Workers' Education, Spillovers and Productivity: Evidence From Plant-Level Production Functions

Enrico Moretti*
UCLA and NBER

January 2004

Abstract

Using a unique firm-worker matched dataset, obtained by combining the Census of Manufacturers with the Census of Population, I assess the magnitude of spillovers from education in US cities by estimating plant level production functions. After controlling for a plant's own skill level, plant fixed effects, and industry-specific and state specific transitory shocks, I find that the output of plants located in cities that experience large increases in the share of college graduates rises more than the output of similar plants located in cities that experience small increases in the share of college graduates. I use three alternative measures of economic distance—input-output flows, technological specialization as measured by distribution of patents across technologies and frequency of patent citations—to test whether spillovers depend on economic distance. I find that spillovers between industries that are in the same city and are economically close are larger than spillovers between industries that are in the same city but are economically distant.

The estimated productivity differences between cities with high and low levels of human capital match remarkably well differences in labor costs between cities with high and low level of human capital. Consistent with a model that includes both standard general equilibrium forces and spillovers, the productivity gains generated by human capital spillover are offset by increased labor costs.

*I am indebted to Dan Ackerberg, Sandy Black, Lee Branstetter, David Card, Janet Currie, Robert Dekle, Paul Devereux, Bronwyn Hall, Chang-Tai Hsieh, Joe Hotz, Daniel Johnson, Phillip Leslie, David Levine, Darren Lubotsky, Marco Manacorda, Charles Mullin, Giovanni Peri, Brian Silverman, two anonymous referees and seminar participants at Berkeley, Rochester and University of Southern California for useful comments on this and previous drafts. The research in this paper was conducted while the author was a Census Bureau research associate at the California Census Research Data Center. Research results and conclusions expressed are those of the author and do not necessarily indicate concurrence by the Bureau of Census. This paper has been screened to insure that no confidential data is revealed. A previous version of this paper circulated under the title: "Human Capital Spillovers in Manufacturing: Evidence From Plant-Level Production Functions", CLE Working Paper 21, November 1999.

1 Introduction

Human capital externalities may arise if the presence of educated workers makes other workers more productive. Marshall (1890) is among the first to recognize that social interactions among workers create learning opportunities that enhance productivity. A growing theoretical literature has since then built on this idea and proposed models where human capital externalities are the main engine of economic growth. In an influential paper, Lucas (1988) argues that human capital externalities in the form of learning spillovers may be large enough to explain long-run income differences between rich and poor countries.

Empirical evidence indicates that spillovers may be important in some high-tech industries.[1] Yet, despite significant policy implications, there is little systematic empirical evidence on the magnitude of human capital spillovers. Only recently have some authors attempted to estimate the size of spillovers from education by comparing the wages of otherwise similar individuals who work in cities or states with different average levels of education (Rauch, 1993; Acemoglu and Angrist, 1999; Ciccone and Peri, 2002; Moretti, 2004).

In this paper, I take a more direct approach to the estimation of human capital externalities and focus on the productivity of manufacturing plants. The idea is quite simple. If externalities exist, we should see that plants located in cities with high levels of human capital can produce a greater output with the same inputs than otherwise similar plants located in cities with low levels of human capital. To test this hypothesis, I estimate plant-level production functions using a unique firm-worker matched dataset, obtained by combining the Census of Manufacturers with the Census of Population.

For each plant and city, I define the overall level of human capital in the city by calculating the fraction of college educated workers among all workers in the city outside the plant. After controlling for a plant's own human capital, I find that the productivity of plants located in cities that experience increases in the overall level of human capital rises more than the productivity of otherwise similar plants located in cities where the overall level of human capital is constant. The key econometric issue in comparing the productivity of plants across metropolitan areas with different overall levels of human capital is the possible presence of unobserved factors that raise productivity and attract a more skilled labor force to a city. It is possible that more productive plants are located in cities with a better educated labor force for reasons independent of human capital spillovers.

A benefit of using longitudinal, plant-level data is that I can deal with some of the most rele-

[1]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). The entry decisions of new biotechnology firms in a city depends on the stock of human capital of outstanding scientists there, as measured by the number of relevant academic publications (Zucker, Darby and Brewer 1998).

vant endogeneity and selectivity issues. By looking at changes over time, I control for permanent unobserved characteristics of plants and cities that might bias a simpler cross-sectional specification. It is still possible that time-varying productivity shocks are correlated with changes in the overall level of human capital in an area. For example, if Southern states that have low levels of productivity at the beginning of the period catch up, and this modernization process in turn attracts a better educated labor force to the South, then the estimated spillover will be too large. To lessen any fear that overall college share is correlated with time-varying unobserved factors, I control for state$\times$industry$\times$year effects. Identification comes by comparing changes over time in the productivity of plants that are in the same state and industry, but in different cities.

According to the most robust estimates, a one percent increase in the city share of college graduates is associated with a 0.5-0.6 percentage point increase in output. This estimate is remarkably robust across specifications. Different assumptions on technology, omitted variables, and variable definitions all yields similar results. Even after controlling for plant fixed effects, industry-specific transitory shocks and state-specific transitory shocks, it is still possible that part of the correlation between plants' productivity and aggregate human capital reflects changes in time-varying unobserved characteristics of cities. I cannot completely rule out this possibility, but I do provide several additional pieces of evidence to further investigate the validity of my conclusions.

I test whether the documented spillovers between two industries that are located in the same city and are *economically close* are larger than the spillovers between two industries that are located in the same city and are *economically distant*. Consistent with the view that measured spillovers represent the transmission of knowledge across related sectors, I find that spillovers generally decline with economic distance. For example, I find that aggregate human capital in the high-tech sector of the city matters more for high-tech plants than aggregate human capital in the low-tech sector of the city; and aggregate human capital in the low-tech sector matters more for low-tech plants than aggregate human capital in high-tech plants.[2]

I probe the relationship between economic distance and spillovers using three direct measures of economic distance. First, I use input-output tables and assume that the economic distance between manufacturing and other industries is proportional to the value of inputs that each industry provides to manufacturing. Second, I use an index of technological distance–first proposed by Jaffe (1986)—based on the distribution of patents across technological fields. According to this metric, two industries are close if the distribution of patents across technologies is similar. Third, I use a metric based on linkages revealed by patents citations (Jaffe

[2]Similarly, human capital in the same 2-digit industry has a larger effect than human capital in the entire manufacturing sector.

et al. 1993). According to this metric, an industry is close to manufacturing if manufacturing patents frequently cite that industry's patents. Using these three metrics, I find that the magnitude of the estimated spillover tends to decline with economic distance, although this relationship is by no means monotonic.

I provide several other specification tests of the estimated spillover effects. Most importantly, I test whether the stock of *physical* capital in a city outside a plant is associated with increased productivity in the plant. If my estimate of human capital spillovers are spurious, or attributable to agglomeration effects rather than human capital externalities, then I may find a similar "spillover" from physical capital. The results show no evidence of such physical capital spillover.

I also use an instrumental variable approach based on the fraction of large plant openings among all the plant openings in a city excluding the relevant 3-digit industry as an instrument for college share in other industries. Openings of large new plant are an important determinant of changes in the aggregate education level of manufacturing workers, explaining 11-18% of the changes in the fraction of college educated workers. Instrumental variable estimates are generally consistent with OLS estimates, although less precise.

In the last section of the paper, I assess the plausibility of the estimated spillover effect by comparing it to the difference in labor costs between cities with high and low levels of human capital. In equilibrium, if firms are really more productive in cities with high levels of human capital, we should observe proportionally higher wages in those cities. Otherwise, firms would relocate from cities with low human capital to cities with high human capital. I find that the estimated productivity differences generated by human capital spillovers are roughly offset by increased labor costs.

Overall, I cannot reject the existence of human capital spillovers in the U.S. manufacturing. However, because the stock of human capital grows slowly over time, the contribution of human capital spillovers to economic growth is not large. The most robust estimates in this paper indicate that human capital spillovers are responsible for an average of 0.1% increase in output per year during the 1980s.

The paper is organized as follows. In section 2, I present a simple general equilibrium framework with spillovers. In section 3, I describe the econometric specification adopted and I discuss the potential sources of bias. In Section 4, I describe the data. Sections 5, 6 and 7 present the empirical results. In section 8, I compare the estimated spillover effects with wage differences across cities. Section 9 concludes.

2 Equilibrium with Spillovers

In this section I present a simple general equilibrium framework to illustrate the nature of a spatial equilibrium in the presence of human capital spillovers. The model is adapted from a well known model by Roback (1982, 1988). The intuition is simple. Firms are more productive in cities with high overall levels of human capital, because of spillovers. In equilibrium, firms are indifferent between cities because wages are higher in cities with a higher overall level of human capital, so that unit costs are the same everywhere. Workers are indifferent because housing prices are higher in cities with a higher overall level of human capital. The model indicates that there are two ways to empirically measure human capital externalities: by comparing the output of firms located in cities with high and low levels of human capital; and by comparing the wages of workers located in cities with high and low levels of human capital. In this paper, I take the former approach. In section 8, I show that the estimated productivity differences between cities with high and low levels of human capital are consistent with observed wage differences between cities with high and low levels of human capital.

Consider two cities and two types of labor, educated and uneducated workers. There are two types of goods, a composite good y —nationally traded— and land h —locally traded. Each city is a competitive economy that produces y using a Cobb-Douglas technology: $y = AH^{\alpha_H} L^{\alpha_L} K^{\beta}$, where H and L are the hours worked by skilled and unskilled workers, respectively and K is capital. To introduce the possibility of human capital spillovers in the model, I allow the productivity of plants in a city to depend on the aggregate level of human capital in the city: $A = f(\overline{S})$. In the empirical part of the paper, for each firm and city, I measure $\overline{S}$ using the fraction of college educated workers in the city, outside the firm. In the absence of human capital spillovers from education, $\frac{\delta f}{\delta S} = 0$. In this case, productivity of a firm increases if more skilled workers are employed in the firm, but holding constant the firm's labor force, increases in the share of educated workers in the city have no effect on productivity. On the other hand, if the college share in a city generates positive human capital spillovers, a rise in college share raises productivity of all plants in the city: $\frac{\delta f}{\delta S} > 0$. Different mechanisms for human capital externalities have been proposed in the theoretical literature, and the model proposed here is consistent with most of these.[3]

[3]Marshall (1890) is often quoted as arguing that social interactions among workers in the same industry and location create learning opportunities that enhance productivity. More recently, an influential paper by Lucas (1988) focuses on the benefits associated with urban areas that come from firms acquiring ideas from their neighbors. In Lucas' words: "We know that there are group interactions that are central to individual productivity. [...] We know that this kind of external effect is common to all the arts and sciences." The external effect of human capital, Lucas adds, in not limited to art and science: "Much of economic life is creative in much the same way as is art and science". Lucas argues that long-run income differences across countries can be explained by human capital externalities in the form of learning spillovers. In other models of learning, individuals augment their human capital through exchanges of ideas in meetings with more skilled neighbors

Because the composite good, y, is traded, its price is the same everywhere. Variation in the cost of living depends only on variation in cost of land, p, which is the same for all workers in a city. Workers maximize utility subject to a budget constraint by choosing quantities of the composite good and residential land. Workers and firms are perfectly mobile, and profit are assumed to be zero. Equilibrium is obtained when the utilities of workers in all cities are equal and firms in different cities have equal unit costs.

The equilibrium for the simple case of only two cities, A and B, is described in Figure 1. The upward sloping lines in each panel represent indifference curves for the two education groups. Indirect utility of skilled and unskilled workers—$V_H(w_H, p)$ and $V_L(w_L, p)$, respectively—is a function of nominal wages and cost of land. The downward sloping lines show combinations of wages and rents that hold constant firms' unit costs: $C(w_H, w_L, p, r, \overline{S})$ =1, where r is the price of capital, which is assumed to be constant across cities. If human capital externalities exist, $\overline{S}$ enters the cost function. In cities with more human capital, firms can produce the same level of output with less labor and capital. In equilibrium, utility of workers is equalized across locations: $V_H(w_H, p) = k_H$ and $V_L(w_L, p) = k_L$ for educated and uneducated workers, respectively. A zero-profit condition for the firm assures that production must take place along the downward sloping curve. Thus the model has three equations (unit cost and indirect utility for each skill group) in three unknowns (w_H, w_L and p). Point 1 in the left panel of Figure 1 represents the equilibrium combination of wage of educated workers and cost of land in city A. Point 1 in the right panel represents the same combination for uneducated workers.[4]

Consider what happens if the share of college educated workers is higher in city B than in city A. For example, suppose that, because of technological differences, skilled workers are particularly productive in city B and demand for them is high. Skilled workers move to B, attracted by higher wages. Even without spillovers, wages are higher. Point 2 represents the equilibrium in city B if there are no spillovers.[5] If the spillover exists, then the isocost curve shifts further to the right. The magnitude of the spillover is the distance from 2 to 3.

In equilibrium, firms in city B are more productive than firms in city A. Since firms are free to relocate from A to B, why is productivity not driven to equality? Wages (and rent) are higher in city B, making firms indifferent between cities. If the cost of land is not very important for firms, the increased productivity in B relatively to A should be offset by increased labor costs

(Glaeser 1999, Jovanovic and Rob 1989). Acemoglu (1996) proposes an alternative model where human capital externalities arise even without learning externalities. The goal of this paper is to test whether spillovers are empirically relevant. Testing which of the explanations proposed in the theoretical literature is valid is beyond the scope of this paper.

[4]I follow Roback (1988) and take the level of utility k_H and k_L as parameters for simplicity. Closure of the model would require that the level of utility is made endogenous. This would complicate the model, without making it more insightful.

[5]In the absence of externalities, the wage of educated workers is higher in B because they are more productive. The wage of uneducated workers is higher because of complementarity (imperfect substitution).

in B relatively to A. I will come back to this point in Section 8, where I compare my estimates of the difference in productivity between cities with high and low human capital with existing differences in labor costs.

3 Econometric Framework

The model in Section 2 indicates that, if human capital spillovers exist, firms in cities with higher overall level of human capital $\overline{S}$ will be more productive. This paper estimates production functions to assess the magnitude of the productivity gains that are generated by human capital spillovers. The fundamental problem in estimating spillovers is the presence of unobservable factors that affect productivity and are correlated with the overall level of human capital across cities. It is possible that more productive firms are located in areas with higher levels of human capital for reasons independent of human capital externalities. I begin this section by introducing heterogeneity into the model described in the previous section. I then describe the econometric specification adopted, and discuss under what conditions human capital spillovers can be empirically identified.

3.1 Empirical Specification

To see the implications of unmeasured productivity shocks, assume, as before, that technology can be described by the following Cobb-Douglas production function:

$$y_{pjct} = A_{pjct} H_{pjct}^{\alpha_{Hj}} L_{pjct}^{\alpha_{Lj}} K_{pjct}^{\beta_j} \tag{1}$$

where y_{pjct} is output of plant p, belonging to industry j, in city c, and year t; j indexes 3-digit industries; H_{pjct} is the number of hours worked by skilled workers in the plant ; L_{pjct} is the number of hours worked by unskilled workers; K_{pjct} is capital. Assume that A_{pjct} is a function of the fraction of college-educated workers outside the firm in the same city. In addition, assume that productivity depends on various industry, city and time components:

$$\ln A_{pjct} = \gamma \overline{S}_{-jct} + \epsilon_p + \epsilon_j + \epsilon_t + \epsilon_c + \epsilon_{jt} + \epsilon_{ct} + \epsilon_{st} + \epsilon_{pjct} \tag{2}$$

where $\overline{S}_{-jct}$ is the share of college graduates among all manufacturing workers in city c with the exception of workers in industry j; and the ϵ's are unobserved productivity shocks at the plant, city, state, industry and year level (s indexes the state where city c is located). Equation 2 captures only spillovers that occur within a city across 3-digit industries. It does not capture potential spillovers that occur within a plant, which are likely to be internalized. Nor does it capture spillovers that occur between plants in the same 3-digit industry, because estimation of these types of spillovers is not empirically feasible due to data limitations. To the extent that spillovers between plants in the same 3-digit industry are large, estimates of γ are to be interpreted as a lower bound on the magnitude of total spillovers.[6] For now, equation 2 captures only spillovers generated within manufacturing. Later, I generalize this assumption and I include the college share in other industries.

In logs the production function becomes

$$\ln y_{pjct} = \gamma \overline{S}_{-jct} + \alpha_{Hj} \ln H_{pjct} + \alpha_{Lj} \ln L_{pjct} + \beta_j \ln K_{pjct} + \epsilon_p + \epsilon_j + \epsilon_t + \epsilon_c + \epsilon_{jt} + \epsilon_{ct} + \epsilon_{st} + \epsilon_{pjct} \quad (3)$$

The main concern in estimating the key coefficient γ is the presence of unobservable productivity shocks that are correlated with college share. Any positive correlation between the ϵ's and $\overline{S}_{-jct}$ will result in overestimates of γ.[7]

A major advantage of using a longitudinal plant-level dataset is that I am able to control for many permanent and time-varying factors that may affect both productivity and overall college share. Specifically, I estimate a production function that includes plant fixed effects (d_p), industry $\times$year effects (d_{jt}) and state$\times$year effects (d_{st}):

$$\ln y_{pjcst} = \gamma \overline{S}_{-jct} + \alpha_{Hj} \ln H_{pjct} + \alpha_{Lj} \ln L_{pjct} + \beta_j \ln K_{pjct} + d_p + d_{jt} + d_{st} + \epsilon_{ct} + \epsilon_{pjct} \quad (4)$$

Equation 4 is the basis of the empirical analysis in this paper. The coefficients on capital and labor are allowed to vary across industries, reflecting technological differences. Plant fixed effects fully absorb any permanent heterogeneity at the plant, city, or industry level (ϵ_p , ϵ_c and ϵ_j). Because of the inclusion of plant fixed effects, identification is based on changes over time in the external college share. State$\times$year effects absorb any state specific time varying shock that are shared by all plants in the same state (ϵ_{st}). Similarly, industry$\times$year absorb any industry specific time-varying shocks (ϵ_{jt}). In the most robust models, I also include industry$\times$state$\times$year effects (d_{jst}). In these models, identification comes from *changes over time*

[6] In theory, a more general specification would allow for spillovers between plants in the same 3-digit industry. This alternative specification would replace $\overline{S}_{-jct}$ in equation 2 with $\overline{S}_{-pjct}$, which is the college share in all manufacturing plants in city c with the exception of plant p. As it will be clear below, this is not feasible because of data limitations. Note, however, that the 3-digit industry classification is very detailed, and in many cases there is only one plant per city in each 3-digit industry, so that $\overline{S}_{-jct} = \overline{S}_{-pjct}$.

[7] For example, the term ϵ_p captures unmeasured plant characteristics that do not change over time, such as the quality of machines, patents, quality of management, and the culture within the firm; ϵ_c captures permanent city characteristics, such as public infrastructure, weather conditions, the presence of a research universities, and efficiency of local authorities; ϵ_t captures general trends in technology that affect all plants as well as variation in productivity over the business cycle; ϵ_j captures fixed industry characteristics; ϵ_{ct} is a time-specific shock that affects productivity of all plants in city c, irrespective of the industry, such as the opening of an airport, the construction of a rail link or a freeway; ϵ_{jt} captures industry and year specific shocks, such as the introduction of an industry-specific new technology; ϵ_{st} captures state and time specific shocks.

within a state and industry. To account for at least some time-varying city specific heterogeneity, in some specifications I control for city characteristics that are potentially correlated with college share, such as city population, unemployment rate, and racial composition.

The main source of heterogeneity that is not controlled for in equation 4 is the time-varying, city specific shock: ϵ_{ct}. The possible correlation between ϵ_{ct} and college share *in all other industries* in the same city, $\overline{S}_{-jct}$ is a concern. Note that correlation between ϵ_{ct} and college share *in the same industry*, $\overline{S}_{jct}$, would not, in itself, result in biased estimates.[8]

3.2 Threats to Identification

The key identifying assumption is that after controlling for plant effects, industry×state×year effects and time-varying city observable characteristics, the fraction of college graduates $\overline{S}_{-jct}$ is uncorrelated with unobserved city-wide shocks ϵ_{ct}. One example where this assumption is valid is if differences in residual changes in college share across cities are driven by changes in tastes of more educated people.

What are examples of situations that would violate my identifying assumption? The possibility that plants with better machines or better management are located in areas with higher college shares is unlikely to constitute a major problem here, as plant fixed effects absorb both plant and city specific permanent heterogeneity. Changes in local business cycle conditions could in theory affect both productivity changes and changes in college share. However, it seems unlikely that differences in the business cycle could constitute a major problem here, since most of the variation in business cycle is absorbed by the state×year and industry×year dummies. Furthermore, in some specifications, I control for state×industry×year dummies.

Many types of skill-biased technological shocks are also unlikely to violate my assumption. One example could be the introduction of computers. If computers are adopted equally by all firms, the productivity of skilled workers will increase in all industries and cities relative to the productivity of unskilled workers. Because I control for skilled and unskilled labor in each plant, this type of shocks should not bias my estimates.[9] Similarly, industry-wide skill-biased technological shocks are unlikely to pose a major threat to my empirical design, because I control for the distribution of skills in each plant and the coefficients on skilled labor and unskilled labor are allowed to vary by industry.[10] A similar conclusion holds in the case

[8] In theory one might think to absorb ϵ_{ct} with city×year effects. In practice, though, this is not feasible, because such a model would be almost completely saturated. Since 3-digit industries are small, most of the variation in $\overline{S}_{-jct}$ is at the city×year level.

[9] The only effect should be an increase in the coefficient on skilled labor. Empirically, I find that, consistent with a skilled bias technical change story, the coefficient on skilled workers does increase in 1990 relative to 1980.

[10] For example, a skill biased technology shock that affects the computer industry will results in a larger

of a skill biased technological shock that is specific to all plants in one industry in one city. For example, consider a shock to the computer industry in San Jose that does not affect the computer industry in San Francisco. If the shock is not transmitted to other industries in San Jose, this shock would not affect college share in other industries, and therefore would not result in biased estimates.[11]

In general, in order for a shock to induce spurious correlation in equation 4, the productivity shock must be (1) city-wide, (2) time-varying, and (3) must be correlated with college share across cities *within a state and industry*. One potential example of such shock is the opening of an airport or a freeway. If the new infrastructure raises the productivity of all existing manufacturing plants in the city, and at the same time attracts more college graduates to the city, estimates of γ will be too large. These kinds of shocks are not likely to be very important in my sample period.

Perhaps the most plausible source of bias is changes in the unobserved quality of workers. Although I control for the skill level of workers in each plant, it is in theory possible that workers of higher unmeasured ability move to cities that experience larger increases in college share. I discuss this issue in greater length in Section 7.2, providing some evidence that heterogeneity in workers quality is not in fact driving my results (see equation 7).

In sum, while I can not completely rule out the possibility that at least some of the estimated effect reflects city-wide, time-varying productivity shocks, it appears that many plausible sources of spurious correlation are accounted for. In Sections 6 and 7, I describe additional specification tests that may help in assessing the validity of my assumptions.

A final concern is that capital and labor inputs should in theory be treated as endogenous.

---

coefficient on skilled labor for plants in that industry. Furthermore, it is plausible that most computer producers across the nation or at least in California would be experiencing similar increases in productivity, so that industry×year×state effects would absorb most of this unobserved shock.

[11]On the other hand, a city-specific skill biased technological shock that is shared by all plants could be problematic. This could be the case, for example, if plants tend to borrow technologies from other industries in the same city. If this type of shock affects all manufacturing plants in some cities but not in other cities in the same state and also raise the aggregate stock of skilled workers in those cities, then my estimates would overstate the magnitude of human capital spillovers.

Another case where my strategy would fail to fully account for skill biased technological shocks is the case of endogenous technological change—where firms choose their technology based on the number of skilled workers in the city. Suppose that a new technology is introduced that raises the productivity of skilled workers, and that there are different intensities of adoption available to firms. Assume also that firms choose the intensity of adoption based both on the skill intensity in the firm *and* on the overall stock of skilled workers in their local economy. In particular, assume that, holding constant the skill intensity in the firm, firms located in cities with a larger fraction of skilled workers choose the version of the technology which has the largest effect on productivity. In this case, two identical firms, employing the same number of skilled and unskilled workers, would experience different unobserved productivity shocks, and these shocks would be proportional to the stock of skilled workers in each city. This scenario depends on the assumption that technology adoption depends not only on the skill intensity inside the firm, but also on the fraction of skilled workers outside the firm in the same city. Although there are theoretical models built on this or similar assumptions (Acemoglu 1996), I am not aware of any empirical study that investigates this hypothesis for US cities.

Unlike the usual case of estimation of production functions, here the focus is not on estimating the coefficients on capital and labor, but it is on estimating γ. Endogeneity of capital and labor is an issue only to the extent that it results in biased estimates of γ. Throughout the paper, I assume that, after controlling for plant effects, industry$\times$year effects and state$\times$year effects, endogeneity of capital and labor does not significantly bias estimates of γ. This assumption is potentially problematic.[12] To assess the sensitivity of my results to this assumption, in section 7.2 I directly measure Total Factor Productivity (TFP) and then explain changes in TFP as a function of changes in $\overline{S}$. This strategy does not involve estimating the production function, but it relies on the assumption that factor prices equal marginal products. Results are generally consistent with results from the main specification obtained by directly estimating the production function, suggesting that endogeneity of capital and labor does not introduce a large bias in the main specification.

4 Data

The data come from a unique match between plant records from the Census of Manufacturing in 1982 and 1992 and worker characteristics from the Census of Population. The Census of Manufacturers is a longitudinal dataset that covers the universe of manufacturing establishments with one paid employee or more. The unit of observation in the Census of Manufacturers is the plant.[13] Two important advantages of the Census of Manufacturers are its panel structure, and that it has a sample size large enough to allow a disaggregation of the data by metropolitan area.

Although the Census of Manufacturers contains detailed information on the number of hours worked in each plant, information on the education level of workers is not reported. To obtain data on workers' education, I match workers in the 1980 and 1990 Censuses of Population to firms in the Census of Manufacturers, by industry and city. Specifically, I assign each plant in the Census of Manufacturers and each worker in the Census of Population to a city-industry cell based on the metropolitan area code and a 3-digit industry definition. The 3-digit industry definition is quite detailed, so the cells are narrow. Examples of 3-digit industries include: iron and steel foundries (SIC 332); engines and turbines (SIC 351); electronic computing equipment (SIC 357); soaps and cosmetics (SIC 284). For each city-industry cell, I use the Census of Population to calculate the fraction of hours worked by individuals with college, some college, high school and less than high school. I combine this information with plant level information

[12] For example, the work by Davis and Haltiwanger (1999) suggests that industry-state shocks do not explain microfluctuations in labor.

[13] A company operating at more than one location is required to file a separate report for each location.

on the total number of hours worked from the Census of Manufacturers to impute the number of hours worked by each education group in each plant. This imputation strategy is similar to the one adopted by Hellerstein, Neumark and Troske (1999).

For city-industry cells for which there is only one plant, the matching is exact. In some cells, however, there is more than one plant. One example is "Motor Vehicles and Passenger Car Bodies" (SIC 371) in Detroit. For cells for which there are more than one plant, the imputation is based on the assumption that the fraction of hours worked by each education group is the same for all plants in the same cell. In the Detroit example, this assumption allows "Motor Vehicles and Passenger Car Bodies" plants in Detroit to have different number of hours worked, but requires the fraction of hours worked by college graduates, individuals with some college, high school graduates and high school drop out to be the same for all plants in that industry in Detroit. The Data Appendix provides a detailed description of the Census of Population and Census of Manufacturers and the matching algorithm. The matched sample is a balanced panel with 40,281 plants. Descriptive statistics for the matched sample are reported in Table 1.

In the preferred specification, I estimate equation 4 controlling for the number of hours worked by individuals belonging to two education groups: high school or less (L in equation 4); and some college or more (H in equation 4). Since plants in cities with a more educated labor force are more likely to employ educated workers, obtaining a good estimate of the skill distribution in each plant is particularly important. Failing to adequately control for the skill level of workers in the plant may result in an upward bias in the estimated spillover.

To assess whether this is an issue, I test whether my results are robust to a finer characterization of the education distribution of workers in the plant. For example, I control for hours worked by 3 education groups: high school or less, some college, college or more. In other models, I also exploit the information available in the Census of Manufacturers on the number of hours worked by production and non-production workers in each plant—basically hours worked by blue and white collar workers. These models separately control for hours worked by production workers belonging to 2 (in some cases 3) education groups and hours worked by non-production workers belonging to 2 (in some cases 3) education groups.[14] Empirically, I find that my estimates are not sensitive to different ways of accounting for the distribution of

[14] I impute hours worked in the plant by production and non-production workers belonging to different education groups using a strategy similar to the one just described. For each city-3-digit industry cell, I use the Census of Population to calculate the fraction of hours worked by production and non-production workers based on occupation. I assume that workers who in the Census of Population have blue-collar occupations are production workers, and workers who have white-collar occupation are non-production workers. I combine this information with plant level information on the total number of hours worked by production and non-production workers from the Census of Manufacturers to impute the number of hours worked by each education-occupation group in each plant.

human capital inside the plant.

As a way to check the reliability of the matched worker-firm data, I estimate plant-level wage equations. If the matching is correct and measurement error is not too large, I expect wage equation coefficients to be close to the ones usually found in the wage equation literature. I show in the Appendix that this in fact seems to be the case.

In interpreting the results I present below, it is important to bear in mind a limitation of the data. The longitudinal dataset that I use is not necessarily representative of the full population of plants, because it only includes plants that are observed both in 1982 and 1992. One consequence is that large plants are more likely to be in the sample. In some models, I re-weight the observations so that the distribution of plant size and other observable plant characteristics reproduces the distribution in the original population (see Section 7.2). Note, also, that by controlling for capital and labor inputs, all models effectively control for plant size.[15]

The key independent variable is college share in the city outside the industry, $\overline{S}_{-jct}$. I use the Census of Population to obtain an estimate of $\overline{S}_{-jct}$. An alternative specification would be to use average years of schooling instead of college share. There is no obvious a priori reason to choose one measure of aggregate human capital over the other.[16] I re-estimated all the models using average schooling instead of college share, and obtained results that are qualitatively similar to the one presented here.[17]

5 Estimates of Human Capital Spillovers

I now turn to the empirical results. As dependent variables, I can use either value of shipments or value added, which is value of shipments minus cost of materials. Previous literature suggests that neither measure is perfect.[18] I present results based on value added, but I have re-estimated all the models using value of shipments and obtained similar results. I report results based on value of shipments in Table 8 below.

[15] I have also looked at whether the probability that a plant exists in 1982 but not in 1992 is correlated with the change in the city college share. I find that this is not the case.

[16] In previous work I have used college share (Moretti 2004).

[17] It is important to note that spillovers may arise not only from the share of college graduates in an area, but also from their total number or their density. In this paper, I focus on spillovers that arise from the share of college graduates. I don't capture spillovers arising from density of human capital. When I control for changes in city population, my estimates do not change significantly.

[18] Hellerstein et al. (1999) point out that value added has two advantages over value of shipments. First, a value-added specification can be derived from polar production functions: one in which the elasticity of substitution between materials and value-added in infinite; and one in which this elasticity of substitution is zero. Second, a value of shipment specification requires one to include value of materials on the right hand side. This specification may be problematic given the potential endogeneity of materials.

Cross-Sectional Estimates: I begin by presenting cross-sectional estimates of plant-level production functions. Columns 1 and 2 in Table 2 refer to a specification where technology is Cobb-Douglas. The coefficient γ on college share outside the industry in columns 1 and 2 is 0.84 in 1992 and 0.81 in 1982, indicating that a one percentage point increase in the overall share of college graduates in the city (excluding the industry a plant belongs to) is associated with an increase in productivity by 0.8%. Through the paper, standard errors are corrected for city-year clustering.

The models control for capital stock, hours worked by skilled and unskilled labor, a dummy equal to one if the plant belongs to a multi-unit firm, and 3-digit industry dummies. Capital stock for equipment and structures is measured from the book values deflated by capital stock deflators.[19] Hours worked by unskilled workers are hours worked by workers who have a high-school degree or less. Hours worked by skilled workers are hours worked by college graduates or workers with some college. Columns 3 and 4 refer to a specification where technology is translog. The coefficient on college share is invariant to this change.[20]

Longitudinal Estimates: I now turn to longitudinal models. Table 3 reports estimates of variants of equation 4. The rows of the Table differ in the way the regressions control for the level of human capital of workers *within the firm.* Like in Table 2, models in row 1 control for hours worked by workers who have a high-school degree or less and for hours worked by workers with at least some college. Column 1 is analogous to the models in Table 2 but adds plant fixed effects. Identification of the spillover comes from *changes* in productivity and college share between 1982 and 1992. The coefficient on college share in column 1 is 0.74.[21] Plant fixed effects purge estimates of permanent plant and city unobserved heterogeneity. The fixed effects estimator may still be biased if there are transitory unobserved factors that affect both changes in college share and changes in productivity. In the specifications in columns 2, 3 and 4, I include respectively industry×year dummies, state×year dummies and industry×state×year dummies. The coefficients are between 0.51 and 0.77.

In the specification used in columns 1 to 4, the intercept of the production function is allowed to vary across plants, but the slope coefficients are constrained to be the same. In reality, however, it is possible that the relative importance of capital and labor varies across industries. In column 5, I relax the restriction that technology is the same across industries and allow the slope coefficients on capital and labor to vary by 2-digit industry. The coefficient

[19]Because capital enters in log, the deflator is fully absorbed when industry dummies are included.

[20]The coefficients on capital and labor appear to vary significantly between 1982 and 1992. I don't have a good explanation for this change.

[21]The coefficients on log capital, log skilled labor and log unskilled labor are, respectively: .185 (.005), .492 (.013), .384 (.011). R^2 is .95.

in column 5 is slightly smaller than the one in column 4.

In column 6 to 10, the assumption of Cobb-Douglas technology is relaxed and a more general Translog production function is estimated. The coefficient on college share is generally lower, but not statistically different from the one obtained from the corresponding Cobb-Douglas specification.

From the results in row 1, I conclude that estimates of the coefficient on college share outside the industry are generally robust to different specifications. After controlling for a plant's own level of human capital, plants located in areas where the overall level of human capital increased became more productive than similar plants located in areas where the overall level of human capital did not change. This increased productivity does not seem to be driven by industry specific or state specific shocks because it is robust to the inclusion of state×year and industry×year dummies.

According to the most robust estimate in columns 5 and 10, an increase of one percentage point in college share outside the industry is associated with a productivity increase equal to 0.6-0.7%. To help interpret the magnitude of the coefficient, consider that the average yearly increase in college share between 1982 and 1992 was about 0.2 percentage points. According to my estimate, an increase in college share of 0.2 percentage points would be associated with an increase in output by about 0.12-0.14%. For the average plant in the U.S., this amounts to about $10,000 per year. I discuss the magnitude of the estimated effect in Section 8.

A key question is whether variation in college share outside the plant's industry is proxying for variation in the education of workers *in the plant*. The specification adopted in row 1 controls for education of workers in the plant by conditioning on the imputed number of hours worked by employees with a high school degree or less and the number of hours worked by employees with some college or more. It is in theory possible that the characterization of the education distribution of workers in the plant based on these two education groups is not fine enough. In particular, it is possible that the educational achievement of workers inside the plant *within each education group* is not constant across cities. For example, the composition of the group of plant employees that I call "skilled"—those with some college or more—may differ across cities, and may be systematically correlated with the overall level of human capital in the city. In other words, the group of plant employees with some college or more could have relatively more college graduates than community college graduates in cities where aggregate college share outside the plant's industry is high. In this case, the estimates of the spillover presented in row 1 would be biased, because they would reflect the correlation between workers education in the plant and workers education outside the plant.

I test whether my estimates are sensitive to a finer characterization of the education distribution of workers in the plant. The model in row 2 controls for hours worked by 3 education groups: high school or less, some college, college or more. If the group of plant employees with some college or more has relatively more college graduates than community college graduates in cities with high aggregate human capital, then estimates in row 2 should be lower than estimates in row 1. I find that estimates in row 2 are slightly lower than estimates in row 1, but that the difference is not statistically significant.

In Panel B, I allow for an even finer characterization of the education distribution of workers in the plant by using the information available on production and non-production workers. Simply controlling for hours worked by production and non-production workers it is not enough to adequately control for human capital in the plant. Although non-production workers tend to have higher education than production workers, the correlation is by no means perfect. Doms, Dunne and Troske (1997) report that only 40% of non-production workers in the Census of Manufacturers have a college degree and more than 27% of production workers have a community college degree. For this reason, in panel B I control not only for the number of hours worked by production and non-production workers, but also for the imputed educational achievement of workers in the two groups. Specifically, models in row 2 control for imputed hours worked by production workers with a high school or less, imputed hours worked by production workers with some college or more, imputed hours worked by non-production workers with a high school degree or less, and imputed hours worked by non-production workers with some college or more. Models in row 4 push this specification even further by allowing for 3 education groups for production workers and 3 for non-production workers. Results in row 3 and 4 are generally consistent with those in Panel A.

Overall, I conclude that my estimates are not very sensitive to different ways to control for human capital of workers in the plant. In the remainder of the paper, I report results based on the most parsimonious specification of row 1 (two education groups inside the plant), although results do not change significantly when I use alternative specifications.

High-Tech vs. Low-Tech: As a first specification test, I test whether human capital spillovers matter more for the production of advanced, high-tech products (computers, scientific equipment, bio-tech, or pharmaceutical) than for the production of mature, low-tech products (cement, steel, or lumber). If I find that human capital spillovers were more important for cement plants than for computer or bio-tech plants, then that would cast doubt on the interpretation of the spillovers. More importantly, I test whether human capital in the high-tech sector of the city matters more for high-tech plants than human capital in the low-tech sector of the city; and whether human capital in the low-tech sector matters more for low-tech plants than human capital in high-tech. Just as I expect people in computers to benefit more from human capital spillovers, I expect them to benefit more from educated people in electronics than from educated people working in the textile sector.

The top panel in Table 4 reports estimates from a regression that includes both aggregate college share in the high-tech sector and aggregate college share in the low-tech sector (excluding the relevant 3-digit industry), separately for high-tech plants and low-tech plants. Entries in each panel come from one regression. For example, the entry in row 1, column 1 is the coefficient on college share in high-tech industries (outside relevant 3-digit industry) interacted with a dummy equal to one if the relevant plant is high-tech. To classify productions as high-tech or low-tech, I used the definition of high-tech industries provided by the American Electronic Association (1997) based on 45 4-digit SIC codes.[22]

The coefficient on high-tech college share for high-tech plants is 1.70, more than double the coefficient on low-tech college share for low-tech plants. This indicates that high-tech plants benefits from spillovers more than low-tech plants. The off-diagonal elements are smaller, indicating that aggregate human capital in high-tech industries has little effect on productivity in low-tech plants, and aggregate human capital in high-tech industries has little effect on productivity in low-tech plants. Estimates in the lower panel, based on a translog specification, are similar.

Spillovers at the 1-Digit and 2-Digit Industry Level: The estimates reported so far are a measure of the spillover generated by college share in the entire manufacturing sector in the relevant city (excluding the relevant 3-digit industry). I now refine the analysis by investigating how the magnitude of the estimated spillover varies when I consider a finer industry breakdown. In particular, I compare the effect of the share of college graduates in the city and 2-digit industry a plant belongs to (excluding the relevant 3-digit industry) with the effect of the share of college graduates in the entire manufacturing sector in the city (excluding the relevant 2-digit industry). Spillovers between plants that belong to similar industries should be larger than spillovers between plants that belong to industries that are different. Finding that the latter effect is larger than the former effect would cast doubt on the validity of my estimates.

Estimates in columns 1 and 2 of Table 5 indicate that the coefficient on the share of college graduates in the 2-digit industry a plant belongs to (excluding the relevant 3-digit industry) is about 0.95-1.00, or about 30% larger than the coefficient on the share of college graduates in the entire manufacturing sector (excluding the relevant 2-digit industry). Estimates in columns 3 and 4, based on a translog specification, yield a similar conclusion.

I try to push this exercise even further by experimenting with a measure of spillovers at

[22] The definition includes computers and office equipment, consumer electronics, communication equipment, electronic components, semiconductors, industrial electronics, photonics, defense electronics, electromedical equipment, software and computer related services, and telecommunication services. According to this definition, about 10% of the plants in the sample are high-tech.

the 3-digit industry level. In theory, spillovers should be larger when measured at the 3-digit level than when measured at the 2-digit or 1-digit level. However, this comparison is made difficult by data limitations. I cannot estimate a model that includes college share at the 3-digit industry level because, at that level of disaggregation, I cannot distinguish between education in the plant and outside the plant. Instead, I use the 3-digit share of non-production workers (excluding the relevant plant) as a proxy for the 3-digit share of college graduates outside the plant.[23] Estimates of models similar to the ones in columns 4 and 5 in Table 3 are 0.667 (0.478) and 0.746 (0.502), respectively. A direct comparison with estimates in Tables 3 and 5 is not possible, because the share of non-production workers is an imperfect proxy for the share of college educated. Although non-production workers do tend to be more educated than production workers, only 40% of non-production workers are college educated (Doms et al. 1997). Because of attenuation bias, the estimated parameters are lower than the parameters one would obtain if it were possible to estimate the same models substituting the 3-digit share of non-production workers with the 3-digit share of college graduates.

6 Do Human Capital Spillovers Decline with Economic and Technological Distance?

The findings on high-tech plants in Table 4 and two-digit industries in Table 5 provide a first piece of evidence that, within a city, the magnitude of the spillover depends on economic proximity. In this section, I use three alternative measures of economic distance to investigate more directly the relationship between economic distance and spillovers. Specifically, I investigate whether human capital spillovers within a city between industries that are economically close are larger than spillovers between industries that are economically distant. Finding that human capital spillovers are large between industries that are located in the same city but are economically distant would be surprising and would cast doubt on the validity of my results.

I modify equation 4 to include college share in manufacturing, as well as college share in Transportation; Communication and Utilities; Trade (retail and wholesale); Services; Finance, Real Estate and Insurance; Mining; and Construction:

$$\ln y_{pjcst} = \sum_{k} \gamma_k \overline{S}_{kct} + \alpha_{Hj} \ln H_{pjct} \alpha_{Lj} \ln L_{pjct} + \beta_j \ln K_{pjct} + d_p + d_{jt} + d_{st} + \epsilon_{ct} + \epsilon_{pjct} \quad (5)$$

where $\overline{S}_{kct}$ is now college share in industry k, city c and year t; and k indexes all 1-digit

[23] To facilitate the comparison with Table 3, I normalize share of non-production workers so that it has the same mean and standard deviation as share of college graduates.

industries (When $k = Manufacturing$, I calculate college share excluding 3-digit industry j). Equation 5 yields estimates of seven $\hat{\gamma}$'s—one for each 1-digit industry. Once I have estimates of the γ's, I can test whether the magnitude of each industry's γ coefficient depends on the economic distance between that industry and manufacturing. Because 1-digit industries are very broad, I also repeat the analysis at the 2-digit industry level.

The three measures of distance that I use capture alternative but not mutually exclusive notions of economic and technological distance between industries. The first measure is based on input-output tables and tries to captures interactions between industries that arise from exchanging goods and services during the production process. According to this metric, the economic distance between manufacturing and each of the 1-digit industries is proportional to the value of inputs that each industry provides to manufacturing. For example, the industry "transportation, communication and utilities" is closer to manufacturing than "finance" because the value of inputs from "transportation, communication and utilities" that are used in manufacturing is larger than the value of inputs from "finance." One limitation of the input-output metric is that is may confound human capital spillovers with pecuniary externalities. The literature on R&D spillovers has preferred measures of technological distance based on patents.[24]

The second measure of distance tries to capture similarities in the distribution of R&D investment and technological expertise across different technical fields, as measured by the number of patents in each field. The US patent and Trademark Office has developed a highly elaborate classification system for technologies to which patented invention belong. By counting the number of patents held by an industry in a technological field, I can obtain a quantitative measure of the industry's level of expertise in that field (Jaffe 1986, Branstetter 2001). According to this metric, two industries are close if the distribution of patents across technological fields is similar.

As a third metric, I use an index based on industry linkages revealed by patents citations. Patent citations serve an important legal function, since they delimit the scope of the property rights awarded by the patent. Thus, if patent B cites patent A, it implies that patent A represents a piece of previously existing knowledge upon which patent B builds. The presumption is that citations are informative of links between patented innovations. The third index of distance is based on the notion that if industry x cites industry's y patents more frequently than industry's z patents, x is closer to y than to z. Patent citations have been used by other authors to document spillovers.[25]

---

[24] See for example Jaffe (1986), Branstetter (2001), Jaffe et al. (1993), Jaffe, Trajtenberg and Henderson (2002).

[25] For example, in an influential paper, Jaffe et al. (1993) compare the geographic location of patent citations with that of the cited patents to measure the extent to which knowledge spillovers are geographically localized.

Input-Output Tables. I rank non-manufacturing industries by distance from manufacturing based on the value of the inputs that each industry provides to manufacturing. The value of inputs provided by each industry to manufacturing is shown in column 2 of Table A1 (top panel).[26] The third column in Table A1 (top panel) shows estimates of the γ_k coefficients in equation 5. The coefficient is largest for Manufacturing and smallest for Finance, Real Estate and Insurance. Although the relationship between the coefficient in column 3 and economic distance in column 2 is by no means monotonic, the estimated coefficients do tend to decrease as we move toward industries that provide fewer inputs into manufacturing.

The negative relationship between the estimated coefficient and economic distance can be better seen in Figure 2 (top panel), which plots the coefficients against the rank based on value of inputs. The OLS fitted line is superimposed. The OLS slope (standard error) is -.109 (.042), and R^2 is 0.54. (A similar figure is obtained if one plots the coefficients against the log value of inputs.) Figure 2 indicates that human capital in industries that are economically close to manufacturing (and presumably interact more with manufacturing) benefits manufacturing plants more than human capital in industries that are economically far from manufacturing (and presumably interact less with manufacturing).

This finding is based on 1-digit industries. I repeat the analysis using a more disaggregated industry definition. Instead of looking at 1-digit industries, I look at 2-digit industries. Everything else remains the same. As before, I rank industries based on value of inputs. Column 3 in the bottom panel of Table A1 shows estimates of a model where the coefficient on college share varies depending on the distance between 2-digit industries. Because there are so many 2-digit industries, I group them in sets of five. In other words, I force the coefficient on college share to be the same for the closest 5 industries, the next five and so on. Note that the industry composition in each 5-industry group is different for each plant. For example, the entry in column 3, row 1 is the coefficient on college share in the 5 2-digit industries that are closest to the relevant plant.

Estimates of the γ_k coefficients show a tendency to decrease as we move from close industries to industries further away. The negative relationship between estimated coefficients and economic distance is more easily seen in Figure 2 (bottom panel), that plots the estimated coefficients against the rank based on value of inputs. The OLS fitted line is superimposed. The slope (standard error) of the line is -0.015 (0.002). R^2 is .85.

See also Jaffe et al. (2002).

[26]The I-O tables are based on national data. I use the "Use" Table, which shows the inputs to industry production and the commodities that are consumed by final users. The Use table is the most frequently requested table because of its applications to the estimates of GDP. Source: $www.bea.gov/bea/industry/iotables/prod/table_list.cfm?anon = 394$.

Distribution of Patents Across Technological Groups. I now repeat the analysis using a measure of economic distance based on the distribution of patents over technological fields. I first divide the patents into 36 technological fields defined in Jaffe et al. (2002) (pp. 452-454). For each industry j, I construct the vector of shares of industry patents in each technological field $s_j = (s_{j1}, s_{j2}, ..., s_{j36})$. For each pair of industries (j,k), I calculate the uncentered correlation coefficient between vector s_j and s_k as follows:

$$\rho_{jk} = \frac{\sum_{h=1}^{36}(s_{jh}s_{kh})}{\sqrt{\sum_{h=1}^{36} s_{jh}^2 \sum_{h=1}^{36} s_{kh}^2}} \tag{6}$$

The uncentered correlation is the angular distance between vectors: two industries with identical distribution of patents across technological fields have a correlation of one, two industries with orthogonal distributions of patents have a correlation of zero. See the Data Appendix for details on the patent dataset.

Column 2 in Table A2 shows the uncentered correlation coefficients ρ at the 1-digit industry level (top panel) and 2-digit industry level (bottom panel), and column 3 shows the corresponding estimates of the γ_k coefficients in equation 5.[27] The relationship between the coefficient in column 3 and economic distance in column 2 is not monotonic, but the estimated coefficients do tend to decrease as we move toward pairs of industries with lower ρ. The negative relationship between estimated coefficients and economic distance is more easily seen in Figure 3, where I plot the estimated coefficients against the rank based on ρ. In the top panel (1-digit industry level), the slope of the fitted line is -0.095 (0.048). R^2 is .43. In the bottom panel (2-digit industry level), the slope of the fitted line is -0.016 (0.005). R^2 is .65.

Patent Citations. Finally, I repeat the analysis using a measure of economic distance based on the frequency of patent citations. For each pair of industries (j,k), I calculate the frequency that a patent assigned to industry j cites a patent assigned to industry k.[28] Column 2 in Table A3 shows the frequency of citations at the 1-digit industry level (top panel) and 2-digit industry level (bottom panel). For example, the first and second entry in column 2 show that manufacturing patents cite manufacturing patents and services patents with frequency equal 74% and 13%, respectively.[29] Column 3 shows the corresponding estimates of the γ_k

[27] While the 1-digit γ_k coefficients are the same as in table A1, the 2-digits are not, because I group 2-digit industries in groups of 5. Since this grouping depends on the specific measure of distance used, the groups are obviously different in Table A1 and A2.

[28] Unlike uncentered correlation, this index of distance is not symmetric, because the frequency that a patent assigned to industry j cites a patent assigned to industry k is not the same as the frequency that a patent assigned to k cites a patent assigned to j. In calculating the index, I do not include self-citations.

[29] Frequencies in the bottom panel (column 2) are lower because 2-digit industries are smaller than 1-digit industries. For example, the first entry in the bottom panel is the average frequency of citations for the 5 2-digit

coefficients.

The relationship between estimated coefficients and economic distance is shown in Figure 4, where I plot the estimated coefficients against the rank based on frequency of citations. In the top panel (1-digit industry level), the slope of the fitted line is -0.086 (0.051). R^2 is .35. In the bottom panel (2-digit industry level), the slope of the fitted line is -0.007 (0.004). Neither slope is statistically different from zero. R^2 is .29. The relationship between spillovers and distance as measured by patent citations appears to be somewhat weaker than the relationship between spillovers and distance as measured by the uncentered correlation coefficient.

7 Additional Results

The results presented in Table 3 are generally consistent with the notion that changes in the aggregate stock of human capital are associated with increased productivity of manufacturing plants. Yet, without a randomized experiment, it is difficult to be completely certain that the estimated parameters are causal. It is always possible that the estimates reflect, at least in part, the presence of city wide, time varying productivity shocks correlated with $\overline{S}$. However, findings in Table 3 show that the estimates of the spillover are robust to a wide variety of assumptions on technology and demand shocks. Moreover, results in Section 6—based on three alternative measures of economic distance—as well as results in Tables 4 and 5, indicate that, within a city, the magnitude of the spillovers decline with economic distance. These results are consistent with the interpretation of my estimates as human capital spillovers. If the documented correlation between college share and productivity were completely spurious, one would not expect to find such a consistent pattern based on economic distance.

In this section, I present three additional pieces of evidence to further investigate the validity of my estimates. I begin by presenting a specification checks based on physical capital. In subsection 7.2, I present estimates from a number of alternative specifications intended to probe the robustness of the results in Table 3. In subsection 7.3, I experiment with an instrumental variable strategy. Taken together, results in this section lend further support to the view that the estimates of the spillovers are not completely spurious.

7.1 A Specification Check

As a specification check, I estimate equation 4 substituting human capital with a measure of overall *physical* capital outside a plant. If my estimate of human capital spillovers are spurious,

industries that receive the most citations by the relevant 2-digit industry. Because entries are an average for 5 industries, they don't sum up to one.

or if they can be explained by agglomeration effects other than human capital externalities, then I may find that plants located in cities where the overall level of physical capital is high are more productive than similar plants located in cities where the overall level of physical capital is low. On the contrary, if my estimates are capturing *only* human capital externalities, there is no reason why physical capital in one plant should be correlated with productivity in other plants.

For each plant and city, I use two alternative measures of density of physical capital: the log of average physical capital outside the plant in the city and the log of per worker average physical capital outside the plant in the city. Cross-sectional estimates in Table 6 suggest that average capital is correlated with productivity, although the sign is positive in 1982 and negative in 1992. However, when plant fixed effects are included the coefficient becomes insignificant, suggesting that plant-level heterogeneity may bias cross-sectional estimates. When state×year effects are added (column 4) or industry×year effects are added (column 5), the coefficients drop to virtually zero. I conclude that overall level of physical capital outside the plant does not have an effect on plants' productivity similar to the one generated by human capital.

7.2 Robustness Checks

In this subsection, I investigate the robustness of the estimates in Table 3 to different assumptions.

Regional vs. National Industries One of the assumptions of the model is that the price of output is constant across locations. This is probably a reasonable assumption for many manufactured goods, because they are traded on the national market. However, some manufactured goods have a more regional distribution, and the assumption of one national price may not be realistic for them.[30] The concern is that the output price of regional industries reflects local production costs, and locations with higher production costs also have higher college share. In this case, my estimates would be biased upward.

I test whether the estimated spillover is different for regional industries and national industries. I define regional industries based on whether the average distance travelled by output is less than 500 (or 300) miles.[31] Examples of regional industries are: hydraulic cement, iron and steel products, metal scrap and waste tailings, icecream and related frozen desserts, prepared feed for animals, and prefabricated wooden buildings. I find that the estimated spillover is lower for regional industries, although the difference is not statistically significant. This finding

[30] Roberts and Supina (1997) find considerable price dispersion in Census of Manufacturing data across a range of industries.

[31] The information on distance is from the Appendix in Weiss (1972). Distance varies between 52 and 1337, with a mean of 498.

suggests that unobserved differences in the costs of production are unlikely to introduce upper bias in my estimates.[32]

Estimates Based on TFP I now turn to estimates of the spillover based on a Total Factor Productivity (TFP) specification. First, I estimate TFP under the assumptions that (1) technology is Cobb-Douglas; (2) factor prices equal marginal products; (3) there are constant return to scale to capital and labor. The labor elasticity is measured at the plant level as the plant-specific ratio of total wages over total output. Having estimated TFP, I then regress TFP on college share in other industries. The advantage of the TFP specification relative to the specifications in Table 3 is that it does not require estimating the production function, and therefore it does not rely on the assumption that capital and labor inputs are exogenous. The disadvantage is that estimates of TFP rely on the assumption that factor prices are paid their marginal product, and that there are constant return to scale.[33]

Estimates in Table 7 show that the estimated spillover varies between 0.255 and 0.693. The most robust specification, in column 4, is not significantly different from the corresponding specification in Table 3, although the standard error is larger.

More Robustness Checks I conclude this sub-section by presenting estimates from a number of alternative specifications intended to probe the robustness of the results in Table 3. The first row in Table 8 reproduces the estimate for the base specification in Table 3, column 1, row 1. The remaining rows present estimates of variants of the base model. The second row reports the estimate from a specification similar to the base specification, where the dependent variable is value of shipments, not value added. The coefficient increases to 0.86.

One concern is that college share is picking up not only human capital spillovers, but also university spillovers, since the density of universities is correlated with college share. To investigate this possibility, I have re-estimated my models controlling for the number of colleges and universities in each city.[34] Variation in the number of institutions between 1982 and 1992 is limited, so the results in row 3 are quite similar to the base case estimates. I have also re-estimated my models controlling for the total number of college degrees awarded (row 4) and

[32]The parameters on external college share and on external college share $\times$ the regional dummy are, respectively, 1.15 (0.39) and -0.52 (0.45) in Cobb-Douglas models that include establishments effects and industry$\times$state$\times$year effects, when the regional dummy is equal to 1 if the plant belongs to an industry where the average distance travelled by output is less than 500 miles. When the regional dummy is equal to 1 if the plant belongs to an industry where the distance travelled by output is less than 300 miles, the corresponding parameters are 0.91 (0.21) and -1.31 (0.92). Translog models yield similar estimates. One possible explanation for the fact the coefficient is lower for regional industries is that regional industries are mostly low-tech, and low-tech industries seems to enjoy lower spillovers.

[33]Under constant returns to scale to K and L, the capital elasticity is simply one minus the labor elasticity. This assumption in useful because, while I observe the capital stock, I don't observe capital elasticity or the rental price of capital *at the plant level.*

[34]Data on colleges and universities are from CASPAR, which is made available by the NSF.

for both the number of universities and the number of degrees awarded (row 5). My estimates are not very sensitive to these additional controls.[35]

A limitation of the data is that capital stock is imputed for plants that are not part of the Annual Survey of Manufacturers in 1982. This is a concern, because it could imply that my models do not adequately control for capital stock. To address this concern, I have re-estimated my models dropping ASM plants. Estimates in row 6 based on non-ASM plants are similar to my main estimates, indicating that my results are not very sensitive to the imputation.

A second data limitation is that output in the computer industry is not easily measured. The BLS output deflator use a hedonic approach and shows a steep fall in recent years. In row 7, I show that when I re-estimate my models excluding plants belonging to the computer industry (electronic computers, SIC 3571; computer terminals, SIC 3572; computer peripheral equipment, SIC 3577), my estimates do not change significantly.

In section 4, I pointed out that results in this paper are based on a selected sample of plants that are observed both in 1982 and 1992. Plants in the selected sample are larger than plants in the population. In row 8, I re-weight the sample to make the plant size distribution look like the distribution in the 1982 population. I assign weights based on plant size: smaller plants receive more weight than larger plants.[36] After the re-weighting, both plant size and other observable characteristics of plants are similar to those of the population in 1982. The coefficient from the weighted regression is 0.694 (row 8).[37] In the next three rows, I run separate regressions based on plant size. (These regressions are not weighted). No clear pattern emerges. For small plants (less than 10 workers), the coefficient is 0.69. It increases to 0.81 for medium sized plants (between 11 and 50 workers), and decreases to 0.75 for large plants (above 50 workers).

Next, I try to address the concern that changes in workers' unobserved ability are correlated with changes in college share. It is in theory possible that workers of higher ability move to cities that experience larger increases in college share. If this is the case, the estimated spillover would reflect higher ability of educated workers in the plant, not higher productivity. By imposing some additional assumptions, it is possible to account for workers heterogeneity in models where the coefficients on skilled and unskilled labor are allowed to vary across cities and time. To see

[35]Including additional controls has little effect on the coefficients. For example, estimating the model in row 5 conditioning on industry×year effects, state×year effects or industry×state×year effects yields, respectively: .504 (.197), .721 (.236), .755 (.240).

[36]I divide the sample into 10 groups, based on the number of workers: less than 10, between 10 and 20, 20-30, etc. I assign a weight to each plant in the longitudinal sample based on the frequency of that plant's group in the 1982 Census of Manufacturers population.

[37]To further investigate the issue of sample selection, I have also tested whether the probability that a plant exists in 1982 but not in 1992 is correlated with changes in the level of human capital in a city, and I found little correlation. Specifically, I divide the cities in the sample into 4 quartiles, according to the change in college share between 1982 and 1992. The average probability that a plant exists in 1982 and not in 1992 for the first quartile, which is the group of cities with the smallest increase in college share is 0.511. The corresponding figures for the second, third and fourth quartile are, respectively, 0.515, 0.512 and 0.544.

this, assume that the production function is $y_{pjct} = A_{pjct}H^{*\alpha_H}_{pjct}L^{*\alpha_L}_{pjct}K^{\beta}_{pjct}$, where H^*_{pjct} and L^*_{pjct} are the true but unobserved skilled and unskilled labor inputs, respectively. Unlike equation 4, here α_H, α_L and β are assumed to be constant across industries. Assume that the true labor inputs are equal to hours worked inflated by an ability coefficient that can vary across cities and over time: $H^*_{pjct} = H^{\theta_{Hct}}_{pjct}$, where H is hours worked (which are observed), and θ_{Hct} is average ability of skilled workers in city c at time t (unobserved). A larger θ_{Hct} implies higher ability. Similarly, $L^*_{pjct} = L^{\theta_{Lct}}_{pjct}$, where θ_{Lct} is average ability of unskilled workers. When estimating a production function that includes H and L (not H^* and L^*), the concern is that changes in unobserved ability, θ_{Hct} or θ_{Lct}, are correlated with changes in $\overline{S}_{-jct}$. The production function becomes $\ln y_{pjcst} = \gamma\overline{S}_{-jct} + (\alpha_H\theta_{Hct})\ln H_{pjct} + (\alpha_L\theta_{Lct})\ln L_{pjct} + \beta\ln K_{pjct} + d_p + d_{jt} + d_{st} + \epsilon_{ct} + \epsilon_{pjct}$. Under these assumptions, one way to account for worker heterogeneity, is to estimate models where the coefficients on skilled and unskilled labor vary by city and over time:

$$\ln y_{pjcst} = \gamma\overline{S}_{-jct} + \alpha'_{Hct}\ln H_{pjct} + \alpha'_{Lct}\ln L_{pjct} + \beta\ln K_{pjct} + d_p + d_{jt} + d_{st} + \epsilon_{ct} + \epsilon_{pjct} \qquad (7)$$

where $\alpha'_{Hct} = \alpha_H\theta_{Hct}$ and $\alpha'_{Lct} = \alpha_L\theta_{Lct}$.

Note that one can think of the θs not only as unobserved ability, but also as unobserved skill-biased, city-specific technological shocks. In this case, one can interpret H^* and L^* as effective labor inputs, i.e. hours worked by skilled and unskilled workers inflated by a technology coefficient that allows workers in a given skill group to be more productive in some cities than in other cities. Row 12 reports an estimate of equation 7. The coefficient is 0.50, lower than the estimate of the more restrictive model where the coefficients on skilled and unskilled labor do not vary across cities, but still positive.[38]

The model in row 13 controls for city density. The coefficient of interest does not change significantly. Including population, unemployment rate, percent black, percent immigrant and percent female also has little effect (row 14).[39]

As a last specification check, in the last two rows, I test whether the magnitude of the estimated spillover varies by multi-unit status. Multi-unit establishments are plants that are part of larger firms with establishments in more than one location. Henderson (2001) argues that single-unit plants should be more sensitive to the characteristics of their local environment than plants that belong to large firms with establishments in several locations. According to this view, plants that belong to multi-establishment firms depend more on internal-firm networks

[38]This finding is consistent with my previous work that uses longitudinal worker level data to address the issue of unobserved worker quality (Moretti 2004).

[39]One reason why it is important to control for population is that larger cities may make firms more productive because they allow for more subcontracting. If return to specialization are important, it is in theory possible that plants in larger cities are more productive.

and therefore are more insulated from local external environments than single-unit plants. For example, while many of the factors that affect the productivity of a General Motors' plant located in St. Louis are probably determined in the General Motors' headquarters in Detroit, all the factors that affect the productivity of a single-unit plant in St. Louis are determined in St. Louis. Spillovers should therefore be larger for single-unit plants that multi-unit plants. I find that the coefficient on college share is 0.91 (0.17) for single-unit plants and only 0.42 (0.32) for multi-unit plants.

7.3 Instrumental Variable Estimates

In this subsection I try to further investigate the validity of my estimates by using an instrumental variable approach. A valid instrument is correlated with changes in $\overline{S}$ in other industries and is orthogonal to unobserved productivity shocks. I propose an instrument based on large plant openings. Specifically, the instrumental variable is the fraction of large plant openings among all the plant openings in a city excluding the relevant 3 digit industry. Large plant openings are defined as plants that exist in 1992 and did not exist in 1982 and that have 1000 or more employees (in some models, I try 500+). Column 1 and 2 in Table 9 show the number of new plants in 1992 and the total employment in these plants, by size. Although large plants are only a small fraction of new plants, they account for 18% of employment generated by new plants.

Large plants have a higher share of skilled workers. The correlation between total employment and share of non-production workers at the plant level is 0.09. Openings of large plants appear to be an important determinant of changes in the aggregate education level of manufacturing workers. Column 3 in Table 9 reports the correlation between the fraction of new plants in a given size group among all new plants in a city (excluding the relevant 3-digit industry) and the 1982-1992 change in manufacturing college share in the city (excluding the relevant 3-digit industry). This correlation is calculated for the sample of 40,281 plants that is used in all the models in this paper. Entries in column 3 suggest that plant opening have a differential impact on aggregate human capital depending on the size of the new plant. While the fraction of small and medium-sized plants is not positively correlated with college share in other industries, the fraction of plants with at least 500 workers and the fraction of plants with at least 1000 workers are positively correlated with college share in other industries.[40] Column

[40]The ten cities with the largest fraction of new plants with at least 1000 workers are: Lafayette, IN; Pine Bluff, AR; Bloomington, IL; Trenton, NJ; Wilmington, DE; Waco, TX; Waterlooo-Cedar Falls, IA; Racine, WI; Flint, MI. The ten cities with the smallest fraction of new plants with at least 1000 workers are: Yakima, WA; Omaha, NE; New Bedford, MA; Nashville, TN; Duluth, MN; Daytona Beach, FL; Monroe, AL; Utica-Rome, NY; Stockton, CA. Example of cities with a fraction of new plants that is close to the sample average are: Toledo, OH; El Paso, TX; Rockford, IL; Akron, OH. In the following 5 Census divisions the fraction of large

4 reports the corresponding regression coefficients.[41]

Is the fraction of new large plants a valid instrument? The instrument is valid if the size distribution of new plants in a city in industries other than the relevant 3-digit industry is orthogonal to the productivity changes in the relevant plant. Note that the instrument is based on openings *outside* the plant's industry. The instrument is not valid if changes in unobserved determinants of plant productivity are correlated with the size distribution of new plants outside the plant's industry.

To investigate the validity of the exclusion restriction, I regress the instrument on 1982 TFP. Finding that the instrument is correlated with the 1982 level of TFP would cast doubt on the validity of the instrument, since the level of productivity in 1982 and the 1982-1992 changes in productivity could be correlated. The coefficient is 0.030 (0.026) and not statistically significant. I also regress the instrument on the number of employees in the relevant plant in 1982 and on the 1982-1992 change in the number of employees in the relevant plant. The concern is that large plant openings may occur close to other large plants (or in areas where plant size is growing), and, at the same time, plants of different size may experience different trends in productivity. If both these facts were true, they would invalidate the instrument. The coefficients on 1982 plant size and on the 1982-1992 changes in plant size are, respectively, 0.123 (0.101) and 0.897 (0.941).

Another possible concern is that the fraction of large plant openings is higher in areas that experience many openings of any size. This could be a problem if areas that experience a large number of openings enjoy positive productivity shocks, that make them particularly attractive (for example: the opening of a port or an airport). To assess this possibility, I regress the instrument on the absolute number of new openings, and the per capita number of new openings. Finding that the fraction of large plants openings is higher in areas that experience a large number of openings would cast doubt on the validity of the instrument. The coefficient on the absolute number of new openings is -1.18 (0.26); the coefficient on the absolute number of new openings normalized by 1982 city population is -0.06 (0.02). This suggests that the fraction of new large plants is higher in cities that experience *fewer* openings of any size.

I also regress the instrument on the average 1982 wage in the relevant plant. The coefficient is 0.102 (0.021). This last result is problematic, because it indicates that the fraction of new large plants outside the relevant industry is positively correlated with the level of wages in

plant openings is above average: New England, Middle Atlantic, East North Central, West North Central, South Atlantic. In the following 4 Census divisions the fraction of large plant openings is below average: East South Central, West South Central, Mountain, Pacific.

[41] Specifically, column 4 reports the coefficient on the fraction of new plants in a given size group among all new plants in a city (excluding the relevant 3-digit industry) in a regression of 1982-1992 changes in manufacturing college share in a city (excluding the relevant 3-digit industry) on fraction of new plants in a given size group among all new plants in a city (excluding the relevant 3-digit industry) in the sample of 40,281 plants.

1982. If the level of wages in 1982 is correlated with the 1982-1992 change in productivity, this would indicate that the instrument is not exogenous.

Table 10 reports instrumental variable estimates.[42] 2SLS estimates in rows 1 and 3 seem to be generally consistent with the corresponding OLS estimates, although standard errors are large and preclude definitive conclusions. The first stage coefficients in row 2 are between 2.22 and 3.34. To help in interpreting the first stage estimates, consider that the instrument has a mean (std deviation) of 0.0025 (0.0024). For the average city-industry, the fraction of large plant openings accounts for a 0.005-0.008 percentage point increase in the aggregate college share outside the relevant industry, or about 11%-18% of the typical increase in college share experienced over a ten year period.[43]

8 Human Capital Spillovers and Wages

The most robust estimates of the spillover indicate that, on average, a one percentage point increase in city college share is associated with a 0.5-0.7 percent increase in productivity. (The average yearly increase in college share is 0.2%). Is this a plausible magnitude? One way to assess the plausibility of the estimated effect is to compare it with the difference in labor costs between cities with high and low human capital. In equilibrium, if firms are really more productive in cities with high levels of human capital, production costs should also be higher. Otherwise, firms would relocate from cities with low human capital to cities with high human capital (see Section 2). The difference in labor costs between cities with high and low human capital is therefore a useful benchmark against which to compare the estimated effect of human capital spillovers on productivity. Finding that the productivity differences between cities with high human capital and low human capital are *larger* than the differences in labor costs (adjusted for the fraction of labor cost to total costs), would suggest that the estimated productivity gains from spillovers are too large, and would cast doubt on the findings in Section 5.[44]

Manufacturing wages are indeed higher in cities where the number of college graduates is high, even after controlling for individual schooling. Compare a city with a large stock of skilled

[42] Because the instrument affects *changes* in aggregate college share, models in Table 10 estimate equation 4 in differences. Specifically: $\Delta \ln y_{pjcs} = \gamma \Delta \overline{S}_{-jc} + \alpha_{Hj} \ln \Delta H_{pjc} + \alpha_{Lj} \Delta \ln L_{pjc} + \beta_j \Delta \ln K_{pjc} + d_{js} + \epsilon_{pjc}$, where Δ represents the 1982-1992 change. Column 1 is equivalent to column 2 in Table 3, column 2 is equivalent to col 3 in Table 3, etc.

[43] Alternatively, compare a city-industry at the 25% percentile with a city-industry at the 75% percentile in terms of fraction of large plant openings. Based on the first stage estimates, the latter has a 0.005-0.008 percentage point increase in the aggregate college share more than the former.

[44] I am abstracting from the cost of capital because it does not vary much across cities. If a substantial part of production costs of manufacturing firms come from land prices, and if cities with higher college share have more expensive land, then it would be possible to find that productivity differences between cities with high human capital and low human capital are larger than the differences in labor costs.

workers like Seattle, WA, with a city with a much smaller stock of skilled workers, like El Paso, TX. The share of college graduates in the Seattle labor force is 0.31, almost double the share of college graduates in El Paso, 0.16. *After controlling for individual schooling*, and other workers' characteristics, average manufacturing wages in Seattle are 20% higher than in El Paso. This implies that an extra percentage point in college share is associated with 1.3 percentage points in higher wages, after controlling for individual observables. As it turns out, the corresponding figure for all U.S. cities is 1.1, not very different.[45]

In equilibrium, this wage difference must reflect productivity differences. Because manufacturing firms produce goods that are traded on the national market, if workers weren't more productive in high-wage cities, manufacturing firms would leave and relocate to low-wage cities. Specifically, we should observe that the productivity increase associated with a one percentage point increase in college share is roughly 1.1*0.7 $\simeq$ 0.75, where 0.7 is the share of output that is typically assumed to go to labor (the remaining 0.3 goes to capital). In other words, this back of the envelope calculation suggests that, in equilibrium, a regression of log output on college share (holding constant other inputs) should yield a coefficient not very far from 0.75.[46]

My most robust estimates in Table 3 place the spillover effect at around 0.5-0.7. I conclude that the estimated productivity differences between cities with high and low levels of human capital are consistent with differences in labor costs that are typically observed between cities with high and low level of human capital.

9 Conclusion

Economists have long speculated that human capital may generate significant spillovers. Lucas (1988), among others, argues that human capital externalities are large enough to explain differences between poor and rich countries in long run growth rates. Yet, despite significant policy implications, systematic empirical evidence on the actual magnitude of externalities is just beginning to emerge. Previous work has focused on differences in education and wages across metropolitan areas.

In this paper, I take a more direct approach by focusing on the productivity of manufacturing

[45] This figure comes from an individual level OLS regression of log wage on city college share, a vector of individual characteristics including education, sex, race, Hispanic origin, U.S. citizenship, a quadratic term in work experience, AFQT score, a vector of family background characteristics, and city and year fixed effects. The coefficient (standard error) on college share is 1.1 (0.21). The sample includes all manufacturing workers in the NLSY. Instrumental variable estimates and panel data estimates that control for individual fixed effects yield similar coefficients. Estimates based on Census data yield similar results. See Moretti (2004) for details.

[46] To see this more formally, consider the simplest possible technology: $y = AL^{\alpha}K^{1-\alpha}$. It easy to see that unit costs are $\ln c = -\ln A + \alpha \ln w + (1-\alpha)\ln r + constant$. In Roback's model, the price of capital is constant across cities. If in equilibrium unit costs are constant across cities, this equation says that any increase in A needs to be offset by a similar increase in α times wages.

establishments. I start from a very simple observation: if human capital spillovers actually exist, then we should observe that plants in cities with a large stock of human capital are more productive than otherwise similar plants in cities with a smaller stock of human capital. My findings suggest that, after controlling for a plant's own level of human capital, plants located in cities where the fraction of college graduates grew faster experienced larger increases in productivity than similar plants in cities where the fraction of college graduates grew more slowly.

Interestingly, the estimated productivity differences between cities with high and low levels of human capital are consistent with differences in manufacturing wages that are typically observed between cities with high and low levels of human capital. Consistent with a model that includes both standard general equilibrium forces and spillovers, the productivity gains generated by human capital spillover appear to be offset by increased labor costs.

Although I control for permanent plant characteristics and state and industry time-varying productivity shocks, I can not completely rule out the possibility that unobserved city heterogeneity may explain part of the estimated effect. However, several pieces of evidence lend credibility to the conclusion that the estimated effect is not completely spurious.

First, the estimated coefficient is remarkably robust across specifications. Different assumptions on technology, omitted variables, and variable definitions all yield similar results.

Second, aggregate human capital in the high-tech sector of the city matters more for high-tech plants than aggregate human capital in the low-tech sector of the city; and aggregate human capital in the low-tech sector matters more for low-tech plants than aggregate human capital in the high-tech sector. More importantly, when I use 3 direct measures of economic distance, I find that, within a city, manufacturing plants benefit more from human capital in industries that are geographically and economically close to manufacturing than from human capital in industries that are geographically close but economically far. This result supports the view that spillovers are related to the amount of interactions between workers in different industries.

Third, unlike density of human capital, density of physical capital outside a plant has no effect on the plant productivity. This indicates that what I am estimating is not simply an agglomeration effect generated by density of economic activity. Finally, an instrumental variable strategy based on the number of large plants openings in the relevant city but outside the relevant 3-digit industry yields estimates that are generally consistent with OLS estimates.

Having established the existence of human capital spillovers, one important direction for future research should be the investigation of the exact mechanisms through which spillovers arise.

References

A., C. and Peri, G. (2002), 'Identifying human capital externalities: Theory with an application to us cities', *Universitat Pompeu Fabra Working Paper 611* .

Acemoglu, D. (1996), 'A microfundation for social increasing returns in human capital accumulation', *Quarterly Journal of Economics* pp. 779–804.

Acemoglu, D. and Angrist, J. (1999), 'How large are human capital externalities? evidence from compulsory schooling laws', *MIT* .

American Electronics Association (1997), 'Aea definition of the high–tech sector', *Cyberstates* .

Branstetter, L. (2001), 'Are knowledge spillovers international or international in scope? microeconomic evidence from us and japan', *Journal of International Economics* **53**, 53–79.

Davis, S. and Haltiwanger, J. (1999), Gross job flows, *in* O. Ashenfelter and D. Card, eds, 'Handbook of Labor Economics'.

Doms, M., Dunne, T. and Troske, K. R. (1997), 'Workers, wages, and technology', *Quarterly Journal of Economics* **112**, 253–290.

Glaeser, E. L. (1999), 'Learning in cities', *Journal of Urban Economics* **46**(2), 254–77.

Hellerstein, J. K., Neumark, D. and Troske, K. R. (1999), 'Wages, productivity and worker characteristics: Evidence from plant-level production functions and wage equations', *Journal of Labor Economics* **17**(3), 409–447.

Henderson, V. (2001), 'Marshall's scale economies', *Brown University mimeo* .

Jaffe, A. B. (1986), 'Technological opportunities and spillovers of r&d: Evidence from firm's patents, profit and market value', *American Economic Review* **76**, 984–1001.

Jaffe, A., Trajtenberg, M. and Henderson, R. (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citation', *Quaterly Journal of Economics* **108**(3), 577–598.

Jaffe, A., Trajtenberg, M. and Henderson, R. (2002), *Patents, Citations and Innovations*, MIT Press.

Jovanovic, B. and Rob, R. (1989), 'The growth and diffusion of knowledge', *Review of Economic Studies* **56**(4), 569–582.

Lucas, R. E. (1988), 'On the mechanics of economic development', *Journal of Monetary Economics* **22**, 3–42.

Marshall, A. (1890), *Principles of Economics*, Macmillan.

McGahan, A. and Silverman, B. (2001), 'How does innovative activity change as industries mature?', *International Journal of Industrial Organization* **19**.

Moretti, E. (2004), 'Estimating the external return to higher education: Evidence from cross-sectional and longitudinal data', *Journal of Econometrics* .

Mowery, D. and Ziedonis, A. (2001), 'The geographic reach of market and non-market channels of technology transfer: Comparing citations and licenses of university patents', *NBER Working Paper 8568* .

Rauch, J. E. (1993), 'Productivity gains from geographic concentration of human capital: Evidence from the cities', *Journal of Urban Economics* **34**, 380–400.

Roback, J. (1982), 'Wages, rents and the quality of life', *Journal of Political Economy* **90**(6), 1257–1278.

Roback, J. (1988), 'Wages, rents and amenities: Differences among workers and regions', *Economic Inquiry* **26**(1), 23–41.

Roberts, M. and Supina, D. (1997), 'Output price and markup dispersion in micro data: The role of producer heterogeneity and noise', *NBER Working Paper 6075* .

Silverman, B. (1999), 'Technological resources and the direction of corporate diversification', *Management Science* **45**.

Weiss, L. W. (1972), 'The geographic size of markets in manufacturing', *The Review of Economics and Statistics* **54**.

Zucker, L. G., Darby, M. R. and Brewer, M. B. (1998), 'Intellectual human capital and the birth of u.s. biotechnology enterprises', *American Economic Review* **88**(1), 290–306.

DATA APPENDIX

CENSUS OF MANUFACTURERS: Plant level data on output, capital, hours worked, industry and metropolitan area are from the Census of Manufacturers. The Census of Manufacturers covers the universe of manufacturing plants with 1 or more employees. Since Standard Metropolitan Statistical Area (SMSA) codes in different years are based on different definitions of metropolitan areas, I correct the 1992 SMSA codes to be consistent with the 1982 definition. I delete all the SMSAs that are new to the '92 sample and were not part of another SMSA in '82.[47] Because the metropolitan area definition was changed after 1982, I also redefine 1992 SMSAs to match the 1982 boundaries. I do this in two steps. First, I make the definition of counties consistent over time because some counties have changed their boundaries during the 1980s and there are coding errors in the Census of Manufacturers county codes. To do so, I use a program written by Randy Becker provided by the CES. Only 5 urban counties are affected (they are located in Georgia, Virginia, Arizona, New Mexico and California). To make sure that all county changes have been captured, I use the County Group Equivalency files. I find 7 more changes in Virginia counties that are not included in the CES program. Once I have a county code that is consistent over time, I use the County Group Equivalency files to identify SMSA boundary changes in the 1992 Census of Manufacturers. In 1982 and 1992 263 SMSAs are identified.

I assign each plant to an industry-city cell based on its 3-digit SIC code and SMSA code. Although 4-digit SIC codes are available, I choose 3-digit industries to maximize consistency with the Census of Population industry classification. For production workers, both the number of workers and the number of hours worked is reported in the Census of Manufacturers. For non-production workers, the number of workers is known, but the number of hours worked is not reported. The number of hours of non-production workers is imputed by assuming that production and non-production workers in the same plant work the same number of hours per capita.

CENSUS OF POPULATION: Data on the skill level of workers in each plant and on the share of college graduates outside the industry come from the 1980 and 1990 Censuses of Population. To maximize sample size, I use the 5% version of the Public Use MicrodataSample (PUMS). The Census industry classification is not the SIC one, but has a similar level of detail as the 3-digit SIC codes. Using the name of the industry, I match the Census industry classification to the SIC one.

As in the Census of Manufacturers, metropolitan area definitions are not consistent across years. To make the 1990 SMSA codes consistent with the 1980 definition, I adopt a procedure consistent with the one described above for the Census of Manufacturers.[48] Years of education

---

[47] I also delete Dayton because it was combined with Springfield, OH and there isn't a good way to separate them and/or to define either one so that it resembles its form in 1982.

[48] Specifically, I assign individuals a metropolitan area on the basis of two geographical identifiers, Public Use Microdata Areas (PUMAs) and metropolitan area codes. The finest geographic units identified in the 5% samples are PUMAs, which are arbitrary geographic divisions that contain no less than 100,000 people each. Most individuals who live in metropolitan areas are also assigned a metropolitan area identifier. However, some PUMAs straddle the boundary of two or more SMSAs and in these 'mixed' PUMAs an SMSA code is not assigned. These 'mixed' PUMAs are assigned a SMSA code on the basis of the County Group Equivalency files. The methodology used to assign SMSA codes and to match MSA across Censuses is identical to the one in Moretti (2004). If over 50% of the PUMA population is attributable to a single MSA, I then assign all individuals in that PUMA to the majority MSA. Since the MSA definition was changed after the 1980 Census,

are assigned to the education codes used in 1990 Census following Table 1 in Kominsky and Siegel (1996). Since 1982 and 1992 are not Census of Population years, linear interpolation is used to estimate the college share for 1982 and 1992.[49]

MATCHING CENSUS OF MANUFACTURERS TO CENSUS OF POPULATION: Plant-level data from the Census of Manufacturers are matched with Census of Population data on workers education by industry and city. I assign each plant in the Census of Manufacturers and each worker in the Census of Population to a city-industry cell based on the metropolitan area code and a 3-digit industry definition. To minimize the amount of measurement error, I exclude all industry-city cells with less than 10 workers. There are a total of 3441 cells. The average number of workers in a cell is 546 in 1980 and 387 in 1990. About 18% of the cells contain only one plant. (The fraction of cells that include only one plant is calculated for the balanced panel used for regressions, not for the population of plants.) The median cell in 1992 includes 100 workers and 4 plants.[50]

The Census of Manufacturers has 381,773 plants in 1982 and 348,385 in 1992. To build the balanced panel used in this paper, I first exclude all plants that do not appear both in 1982 and 1992. A total of 161,321 plants exist in both years. I then delete plants for which some of the relevant variables are missing in at least one year. I also exclude from the sample all plants that have capital or production hours or non production hours equal zero. With Cobb-Douglas or Translog production functions, output is zero for any plants where one of the inputs is zero. Finally, I delete industry-city cells with less than 10 workers from the Census of Population. The resulting balanced panel sample has 40,281 plants in 1982 and 1992. This sample covers approximately 24% of average annual manufacturing employment over the period from 1982 to 1992. Large plants are overrepresented in the worker-firm matched sample. For example, the average number of hours worked by all plants in 1982 is 105.2, less than half than the average number of hours worked by plants in the matched sample. Similarly, output, value added, value of capital and wages are lower in the population than in my sample. The average output, value added, capital and wages in the population of plants in 1982 are, respectively: 9018; 3828; 3336; 12.2. However, the non-representativeness of the worker-firm matched sample does not seem to bias the estimates in any significant way (see Section 7.2).

In theory, the Worker Establishment Characteristics Database (WECD) could have been used instead of the sample used here. WECD matches the Census of Manufacturers to the Census of Population using a more precise algorithm that requires eliminating from the sample all observations located in cells with more than one plant (Hellerstein et al. 1999). The main

---

I redefine 1990 SMSAs to match the 1980 boundaries. The County Group Equivalency files are used to identify PUMAs that contain the affected counties in the 1990 Census. If the counties in question comprise more than half of the PUMAs population, all respondents are assigned to the pertinent SMSA. If more than 10% of a SMSAs 1990 population is affected by the boundary changes and is unrecoverable from the County Equivalency files, I drop the city from the analysis. Dayton and Springfield, Ohio are the only such cities. 282 SMSAs are identified in 1980 and 1990. The computer code for this assignment is available on request.

[49] An alternative would have been to use averages obtained yearly from the Current Population Survey. Given the smaller sample size of the CPS, results obtained by interpolating Census averages turn out to be more precise than results obtained from CPS averages.

[50] There is a wide variation in cell size across industries. For example, Petroleum Refining (SIC 291) and Engine and Turbines (SIC 351) have typically only one plant per cell, while Plastic Products (SIC 308) and Scientific Instruments (SIC 381, 382) have 7 plants per cell. Not all industries are present in all cities. For example, Office and Accounting Machines (SIC 357) plants are present in only 29 cities, while there are Electrical Machinery (SIC 361, 362, 364, 367, 369) plants in 197 different cities.

reason why I do not use WECD is that it is available only for 1992 and does not allow for a longitudinal analysis.

PLANT-LEVEL WAGE EQUATIONS: In order to assess the quality of the match between workers and plants, I have estimated plant-level wage equations. Although the focus of this paper is not on wages, plant-level wage equations provide an indirect test of the quality of the matching. If the matching is correct and measurement error is not too large, one would expect wage equation coefficients to be close to the ones usually found in the wage equation literature.[51] Data on wages, from the Census of Manufacturers, are plant averages obtained by dividing the total wage bill by the number of hours worked. Data on workers are cell averages from the Census of Population. For example, "percentage female" is the fraction of women in the industry and city to which the plant belongs. The coefficients are roughly similar to the ones found in the literature based on individual level regressions and the ones found in Hellerstein et al. (1999), based on a plant-level regression. For example, the coefficients on years of schooling are 0.078 (0.004) and 0.086 (0.006) for 1992 and 1982 respectively. These coefficients are slightly smaller—probably because of measurement error—but not completely different from the standard estimates of the return to education obtained from worker-level data.[52] Women and blacks are paid less, and older workers more. The coefficients on female, black and age in 1992 are, respectively: -0.304 (0.029), -0.123 (0.056), 0.011 (0.001). The coefficients on female, black and age in 1982 are, respectively: -0.449 (0.026), -0.026 (0.058), 0.012 (0.001).

I conclude that the matched worker-firm sample contains some measurement error, but can roughly reproduce standard individual level wage equation results.

PATENTS: To construct the two measures of economic distance based on patents, I use the NBER patent dataset. I use all patents granted after 1970. For the index based on patent citations, I exclude self-citations. A major problem in linking patents to the Census of Manufacturers is that patents are not directly assigned industry codes. I use the concordance that links the International Patent Classification (IPC) system to the SIC system at the four-digit SIC level developed by Brian Silverman. The concordance has been used by various scholars to assess the specific industries in which firms have technological strength (Silverman 1999), patenting activity through the industry life cycle (McGahan and Silverman 2001), and industry-specific effects in university-industry technology transfer (Mowery and Ziedonis 2001). The concordance and a detailed explanation on how it was constructed are available at *www.rotman.utoronto.ca/ silverman*. In interpreting my results, it is important to keep in mind that the patent-SIC code is not one-to-one. Silverman's concordance assigns multiple SIC codes to each patent. I use the variable *usefreq* to select the SIC code that is most important for each patent, and ignore all the other SIC codes. This is likely to introduce some measurement error, which could bias downward the documented relationship between spillovers and the two measures of economic distance based on patents.

[51] Hellerstein et al. (1999) show that plant-level wage equations represent the aggregation of individual-level wage equations over workers employed in a plant and hence should provide coefficients similar to the ones obtained from their individual-level counterparts.

[52] One difference is that the 1992 estimates are usually found to be larger than the 1982 ones.

Table 1: Summary Statistics

	1982		1992	
	Mean	Std. Dev.	Mean	Std. Dev.
	(1)	(2)	(3)	(4)
Value of Output (x 1000)	19944.0	163592.5	20938.7	174087.5
Added Value (x 1000)	8019.1	55307.8	9412.53	69566.91
Capital (x 1000)	7042.1	62003.4	8007.13	62502.58
Hours Worked (x1000)	223.2	432.4	222.7	435.7
Hours worked by College Graduates (x1000)	38.9	309.5	39.0	310.8
Hours worked by Workers with Some College (x1000)	43.5	338.6	67.7	333.7
Hours worked by High School Graduates (x1000)	90.4	342.4	73.7	310.0
Hours worked by High School Drop Outs (x1000)	50.2	249.4	42.4	136.8
High-Tech	0.099	0.298	0.099	0.298
Average Hourly Wage	13.54	5.50	13.73	5.40
Belong to Multi-Units Firm	0.25	0.43	0.29	0.45
College Share in Other Industries	0.161	0.042	0.191	0.061
Number of Plants	40,281		40,281	

NOTES: Monetary values are in 1992 dollars.

Table 2: Estimates of Production Functions: Cross-Sectional Specification

	Cobb-Douglas		Translog	
	1992	1982	1992	1982
	(1)	(2)	(3)	(4)
College Share in Other Industries	0.846	0.812	0.834	0.807
	(0.102)	(0.113)	(0.107)	(0.133)
ln Capital	0.178	0.476	0.501	0.657
	(0.004)	(0.010)	(0.050)	(0.057)
ln Unskilled Labor	0.470	0.333	0.606	0.332
	(0.014)	(0.012)	(0.040)	(0.050)
ln Skilled Labor	0.382	0.196	0.465	0.265
	(0.015)	(0.010)	(0.039)	(0.060)
ln Unskilled Labor sq.			0.098	0.096
			(0.011)	(0.010)
ln Skilled Labor sq.			0.068	0.053
			(0.011)	(0.012)
ln Capital sq.			0.048	0.024
			(0.002)	(0.002)
ln Unskilled × ln Skilled			-0.111	-0.022
			(0.018)	(0.014)
ln Unskilled × ln Capital			-0.075	-0.095
			(0.010)	(0.016)
ln Skilled × ln Capital			-0.028	-0.047
			(0.011)	(0.016)
Multi-Unit	0.150	0.073	0.122	0.069
	(0.008)	(0.011)	(0.012)	(0.012)
Industry Effects	Yes	Yes	Yes	Yes
R. sq.	0.89	0.89	0.91	0.90

NOTES: Standard errors adjusted for clustering in parenthesis. Each column is a separate regression. All labor inputs are measured in number of hours worked. Specifically, unskilled labor is hours worked by workers who have a high-school degree or less; skilled labor is hours worked by college graduates or workers with some college. Industry effects are dummies for 3-digit industries. N = 40281. See text for details.

Table 3: Longitudinal Estimates of Human Capital Spillovers

	Cobb-Douglas					Translog				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PANEL A										
(1) Controlling for Two Education Groups Inside Plant										
College Share in Other Industries	0.743	0.510	0.734	0.777	0.702	0.672	0.457	0.660	0.711	0.604
	(0.183)	(0.160)	(0.226)	(0.226)	(0.220)	(0.172)	(0.154)	(0.212)	(0.215)	(0.211)
(2) Controlling for Three Education Groups Inside Plant										
College Share in Other Industries	0.684	0.511	0.688	0.736	0.694	0.584	0.441	0.593	0.656	0.539
	(0.195)	(0.171)	(0.245)	(0.207)	(0.221)	(0.181)	(0.225)	(0.225)	(0.210)	(0.211)
PANEL B										
(3) Controlling for Two Education Groups for PW + Two Education Groups for Non-PW										
College Share in Other Industries	0.847	0.595	0.834	0.883	0.914	0.747	0.498	0.724	0.757	0.719
	(0.181)	(0.160)	(0.231)	(0.222)	(0.211)	(0.169)	(0.152)	(0.211)	(0.214)	(0.212)
(4) Controlling for Three Education Groups for PW + Three Education Groups for Non-PW										
College Share in Other Industries	0.687	0.557	0.669	0.795	0.888	0.603	0.464	0.568	0.671	0.589
	(0.215)	(0.194)	(0.259)	(0.254)	(0.235)	(0.202)	(0.182)	(0.232)	(0.237)	(0.236)
Establishment Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind. × Year Effects		Yes					Yes			
State × Year Effects			Yes					Yes		
Ind. × State × Year				Yes	Yes				Yes	Yes
Technology Varies by Ind.					Yes					Yes

NOTES: Standard errors adjusted for clustering in parenthesis. The equation estimated is variants of equation 4. In row 1, models control for hours worked inside the plant by 2 education groups (high school or less, some college or more). In row 2, models control for hours worked inside the plant by 3 education groups (high school or less, some college, college). In row 3, models control for hours worked inside the plant by production workers belonging to 2 education groups and for hours worked inside the plant by non-production workers belonging to 2 education groups (high school or less, some college or more). In row 4, models control for hours worked inside the plant by production workers belonging to 3 education groups and for hours worked inside the plant by non-production workers belonging to 3 education groups (high school or less, some college, college). All models also control for capital. Models in column 6 to 10 also control for capital squared, hours worked by each education group squared and all the interactions. Each entry is a separate regression. There are 40,281 plants, observed in both 1982 and 1992.

Table 4: Longitudinal Estimates of Human Capital Spillovers, by High-Tech Status

	Plant is High-Tech (1)	Plant is Low-Tech (2)
REGRESSION A: COBB-DOUGLAS		
College Share in High-Tech Outside Relevant 3-digit Industry	1.70	0.14
	(0.31)	(0.22)
College Share in Low-Tech Outside Relevant 3-digit Industry	0.22	0.80
	(0.88)	(0.50)
REGRESSION B: TRANSLOG		
College Share in High-Tech Outside Relevant 3-digit Industry	1.60	0.07
	(0.30)	(0.22)
College Share in Low-Tech Outside Relevant 3-digit Industry	0.26	0.89
	(0.87)	(0.53)
Establishment Effects	Yes	Yes
Ind. × State × Year	Yes	Yes

NOTES: Standard errors adjusted for clustering in parenthesis. Both models control for capital, hours worked by skilled and unskilled workers. All entries in each panel are from the same regression. For example, the entry in row 1, column 1 is the coefficient on college share in high-tech industries (outside relevant 3-digit industry) interacted with a dummy equal one if the relevant plant is high-tech. The entry in row 1, column 2 is the coefficient on college share in high-tech industries (outside relevant 3-digit industry) interacted with a dummy equal one if the relevant plant is low-tech. There are 40,281 plants, observed in both 1982 and 1992. See text for details.

Table 5: Longitudinal Estimates of Human Capital Spillovers at the 2-Digit and 1-Digit Industry Level

	Cobb-Douglas		Translog	
	(1)	(2)	(3)	(4)
College Share in 2-Digit Industry	1.008	0.956	0.917	0.879
Excluding Relevant 3-Digit Industry	(0.300)	(0.315)	(0.304)	(0.310)
College Share in Manufacturing	0.751	0.683	0.632	0.579
Excluding Relevant 2-Digit Industry	(0.242)	(0.293)	(0.271)	(0.286)
Establishment Effects	Yes	Yes	Yes	Yes
Ind. × State × Year	Yes	Yes	Yes	Yes
Technology Varies by Ind.		Yes		Yes

NOTES: Standard errors adjusted for clustering in parenthesis. Row 1 reports the coefficients on the share of college graduates in 2-digit industry the plant belongs to, calculated excluding the 3-digit industry the plant belongs to. Row 2 reports the coefficients on the share of college graduates in manufacturing, calculated excluding the 2-digit industry the plant belongs to. All models control for capital, hours worked by skilled and unskilled workers and establishment effects. Each column is a separate regression. There are 40,281 plants, observed in both 1982 and 1992.

Table 6: The Effect of Physical Capital Outside the Plant on Plant Productivity

	Cross-Section		Panel		
	1982	1992			
	(1)	(2)	(3)	(4)	(5)
Model 1:					
Coeff on ln Average Capital Outside Plant	0.349	-0.150	0.012	0.007	0.007
	(0.067)	(0.065)	(0.017)	(0.017)	(0.018)
Model 2:					
Coeff on ln Average Capital Per Worker Outside Plant	1.871	-0.724	-0.005	-0.006	-0.004
	(0.661)	(0.404)	(0.021)	(0.021)	(0.022)
Establishment Effects			Yes	Yes	Yes
State×Year Effects				Yes	
Industry×Year Effects					Yes

NOTES: Standard errors adjusted for clustering in parenthesis. Each entry is from a separate regression. The equation estimated is equation 4, where human capital $\overline{S}$ is substituted with a measure of physical capital. Entries in row 1 are the coefficients on the log of average physical capital outside the plant in a city. Entries in row 2 are the coefficients on the log of per worker average physical capital outside the plant in the city. All models control for capital in the plant and hours worked by skilled and unskilled workers in the plant. There are 40,281 plants, observed in both 1982 and 1992.

Table 7: Longitudinal Estimates of Human Capital Spillovers Based on TFP

	(1)	(2)	(3)	(4)
College Share in Other Industries	0.461	0.255	0.636	0.693
	(0.245)	(0.241)	(0.295)	(0.311)
Establishment Effects	Yes	Yes	Yes	Yes
Ind. × Year Effects		Yes		
State × Year Effects			Yes	
Ind. × State × Year				Yes

NOTES: Standard errors adjusted for clustering in parenthesis. Estimates are obtained as follows. First TPF is estimated under the assumptions that (1) technology is Cobb-Douglas; (2) factor prices equal marginal products; (3) there are constant returns to scale. The labor elasticity is measured at the plant level as the ratio of total wages over total output. The capital elasticity is one minus the labor elasticity. Second, TFP is regressed on college share in other industries. Each column is a separate regression. There are 40,281 plants, observed in both 1982 and 1992.

Table 8: Robustness Checks

	Coeff. on College Share in Other Industries
(1) Base Specification	0.743
	(0.183)
(2) Shipments	0.866
	(0.198)
(3) Number of Colleges	0.715
	(0.184)
(4) Number of College Degrees Awarded	0.747
	(0.188)
(5) Number of Colleges + Degrees Awarded	0.750
	(0.189)
(6) Drop ASM plants	0.860
	(0.280)
(7) Drop computer plants	0.723
	(0.175)
(8) Weighted regression	0.694
	(0.202)
(9) Small plants (1-10 workers)	0.692
	(0.259)
(10) Medium plants (11-50 workers)	0.816
	(0.206)
(11) Large plants (51+ workers)	0.755
	(0.319)
(12) Coeff on labor inputs vary across cities, time	0.501
	(0.202)
(13) City density	0.731
	(0.187)
(14) City pop. + other city characteristics	0.703
	(0.188)
(15) Single-unit plants	0.919
	(0.178)
(16) Multi-unit plants	0.428
	(0.323)

NOTES: Standard errors adjusted for clustering in parenthesis. Each entry is a separate regression.
(1) The base case is from Table 3, column 1.
(2) The dependent variable is value of shipments.
(3) Model controls for number of colleges in city.
(4) Model controls for number of college degrees awarded.
(5) Model controls for number of colleges and number of degrees.
(6) Sample does not include plants in Annual Survey of Manufacturers.
(7) Sample does not include computer and computer accessories plants.
(8) Weights are based on the distribution of plant size in the 1982 population.
(9) Sample includes only plants with 10 workers or less.
(10) Sample includes only plants with 11-50 workers.
(11) Sample includes only plants with more than 50 workers.
(12) The coefficient on skilled and unskilled labor can vary across cities and over time.
(13) Model controls for city density.
(14) Model controls for population, percent unemployed, black, immigrant and female.
(15) Sample includes only single-unit plants.
(16) Sample includes only multi-unit plants.

Table 9: The Size Distribution of New Plants in 1992 and Their Impact on Changes in the Aggregate College Share

	Number of New Plants in 1992	Total Employment	Correlation with Changes in College Share in Other Industries	Coeff.
	(1)	(2)	(3)	(4)
1-10 Workers	144,645	483,515	0.02	0.013
				(0.011)
11-100 Workers	65,347	2,026,164	-0.07	-0.054
				(0.012)
101-500 Workers	9278	1,797,210	0.05	0.102
				(0.039)
501-1000 Workers	773	522,462	0.13	1.329
				(0.189)
1000 + Workers	413	969,952	0.19	2.249
				(0.325)

NOTES: Entries in columns 1 and 2 refer to all the plants in the LRD that exist in 1992 but did not exist in 1982. For columns 3 and 4, the fraction of new plants in a given size group among all new plants in a city (excluding the relevant 3-digit industry) was assigned to each of the plants in the sample of 40,281 plants used for all the regressions in this paper. Column 3 reports the correlation between the fraction of new plants in a given size group among all new plants in a city in 1992 (excluding the relevant 3-digit industry) and the 1982-1992 change in manufacturing college share in the city (excluding the relevant 3-digit industry) in the sample of 40,281 plants. Column 4 reports the coefficient on the fraction of new plants in a given size group among all new plants in a city (excluding the relevant 3-digit industry) in a regression of 1982-1992 changes in manufacturing college share in a city (excluding the relevant 3 digit industry) on fraction of new plants in a given size group among all new plants in a city (excluding the relevant 3-digit industry) in the sample of 40,281 plants.

Table 10: Instrumental Variable Estimates of Human Capital Spillovers

	Cobb-Douglas				Translog			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) IV is Based on Fraction of New Plants with 1000 or More Workers								
(1) College Share in Other Industries	0.84	1.03	1.28	1.21	0.95	0.87	1.29	1.15
	(0.66)	(0.62)	(0.57)	(0.57)	(0.62)	(0.66)	(0.56)	(0.57)
(2) First Stage	2.23	3.39	3.24	3.23	3.37	2.22	3.23	3.21
	(0.29)	(0.28)	(0.30)	(0.29)	(0.29)	(0.28)	(0.29)	(0.29)
(2) IV is Based on Fraction of New Plants with 500 or More Workers								
(3) College Share in Other Industries	0.85	0.87	1.59	1.43	0.88	0.79	1.59	1.32
	(0.70)	(0.72)	(0.64)	(0.65)	(0.71)	(0.69)	(0.64)	(0.65)
(4) First Stage	1.15	1.57	1.55	1.54	1.56	1.15	1.55	1.53
	(0.15)	(0.13)	(0.15)	(0.15)	(0.15)	(0.14)	(0.15)	(0.15)
Industry Effects	Yes				Yes			
State Effects		Yes				Yes		
Ind. × State			Yes	Yes			Yes	Yes
Technology Varies by Ind.				Yes				Yes

NOTES: Standard errors adjusted for clustering in parenthesis. The equation estimated is equation 4 in differences. Specifically: $\Delta \ln y_{pjcs} = \gamma \Delta \overline{S}_{-jc} + \alpha_{Hj} \ln \Delta H_{pjc} + \alpha_{Lj} \Delta \ln L_{pjc} + \beta_j \Delta \ln K_{pjc} + d_{js} + \epsilon_{pjc}$. (Column 1 is equivalent to column 2 in Table 3, column 2 is equivalent to column 3 in Table 3, etc.) The instrumental variable is the fraction of large plant openings among all the plant openings in a city excluding the relevant 3-digit industry. In model 1, large plant openings are defined as openings of plants with 1000 workers or more. There are 413 such openings, and they account for 18% of employment in new plants. In model 2, large plant openings are defined as plant openings with 500 workers or more. There are 1186 such openings, and they account for 26% of employment in new plants. The dependent variable in the first stage is the college share in other industries. Each entry is a separate regression. There are 40,281 plants in the sample, observed in both 1982 and 1992.

Appendix Table A1: The Spillover Effect of College Share in 1-Digit and 2-Digit Industries, by Economic Distance. Distance Based on Input-Output Tables

	Distance (Rank)	Inputs in Manuf. (Billions of Dollars)	Coefficient on College Share in Specified Industry
	(1)	(2)	(3)
Model A: 1-Digit Industries			
Manufacturing	1	841	0.802
			(0.192)
Transportation, Communication, Utilities	2	122	0.488
			(0.213)
Trade	3	119	0.705
			(0.335)
Services	4	112	0.213
			(0.284)
Mining	5	93	-0.004
			(0.042)
Finance	6	29	0.048
			(0.154)
Construction	7	14	0.273
			(0.339)
Model B: 2-Digit Industries			
	1-5	30.7	0.577
			(0.187)
	6-10	9.8	0.314
			(0.233)
	11-15	4.2	0.413
			(0.182)
	16-20	2.6	0.230
			(0.186)
	21-25	1.8	0.232
			(0.161)
	26-30	1.1	0.076
			(0.119)
	30+	0.9	0.083
			(0.074)

NOTES: Model A: Column 1 reports the rank based on the value of inputs used in manufacturing from the specified 1-digit industry. Column 2 reports the value of inputs used in manufacturing from the specified industry. Entries in column 3 are the γ_k coefficients in equation 5, where k indexes 1-digit industries. (When $k = Manufacturing$, I calculate college share excluding the relevant 3-digit industry.) Figure 2 (top panel) plots column 3 against column 1. Model B: Column 1 reports the rank based on the value of inputs used in the relevant 2- digit industry from the specified 2-digit industry group. Column 2 reports the average value of inputs used in the relevant 2-digit industry from the specified 2-digit industry group. Entries in column 3 are the γ_k coefficients in equation 5, where k indexes 2-digit industries. For example, the entry in row 1 is the coefficient on college share in the 5 2-digit industries that are closest to the relevant plant. Figure 2 (bottom panel) plots column 3 against column 1. Both models control for capital, hours worked by skilled and unskilled workers, establishment effects, industry×year and state×years effects. There are 40,281 plants in the sample, observed in both 1982 and 1992.

Appendix Table A2: The Spillover Effect of College Share in 1-Digit and 2-Digit Industries, by Economic Distance. Distance Based on the Distribution of Patents Across Technological Groups

	Distance (Rank) (1)	Uncentered Correlation Coefficient ρ (2)	Coefficient on College Share in Specified Industry (3)
Model A: 1-Digit Industries			
Manufacturing	1	1	0.802
			(0.192)
Trade	2	.618	0.705
			(0.335)
Mining	3	.577	-0.004
			(0.042)
Construction	4	.450	0.273
			(0.339)
Transportation, Communication, Utilities	5	.398	0.488
			(0.213)
Finance	6	.363	0.048
			(0.154)
Services	7	.345	0.213
			(0.284)
Model B: 2-Digit Industries			
	1-5	0.694	0.760
			(0.180)
	6-10	0.461	0.407
			(0.187)
	11-15	0.402	0.313
			(0.178)
	16-20	0.360	0.388
			(0.205)
	21-25	0.229	0.225
			(0.303)
	26-30	0.106	0.369
			(0.371)
	30+	0.063	0.076
			(0.356)

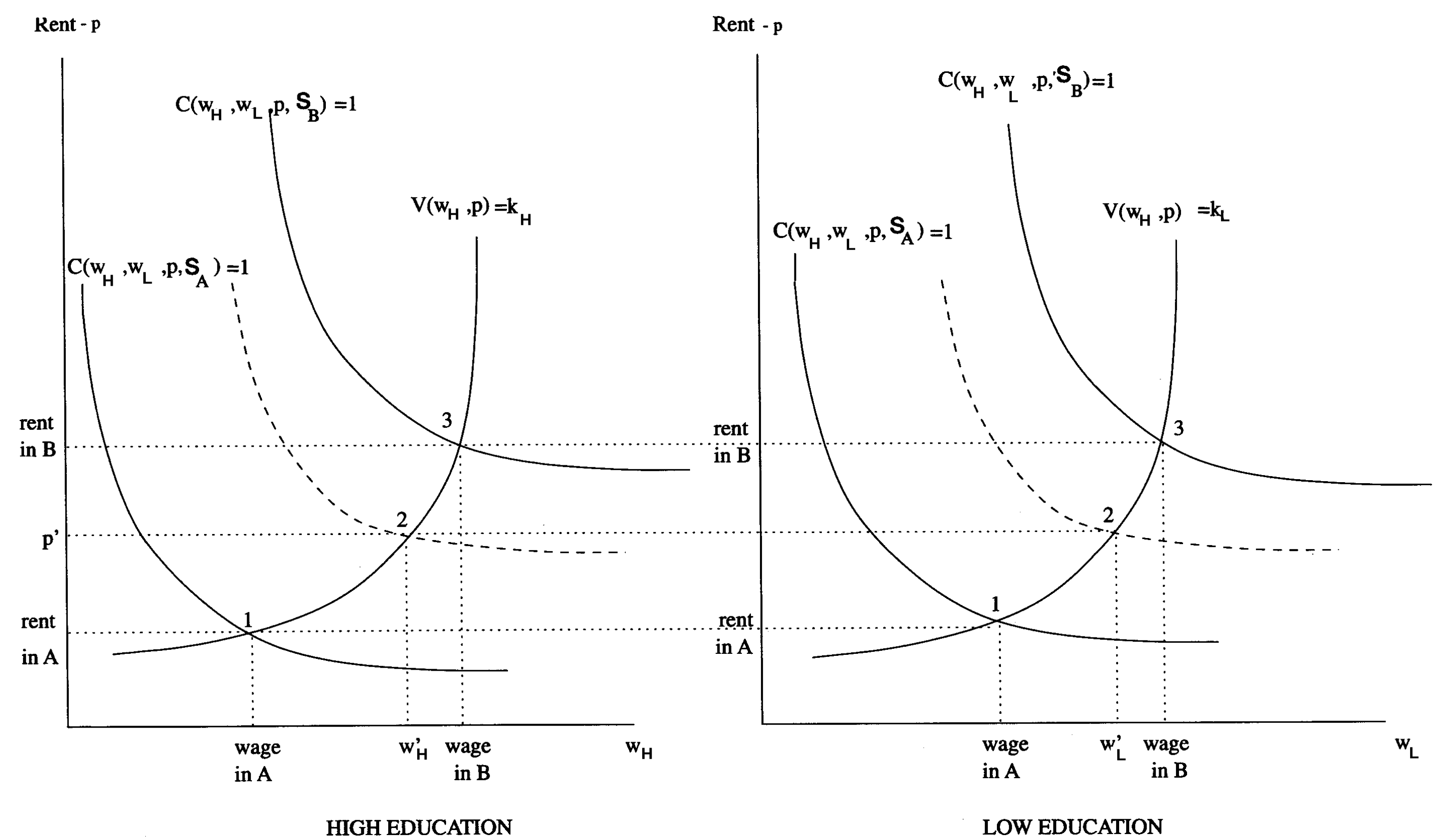
NOTES: Model A: Column 1 reports the rank based on the uncentered correlation coefficient between manufacturing and the specified 1-digit industry. Column 2 reports the uncentered correlation coefficient between manufacturing and the specified 1-digit industry. The uncentered correlation coefficient is defined in equation 6. Entries in column 3 are the γ_k coefficients in equation 5, where k indexes 1-digit industries. (When $k = Manufacturing$, I calculate college share excluding the relevant 3-digit industry.) Figure 3 (top panel) plots column 3 against column 1. Model B: Column 1 reports the rank based on the uncentered correlation coefficient. Column 2 reports the uncentered correlation coefficient between the specified 2-digit industry group and the relevant 2-digit industry. Entries in column 3 are the γ_k coefficients in equation 5, where k indexes 2-digit industries. For example, the entry in row 1 is the coefficient on college share in the 5 2-digit industries that are closest to the relevant plant. Figure 3 (bottom panel) plots column 3 against column 1. Both models control for capital, hours worked by skilled and unskilled workers, establishment effects, industry×year and state×years effects. There are 40,281 plants in the sample, observed in both 1982 and 1992.

Appendix Table A3: The Spillover Effect of College Share in 1-Digit and 2-Digit Industries, by Economic Distance. Distance Based on Patent Citations.

	Distance (Rank)	Frequency of Patent Citations	Coefficient on College Share in Specified Industry
	(1)	(2)	(3)
Model A: 1-Digit Industries			
Manufacturing	1	0.743	0.802
			(0.192)
Services	2	0.134	0.213
			(0.284)
Transportation, Communication, Utilities	3	0.064	0.488
			(0.213)
Construction	4	0.029	0.273
			(0.339)
Trade	5	0.016	0.705
			(0.335)
Mining	6	0.010	-0.004
			(0.042)
Finance	7	0.009	0.048
			(0.154)
Model B: 2-Digit Industries			
	1-5	0.091	0.337
			(0.122)
	6-10	0.043	0.249
			(0.180)
	11-15	0.032	0.213
			(0.208)
	16-20	0.020	-0.032
			(0.185)
	21-25	0.009	0.033
			(0.208)
	26-30	0.002	0.260
			(0.196)
	30+	0.0002	0.071
			(0.269)

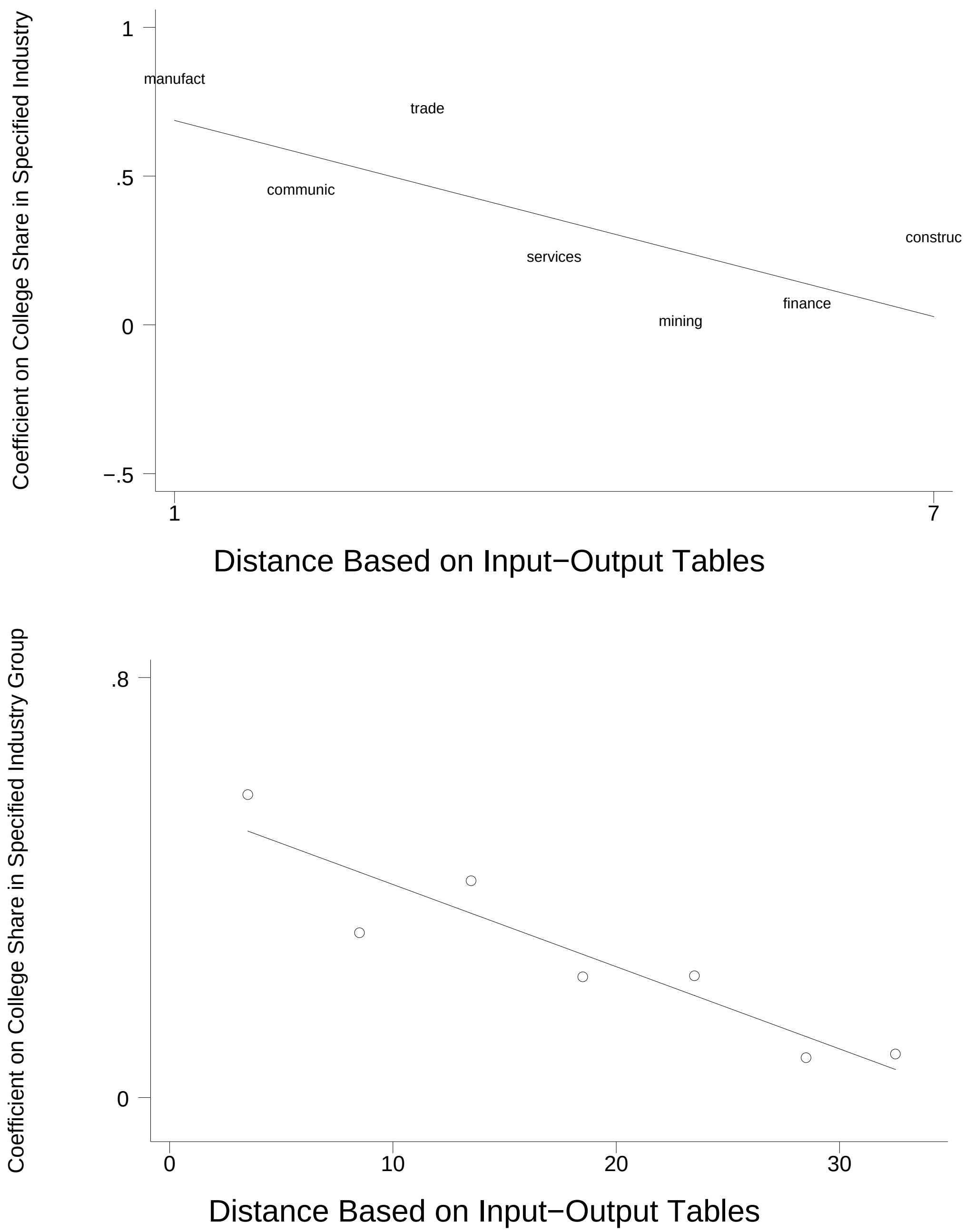
NOTES: Model A: Column 1 reports the rank based on the frequency of patent citations between the manufacturing and the relevant 1-digit industry. Column 2 reports the probability that patents assigned to manufacturing firms cite patents assigned to firms in the specified industry. Entries in column 3 are the γ_k coefficients in equation 5, where k indexes 1-digit industries. (When $k = Manufacturing$, I calculate college share excluding the relevant 3-digit industry.) Figure 4 plots column 3 against column 1. Model B: Column 1 reports the rank based on the frequency of patent citations between the specified 2-digit industry group and the relevant 2-digit industry. Column 2 reports the frequency that the relevant 2-digit industry cites patents assigned to the specified group. Entries in column 3 are the γ_k coefficients in equation 5, where k indexes 2-digit industries. For example, the entry in row 1 is the coefficient on college share in the 5 2-digit industries that are closest to the relevant plant. Figure 4 (bottom panel) plots column 3 against column 1. Both models control for capital, hours worked by skilled and unskilled workers, establishment effects, industry$\times$year and state$\times$years effects. There are 40,281 plants in the sample, observed in both 1982 and 1992.

Figure 1: Equilibrium Wages and Rent in Two Cities



NOTES: Point 1 is the equilibrium in city A. Point 2 is the equilibrium in city B without externality. Point 3 is the equilibrium in city B with externality. The dashed lines in both panels are the isocost curves in city B without externality. w_H and w_L are the nominal wages of educated and uneducated workers, respectively.

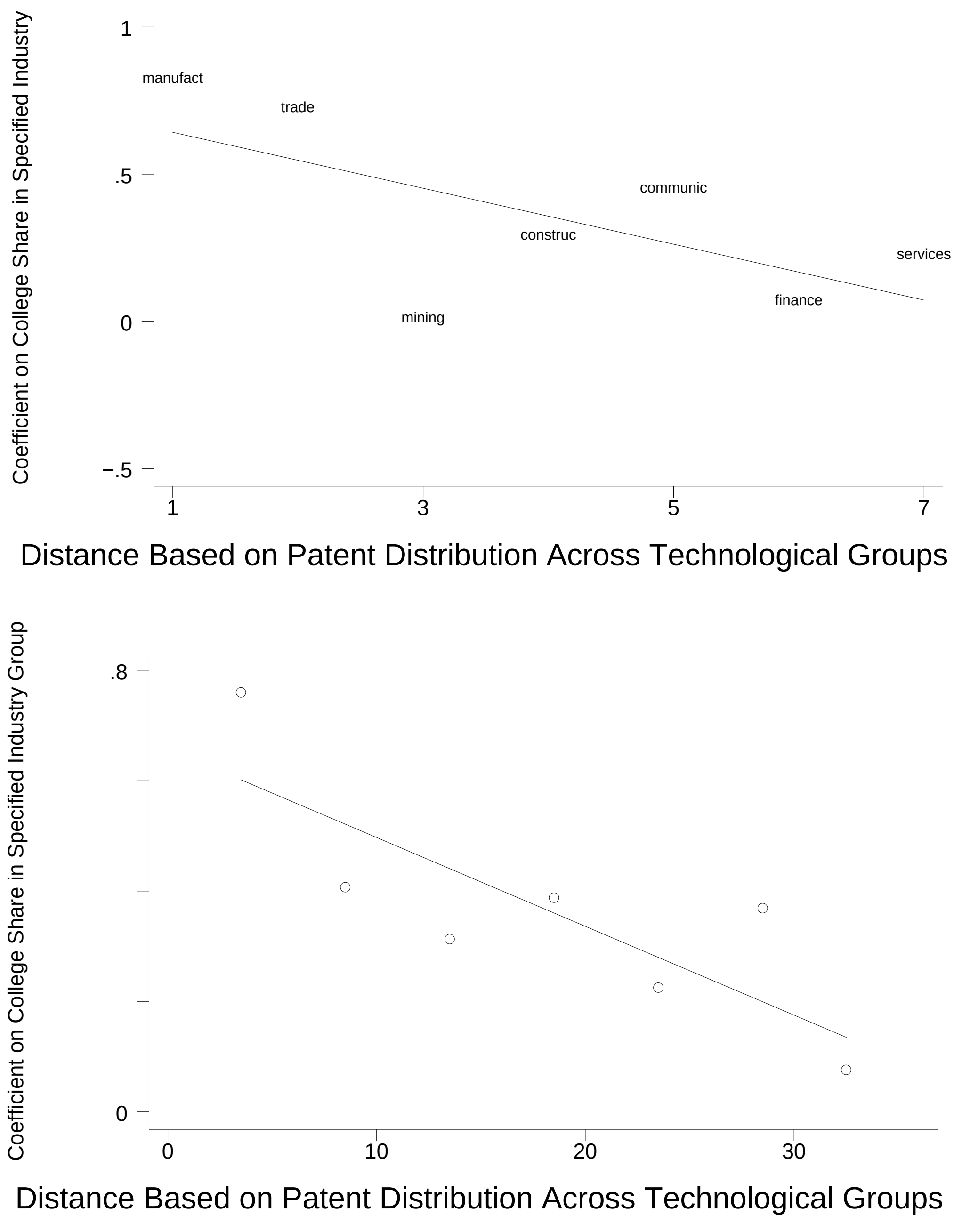
Figure 2: The Spillover Effect of College Share in 1-Digit and 2-Digit Industries, by Economic Distance. Distance Based on Input-Output Tables



Top panel: The figure plots the estimated coefficients on college share in each 1-digit industry (from Table A1, column 3) on the y-axis against the rank in value of inputs, on the x-axis (from Table A1, column 1). For example, manufacturing has rank 1, because the value of inputs is highest for manufacturing. Transportation has rank 2, trade has rank 3, etc. The OLS fitted line is superimposed. The slope (standard error) of the line is -0.109 (0.042). R^2 is .57.

Bottom Panel: The figure plots the estimated coefficients on college share in each 2-digit industry group (from Table A1, column 3) on the y-axis against the rank in value of inputs, on the x-axis (from Table A1, column 1). The OLS fitted line is superimposed. The slope (standard error) of the line is -0.015 (0.002). R^2 is .85.

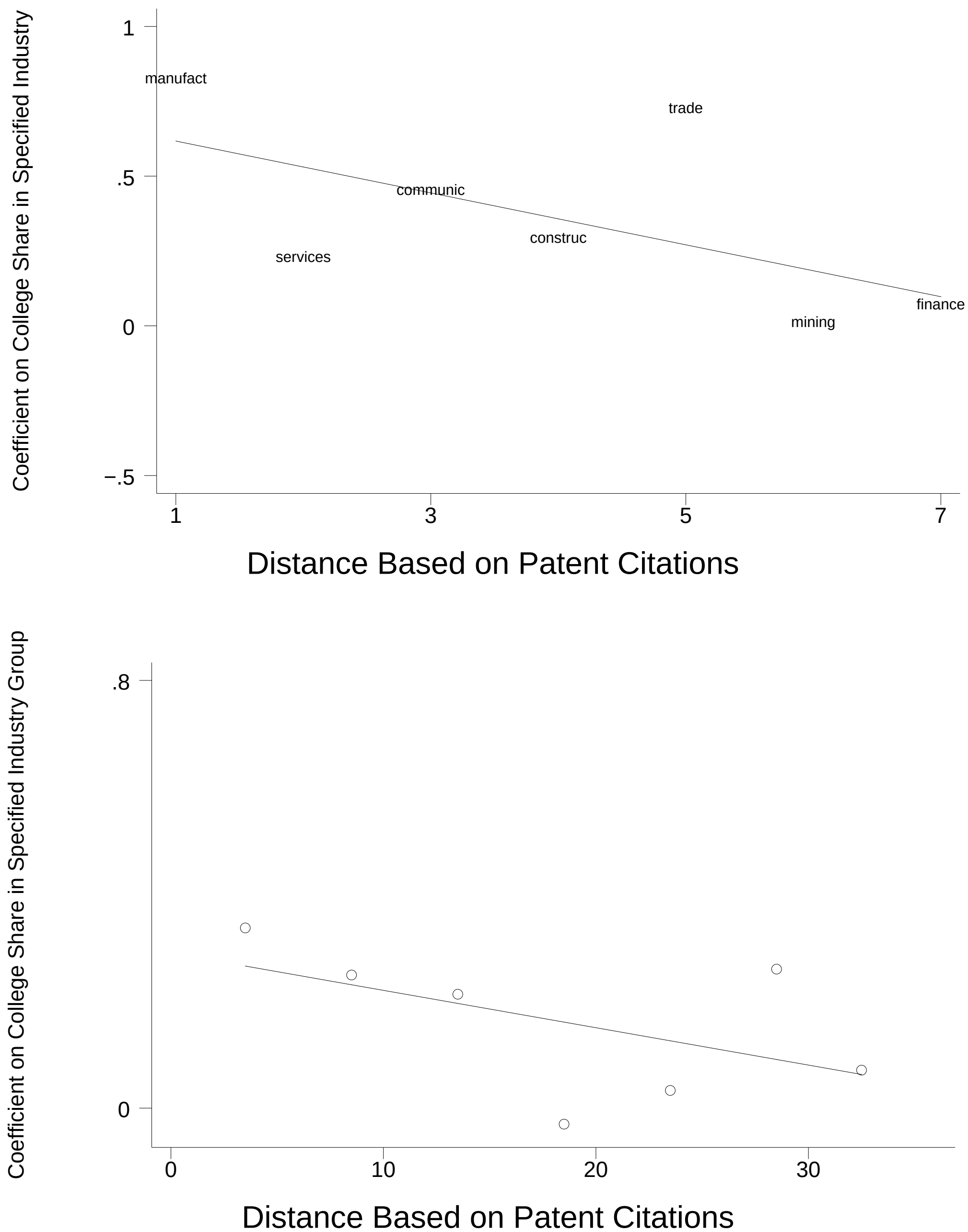
Figure 3: The Spillover Effect of College Share in 1-Digit and 2-Digit Industries, by Economic Distance. Distance Based on the Distribution of Patents Across Technological Groups



Top panel: The figure plots the estimated coefficients on college share in each industry (from Table A2, column 3) on the y-axis against the rank in the uncentered correlation coefficient based on differences in the distribution of patents across technological groups, on the x-axis (from Table A2, column 1). For example, manufacturing has rank 1, because the uncentered correlation coefficient is highest for manufacturing. Trade has rank 2, mining has rank 3, etc. The OLS fitted line is superimposed. The slope (standard error) of the line is -0.095 (0.048). R^2 is .43.

Bottom Panel: The figure plots the estimated coefficients on college share in each industry group (from Table A2, column 3) on the y-axis against the rank in the uncentered correlation coefficient based on differences in the distribution of patents across technological groups, on the x-axis (from Table A2, column 1). The OLS fitted line is superimposed. The slope (standard error) of the line is -0.016 (0.005). R^2 is .65.

Figure 4: The Spillover Effect of College Share in 1-Digit and 2-Digit Industries, by Economic Distance. Distance Based on Patent Citations



Top panel: The figure plots the estimated coefficients on college share in each 1-digit industry (from Table A3, column 3) on the y-axis against the rank in the frequency of patent citations, on the x-axis (from Table A3, column 1). For example, manufacturing has rank 1, because the frequency of patent citations is highest for manufacturing. Services has rank 2, communication has rank 3, etc. The OLS fitted line is superimposed. The slope (standard error) of the line is -0.086 (0.051). R^2 is .35.
Bottom Panel: The figure plots the estimated coefficients on college share in each 2-digit industry group (from Table A3, column 3) on the y-axis against the rank in frequency of patent citations, on the x-axis (from Table A3, column 1). The OLS fitted line is superimposed. The slope (standard error) of the line is -0.007 (0.004). R^2 is .29.