. 2000. "Does Public Infrastructure Affect Economic Activity? Evidence from the Rural Interstate Highway System." *Regional Sci. and Urban Econ.* 30:457–90.

Chiang, Hyowook. 2004. "Learning by Doing, Worker Turnover, and Productivity Dynamics." Working Paper no. 593, Econometric Society, Far Eastern meeting, Seoul, July 2.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction.* Cambridge, MA: MIT Press.

Dumais, Guy, Glenn Ellison, and Edward L. Glaeser. 2002. "Geographic Concentration as a Dynamic Process." *Rev. Econ. and Statis.* 84 (May): 193–204.

Duranton, Giles, and Diego Puga. 2004. "Micro-foundations of Urban Agglomeration Economies." In *Handbook of Urban and Regional Economics*, vol. 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." *J.P.E.* 105 (October): 889–927.

———. 1999. "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *A.E.R. Papers and Proc.* 89 (May): 311–16.

Ellison, Glenn, Edward L. Glaeser, and William Kerr. Forthcoming. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *A.E.R.*

Foster, Lucia, John C. Haltiwanger, and C. J. Krizan. 2000. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." Manuscript, Bur. Census, Washington, DC.

Foster, Lucia, John C. Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *A.E.R.* 98 (March): 394–425.

Glaeser, Edward L. 1999. "Learning in Cities." *J. Urban Econ.* 46 (September): 254–77.

———. 2001. "The Economics of Location-Based Tax Incentives." Working Paper no. 1932, Harvard Inst. Econ. Res.

Glaeser, Edward L., and Joshua D. Gottlieb. 2008. "The Economics of Place-

Making Policies." Working Paper no. 14373 (October), NBER, Cambridge, MA.

———. 2009. "The Wealth of Cities: Agglomeration Economics and Spatial Equilibrium in the United States." *J. Econ. Literature* 47 (December): 983–1028.

Glaeser, Edward L., and Janet E. Kohlhase. 2003. "Cities, Regions and the Decline of Transport Costs." *Papers Regional Sci.* 83 (January): 197–228.

Greenstone, Michael, and Enrico Moretti. 2004. "Bidding for Industrial Plants: Does Winning a 'Million Dollar Plant' Increase Welfare?" Working Paper no. 9844 (July), NBER, Cambridge, MA.

Griliches, Zvi. 1958. "Research Cost and Social Returns: Hybrid Corn and Related Innovations." *J.P.E.* 66 (October): 419–31.

Griliches, Zvi, and Jacques Mairesse. 1995. "Production Functions: The Search for Identification." Working Paper no. 5067 (March), NBER, Cambridge, MA.

Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy.* Cambridge, MA: MIT Press.

Henderson, J. Vernon. 2003. "Marshall's Scale Economies." *J. Urban Econ.* 55 (January): 1–28.

Jaffe, Adam B. 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value." *A.E.R.* 76 (December): 984–1001.

Jaffe, Adam B., Manuel Trajtenberg and Rebecca M. Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citation." *Q.J.E.* 108 (August): 577–98.

Jovanovic, Boyan, and Rafael Rob. 1989. "The Growth and Diffusion of Knowledge." *Rev. Econ. Studies* 56 (October): 569–82.

Krugman, Paul. 1991*a. Geography and Trade.* Cambridge, MA: MIT Press.

———. 1991*b.* "Increasing Returns and Economic Geography." *J.P.E.* 99 (June): 483–99.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Rev. Econ. Studies* 70 (April): 317–42.

Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *J. Monetary Econ.* 22 (July): 3–42.

Marshall, Alfred. 1890. *Principles of Economics.* New York: Macmillan.

Moretti, Enrico. 2004*a.* "Estimating the External Return to Higher Education: Evidence from Cross-Sectional and Longitudinal Data." *J. Econometrics* 120 (July–August): 175–212.

———. 2004*b.* "Human Capital Externalities in Cities." In *Handbook of Urban and Regional Economics,* vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse. Amsterdam: North-Holland.

———. 2004*c.* "Workers' Education, Spillovers and Productivity: Evidence from Plant-Level Production Functions." *A.E.R.* 94 (June): 656–90.

———. Forthcoming. "Local Labor Markets." In *Handbook of Labor Economics,* edited by Orley Ashenfelter and David Card.

Olley, Steven G., and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64 (November): 1263–98.

Petrongolo, Barbara, and Christopher A. Pissarides. 2005. "Scale Effects in Markets with Search." *Econ. J.* 116 (January): 21–44.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *J.P.E.* 90 (December): 1257–78.

Saxenian, Anna Lee. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128.* Cambridge, MA: Harvard Univ. Press.

Syverson, Chad. 2004a. "Market Structure and Productivity: A Concrete Example." *J.P.E.* 112 (December): 1181–1222.

———. 2004b. "Product Substitutability and Productivity Dispersion." *Rev. Econ. and Statis.* 86 (May): 534–50.

van Biesebroeck, Johannes. 2004. "Robustness of Productivity Estimates." Working Paper no. 10303 (February), NBER, Cambridge, MA.

Weiss, Leonard W. 1972. "The Geographic Size of Markets in Manufacturing." *Rev. Econ. and Statis.* 54 (August): 245–57.

Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *A.E.R.* 88 (March): 290–306.