Technology Diffusion and the Spatial Distribution of Wages in the US

by

Antonio Ciccone*

December 1999 (First Version: October 1997)

Summary: What explains the spatial distribution of wages across US counties? I find that two of the most important factors are spatial technology diffusion and externalities due to the aggregate scale of production. One empirical finding supporting the importance of spatial technology diffusion is that average wages in a county decrease with the average level of schooling in neighboring counties when employment in the county and average wages in neighboring counties are held constant. All empirical results are obtained using a novel instrument for (endogenous) employment at the county-level and take into account other factors (e.g. productivity-differences across states, climate) that may determine wages.

Keywords: US Spatial Wage Distribution, Spatial Technology Diffusion, Dynamic Spatial Externalities, Spatial Fixed Effects.
JEL Codes: O30, O40, R10.

---

* Universitat Pompeu Fabra. Email: Antonio.Ciccone@Econ.Upf.Es. Phone: 34-93-542-1669. Address: Department of Economics, Universitat Pompeu Fabra, C. Ramon Trias Fargas 25-27, 08005 Barcelona, Spain. This research has been supported by CREI at UPF. The paper could not have been written without Walter Garcia-Fontes' help with the data. Many thanks also to Douglas Almond outstanding research assistance, and to David Card, Ken Chay, and Jordi Gali for comments.

1 Introduction

Spatial differences in wages across the US are very large. Average wages in the private sector, without agriculture and mining, of the top 50 US counties in 1990 were more than three times average wages of the bottom 50 counties. Average wages of the top 300 counties, 10 percent of all US counties, were more than twice average wages of the bottom 300 counties. Average wages in New Jersey were one third greater than average wages in Illinois, and average wages in Illinois one third greater than average wages in South Dakota. What explains these spatial differences in wages? Research on differences in labor productivity across US cities and states suggests that spatial differences in wages may be partly determined by spatial differences in total factor productivity (TFP). Spatial differences in TFP may in turn be driven by spatial externalities due to the aggregate scale of production (Sveikaukas, 1975; Henderson, 1986), the average level of human capital (Lucas, 1988; Rauch 1993), or the density of economic activity (Ciccone and Hall, 1996). Which of these factors is most important is not clear. Other factors that may explain spatial differences in wages, but have not been explored so far, are related to the spatial diffusion of technology. For example, if technology spreads more rapidly among neighboring counties, then counties will have higher wages the higher levels of TFP of their neighbors. Furthermore, human capital externalities across neighboring counties imply that counties whose neighbors have high levels of human capital will have higher wages (Lucas, 1999).

I look at these issues using data on basically all US counties (2826) in 1990. My empirical results indicate that differences in TFP across counties are key for explaining differences in average wages across counties. Furthermore, the two main determinants of a county's level of TFP are the levels of TFP of neighboring counties and the aggregate scale of production in the county. One empirical finding supporting the importance of neighboring counties' levels of TFP is that average wages in a county decrease with the average level of schooling in neighboring counties when employment in the county and average wages in neighboring counties are held constant. This finding is robust and easily explained with spatial technology diffusion as higher schooling in neighboring counties implies lower TFP in neighboring counties when their wages are held constant. My point estimate of the long-run rate of spatial technology diffusion indicates that a 10 percent increase in levels of TFP of a county's neighbors increases wages in the county by 6 percent. Doubling the aggregate scale of production (measured by employment) in a county increases wages by 3.6 percent. I do not find evidence suggesting that counties with high schooling or high density of economic activity (measured by employment per acre) have high TFP. Schooling does increase wages however. One additional year of average schooling in a county increases average wages by 6.9 percent.

The main problem in estimating the effect of the aggregate scale of production or the density of economic activity on TFP at the county-level is reverse causation. Suppose for example that we suspect that high wages in some counties are driven by high TFP due to externalities from the density of economic activity. The reverse causation problem is that a positive partial correlation between average wages and employment per acre across counties may not necessarily indicate such externalities. Instead, the positive partial correlation may be due to workers moving to counties where TFP is high for reasons not controlled for. A natural way to identify the effect of employment per acre on wages would be an instrumental-variables approach. This requires variables that are correlated with employment-density across counties but uncorrelated with determinants of TFP not controlled for. However, no such variables have been identified at the county-level so far. I suggest that one possible candidate would be employment-density across US counties in the early 19th-century. Let me neglect for a moment that such data is not available and explain the logic of this instrument. There is little doubt that the spatial distribution of employment in the early 19th-century in the US was driven by factors that may have affected the spatial distribution of TFP then. For example, people migrating to North

America from Europe probably settled close to the eastern seaboard, railroads, natural ports, and river-crossings mainly because of lower transportation-costs given the transportation-technology. Furthermore, people moved to areas rich in natural resources or areas where agricultural productivity was high because of land-quality and climate. It seems however reasonable to suspect that most of the potential determinants of the spatial distribution of TFP in the early 19th-century (except maybe coastal location or climate, which can be controlled for) do not have much of an effect on TFP outside of agriculture and mining today. This is because today, railways, highways, and airports are available all over the US; the relative importance of agriculture and mining has much declined; air-conditioning and heating has improved and so on. In fact, these factors combined explain in part why production in the last 50 years or so grew rapidly in regions of the US that were sparsely populated in the early 20th-century.

This leaves the question of how to deal with the lack of data on employment-density at the county-level in the early 19th-century. I suggest a proxy. The variable used in the empirical work is the size of the administrative territory of counties—i.e. their total land-area. Total land-area of counties is a historically predetermined variable, as more than 95 percent of the counties in the US today already existed and had the same territory in the late 19th-century. The vast majority of counties is even older (Encyclopedia Americana, 1918). For example, the nine counties that make up the consolidated metropolitan area south and east of San Francisco (S.F.-Oakland-San Jose)—including Santa Clara County which contains Silicon Valley—have not changed since 1856 (Coy, 1917). Furthermore, the few adjustments to the land-area of US counties made during the late 19th-century and the 20th-century have to do with geographic considerations and are unrelated to population. This is evidenced by the fact that Los Angeles County, by far the most populated county today, has not been broken up despite the more than thousandfold increase in population.

Despite being historically pre-determined, total land-area of counties is strongly negatively correlated with current employment-density. Regressing the logarithm of employment-density of counties in 1990 on the logarithm of their total land-area and state dummies yields a highly significant coefficient on total land-area of -0.48. This is due to two main reasons. First, one of the criteria used for dividing up states into counties was the equalization of population-size of counties. General Vallejo, chairman of the 1850 commission that formed the first counties in California, stated so explicitly when asked to explain why El Dorado County, where gold was found in 1848, was so small and Los Angeles County, largely uninhabited, so large (Coy, 1917). This criterion induced a negative correlation between population-density and land-area of counties. Second, the spatial population-distribution in the US has proven to be persistent. This may be because populated places became increasingly productive as a result of externalities or because people acquired a taste for the location where they grew up. Equalization of population-size was however not the only criterion used to divide the territory into counties. Other criteria induced a positive correlation between population and land-area of counties that also endured. Regressing the logarithm of employment of counties in 1990 on the logarithm of their total land-area and on state dummies yields a highly significant coefficient on land-area of +0.57.

There seems to be no previous work analyzing the role of technology diffusion and spatial externalities for the wage-distribution across US counties.[1] Empirical studies of spatial externalities in the US generally focus on MSAs (metropolitan statistical areas; MSAs contain three counties on average) or states. Sveikauskas (1975), Segal (1976), and Moomaw (1981,1985) combine data on employment or population with data on output to estimate the elasticity of productivity with respect to employment or population. Henderson (1986) studies the elasticity of industry productivity with respect to industry employment at the same level of geographic detail. Glaeser et al. (1992) and Henderson et al. (1995) also focus on MSAs but

[1] Hanson (1998) is a very interesting contribution using data on counties. Following Krugman (1991), he relates (the change over time of) average wages and employment across counties to (the change over time of) income across states. His theoretical and empirical approach is very different from my approach however and therefore difficult to compare.

examine the determinants of employment growth rates instead of productivity. Rauch (1993) and Glaeser and Mare (1994) look at human capital externalities using data on individuals in MSAs. Ciccone and Hall (1996) explore the role of spatial externalities due to the density of economic activity for explaining differences in labor productivity across states.

The rest of the paper is organized in the following way. Section 2 presents the model and Section 3 discusses estimation. Section 4 describes the data and Section 5 presents results and specification tests. Section 6 summarizes.

2 The Model

The main features of the model that serves as a basis for estimation are summarized in the next subsection. The summary is followed by details on the derivation.

2.A Summary of the Model

The model reduces to two equations. The first is basically aggregate labor-demand across counties. The equation states that the log of wages across counties (the vector ***w***) increases with average years of schooling across counties (the vector ***SCHOOL***) and the log of TFP across counties (the vector ***TFP***) but decreases with the log of employment per acre of land across counties (the vector ***DEN***):

$$\boldsymbol{w} = \boldsymbol{TFP} + n\,\boldsymbol{SCHOOL} - b\,\boldsymbol{DEN}\ , \tag{1}$$

$n, b \geq 0$. It is clear why TFP and schooling enter positively. Density enters negatively because, given TFP and schooling, more employment per acre implies that less land is available per worker which lowers workers' productivity. The derivation of (1) assumes that markets for labor and land are competitive in each county and that all firms face the same rental cost of capital.

The second equation of the model captures the potential determinants of TFP across counties. To capture spatial externalities, log-TFP across counties is allowed to depend on log-employment (the vector ***EMP***), log-density, and schooling across counties as well as on the unweighted average level of schooling in neighboring counties ($\boldsymbol{NSCHOOL}$). To capture spatial technology diffusion, log-TFP across counties is also allowed to depend on the unweighted average level of log-TFP in neighboring counties ($\boldsymbol{NTFP}$):

$$\boldsymbol{TFP} = \boldsymbol{CNTR} + g\,\boldsymbol{EMP} + q\,\boldsymbol{DEN} + h\,\boldsymbol{SCHOOL} + f\,\boldsymbol{NSCHOOL} + s\,\boldsymbol{NTFP}\;; \qquad (2)$$

CNTR stands for other control variables. Equation (2) can be derived as the steady-state equation of a model with dynamic externalities in counties and technology diffusion across neighboring counties. Parameters on the right-hand side can therefore be interpreted as measuring long-run effects. For example, g can be interpreted as the long-run externality from employment in counties and s as the long-run degree of technology diffusion across neighboring counties.

The equation that will serve as the basis for estimation combines (1) and (2) with the fact that b is equal to the share of land in income paid to labor and land:

$$\begin{aligned} \boldsymbol{AdjW} &= \boldsymbol{CNTR} + g\,\boldsymbol{EMP} + q\,\boldsymbol{DEN} + (n+h)\boldsymbol{SCHOOL} \\ &\quad + s\,\boldsymbol{NAdjW} + (f - ns)\boldsymbol{NSCHOOL} + \boldsymbol{u} \end{aligned} \qquad (3)$$

where $\boldsymbol{AdjW} = \boldsymbol{w} + \hat{b}\,\boldsymbol{DEN}$, with $\hat{b}$ equal to the share of land in income paid to labor and land; ***NAdjW*** denotes the unweighted average of adjusted log-wages of neighbors; $\boldsymbol{u}$ will capture, among other things, differences in log-TFP across counties not explained by the model ("exogenous" TFP).

To better understand some of the implications of the model, suppose that average levels of schooling across counties do no matter for TFP, i.e. $h = f = 0$ in (2). In this case, (3) implies that a high average level of schooling of a county's neighbors should *decrease* adjusted wages in the county. This is an implication of spatial technology diffusion because the higher the average level of schooling of neighbors, the lower the average level of TFP of neighbors once their wages are controlled for. Spatial technology diffusion not only predicts the sign of the effect of average schooling of neighbors on adjusted log-wages but also the magnitude. In the case where $h = f = 0$, the effect of average schooling of neighbors on adjusted log-wages must be just offset by the effect of average adjusted log-wages of neighbors multiplied by the effect of schooling.[2] This restriction will not be satisfied however if there are strictly positive schooling externalities. To see this, denote the coefficient on ***SCHOOL*** with *B* and the coefficient on ***NSCHOOL*** with *C*. (3) implies that $s\,B + C = s\,h + f$. Schooling externalities in counties ($h > 0$) and across neighboring counties ($f > 0$) therefore imply that $C + s\,B$ will be strictly positive instead of zero. Notice also that if spatial technology diffusion is irrelevant, i.e. $s = 0$, then (3) predicts that high levels of schooling of neighbors *increase* adjusted wages in the county if there are schooling externalities across neighboring counties.

2.B Details of the Model

Assume that firms produce goods with labor, human capital, physical capital, and land. Their production function is $Y_{fc} = A_c (H_{fc} L_{fc})^{a(1-b)} K_{fc}^{ab} M_{fc}^{1-a}$ with $0 \le a \le 1$ and $0 \le b \le 1$ where Y_{fc} denotes output of firm f in county c; A_c denotes the level of TFP in county c; H_{fc} denotes the employment-weighted average quantity of human capital of firm f in county c; L_{fc} denotes employment; K_{fc} denotes the total quantity of physical capital; and M_{fc} denotes the total quantity of land. This production function implicitly assumes that there are constant returns to scale to labor, physical capital, and land at the firm-level. Assume also that firms in all counties face the same rental cost of physical capital. Output at the firm-level can then be written as

$$Y_{fc} = DA_c^{a} (H_{fc} L_{fc})^{1-b} M_{fc}^{b} \qquad (4)$$

where D depends on the cost of physical capital, and $a = 1/(1-ab)$ and $b = a(1-a)$.

Suppose that markets for labor and land are perfectly competitive in each county. Workers with the same level of human capital in the same county will therefore earn the same wage. I make no assumptions about wages of workers with the same level

[2] This is because levels of TFP of neighbors remain unchanged if log-wages of each neighbor are increased by 1 at the same time as levels of schooling are increased by $1/n$.

of human capital in different counties. They may or may not differ. How can counties be in equilibrium at different wages? The simplest answer is that lower prices of housing compensate workers in low-wage counties. Furthermore, different workers may value the degree of urbanization, air pollution, and crime of different counties as well as their climate or geographic characteristics in different ways. The assumption of competitive labor markets in each county combined with (4) implies that average labor productivity is proportional to average wages. Average wages across counties are therefore characterized by

$$\boldsymbol{w}_t = d\,\boldsymbol{1} + a\,\boldsymbol{a}_t + (1-b)\boldsymbol{h}_t - b\,\boldsymbol{d}_t = a\,\boldsymbol{a}_t + n\,\boldsymbol{S}_t - b\,\boldsymbol{d}_t \tag{5}$$

where $\boldsymbol{w}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{d}, \boldsymbol{S}$ denote vectors (with as many elements as counties in the sample) that collect the logarithm of average wages, the logarithm of levels of TFP, the logarithm of average levels of human capital, the logarithm of employment-densities L_{ct}/M_{ct}, and average years of schooling across counties. d is an unimportant constant and $\boldsymbol{1}$ denotes the unit-vector.[3] The second equality in (5) has been obtained by following Mincer (1974) in assuming that levels of human capital depend exponentially on years of schooling: $\boldsymbol{h} = l\,\boldsymbol{S}$. It is straightforward to show that b in (5) is equal to the share of land in income paid to labor and land.

Different theories in the literature suggest that the level of TFP in a county may be endogenous and driven by aggregate employment, aggregate employment-density, or aggregate human capital because of externalities. For some of these theories see Arrow (1962), Fujita (1989), Henderson (1986), Jacobs (1969), Lucas (1988), Marshall (1890), and Porter (1990).[4] The exposition that follows will for simplicity assume that externalities are driven by aggregate employment only.

[3] Details of this derivation can be found in the Appendix.

[4] Romer (1986) suggests that externalities are driven by the aggregate stock of physical capital. There is no data on physical capital at the county-level however. This is why I made use of the assumption that all firms face the same rental cost of capital. It can be shown that this approach implies that physical capital externalities will show up as externalities related to aggregate employment or average schooling in the county. This is because the complemetarity between labor and human capital on the one hand and physical capital on

Furthermore, the level of TFP in a county may also depend on levels of TFP in neighboring counties because of spatial technology diffusion. A simple and convenient way to capture the determinants of TFP in county c at time t is the following constant-elasticity specification

$$\boldsymbol{a}_t = \boldsymbol{f}_t + \Gamma(L)\boldsymbol{l}_{t-1} + \Sigma(L)\boldsymbol{N}\boldsymbol{a}_{t-1}, \tag{6}$$

where $\boldsymbol{l}$ denotes the vector that collects the logarithm of employment across counties and $\boldsymbol{f}$ captures difference in log-TFP unrelated to log-employment or log-TFP of neighboring counties. $\Gamma(L)$ denotes the distributed lag,

$$\Gamma(L) = (g_1 / a) + (g_2 / a)L + \cdots + (g_G / a)L^G, \tag{7}$$

and captures that externalities may take some time to set in. $\boldsymbol{N}$ in (6) denotes a square matrix collecting weights w_{cm}. These weights will be equal to the inverse of the number of county c's neighbors if county c and m are neighbors. $\boldsymbol{N}\boldsymbol{a}_{t-g}$ therefore denotes the unweighted average level of log-TFP of each county's neighbors at time $t-g$. Finally, $\Sigma(L)$ denotes

$$\Sigma(L) = s_1 + s_2 L + \cdots + s_I L^I \tag{8}$$

and captures lags in technology diffusion. According to (6), TFP in county c at time t may depend on past employment in the county as well as past TFP in neighboring counties. The specification implies that the elasticity of TFP in county c at time t with respect to employment g periods back in the same county is g_g / a; the elasticity with respect to TFP i periods back in neighboring county m is $s_i w_{cm}$.

Denote the long-run effect of employment on TFP in (6) with $g / a = \sum_{g=1}^{G} g_g / a$ and the long-run effect of the average level of TFP of neighboring counties with $s = \sum_{i=1}^{I} s_i$. These long-run effects can be estimated with data across counties. To see how, suppose that $\boldsymbol{f}_t$ and $\boldsymbol{l}_t$ follow stationary stochastic processes around a

---

the other in the production function implies that counties with higher employment and levels of schooling will also have higher levels of physical capital.

trend where exogenous TFP in all counties grows at the same rate and employment in all counties grows at same rate. Then—under a technical condition ensuring that the effect of past TFP on current TFP in (6) is not too strong—(5) and (6) can be combined to

$$\mathbf{w}_t + b\,\mathbf{d}_t = d\,\mathbf{1} + g\,\mathbf{l}_t + v\mathbf{S}_t + s\,\mathbf{N}\left(\mathbf{w}_t + b\,\mathbf{d}_t - v\mathbf{S}_t\right) + a\,\mathbf{f}_t + \mathbf{v}_t \;, \tag{9}$$

where $\mathbf{v}_t$ depends on the deviations of current and past values of TFP and employment from trend.[5] Average wages will therefore be positively related to employment if the long-run externality of employment is positive $g > 0$; average wages will be positively related to the average level of TFP of neighboring counties if the parameter capturing long-run technology diffusion is positive $s > 0$.

To see in some more detail how (9) can be derived, suppose that $\mathbf{f}_t - E_t\mathbf{f}_t$ and $\mathbf{l}_t - E_t\mathbf{l}_t$ follow stationary stochastic processes. Assume also that different counties have expected levels of exogenous technology and employment given by $E_t\mathbf{f}_t = \mathbf{f} + k\,\mathbf{1}t$ and $E_t\mathbf{l}_t = \mathbf{l} + j\,\mathbf{1}t$. Then, (6) implies under the aforementioned technical condition that $E_t\mathbf{a}_t$ will converge to a steady-state $\mathbf{a}_t^*$ given by $\mathbf{a}_t^* = E_t\mathbf{f}_t + g\,E_t\mathbf{l}_t + s\,\mathbf{N}\mathbf{a}_t^*$. This equation basically states that the strength of long-run externalities and the long-run degree of technology diffusion show up as static externalities and technological interdependence in steady-state. The equation cannot be implemented empirically however because we neither observe the relevant expectations nor $\mathbf{a}_t^*$. Substituting the actual variables into the last equation yields a new, blurry version

$$\mathbf{a}_t = \mathbf{f}_t + g\,\mathbf{l}_t + s\,\mathbf{N}\mathbf{a}_t + \mathbf{u}_t; \tag{10}$$

$\mathbf{u}_t$ arises because we do not observe the relevant expectations but only realized values. Also, $E_t\mathbf{a}_t$ is never exactly equal to the steady-state value $\mathbf{a}_t^*$. Inferring TFP from (5) and substituting in (10) yields (9).

[5] The condition is that the solutions of $\left|\mathbf{I} - s_1\mathbf{N}z - s_2\mathbf{N}z^2 - \cdots - s_l\mathbf{N}z^l\right| = 0$ lie outside the unit-circle.

3 Estimation

The equation that will serve as the basis for estimation is (3). The control variables used are 44 state dummies (one dummy for each state in the sample) and a dummy for Washington DC, 17 CMSA dummies (one dummy for each consolidated metropolitan statistical area in the sample), a MSA dummy, and 24 variables describing the climate in each county. These dummies allow counties in different states and CMSAs as well as counties in an MSA to have different levels of TFP and hence different wages for reasons that are unrelated to employment, density, or schooling in the county or in neighboring counties. Estimation also includes an Ocean-Great Lakes dummy to allow for exogenous TFP differences that arise because some counties touch on the Atlantic Ocean, the Pacific Ocean, or the Great Lakes.

The main problem in estimating (3) is that employment across counties is endogenous. Furthermore, the adjusted average wage of neighbors is also endogenous. The method of estimation used will therefore be based on instrumental variables. Instruments used for estimation are total land-area, average total land-area of neighbors, years of schooling, average years of schooling of neighbors, and the control variables. The implicit assumption is that land-area and schooling are unrelated to TFP not explained by (2), i.e. neither explained by the control variables (climate, state/Washington DC dummies, and CMSA/MSA dummies) nor explained by externalities associated with employment, employment-density, schooling, neighbors' schooling, or neighbors' TFP. Furthermore, land-area and schooling must also be unrelated to $\boldsymbol{u}_t$ in (10). The assumption that schooling can be used as an instrument can be tested in a restricted version of the model.[6] It is also possible to estimate a restricted version of the model without schooling as an instrument. Notice

[6] It may be worthwhile to emphasize that to use average years of schooling across counties as an instrument it is not necessary to assume that schooling across counties is unrelated to TFP across counties. Schooling may be related to differences in TFP captured by (2).

that the number of instruments available is insufficient to estimate the unrestricted version of (3). Estimation will therefore be based on restricted versions.

Another concern in the estimation of (3) is heteroskedasticity and spatial correlation in exogenous TFP as captured by $\boldsymbol{u}$. The preferred method of estimation is therefore generalized two-stage least-squares (G2SLS).

4 Data

All the data required for estimation of (3) is available for 1990. The data covers total land-area, average wage level, level of employment, and average years of schooling by county. Average years of schooling is prepared by the *U.S. Bureau of Census* (1992) and comes from the *1990 Census of Population*. The data on wages, employment, and area by county is prepared by the *Regional Economic Measurement Division* of the *U.S. Bureau of Economic Analysis* (1993). Wages reflect firms' cost of labor: wages are measured before deductions such as social security and union dues, and include commissions, bonuses, tips, and pay-in-kind; they also include employers' contributions to pension benefits, social security, and health benefits. Employment is measured as the number of full-time and part-time jobs.[7] Average wages and employment used are for the private economy without agriculture and mining, as the model assumes that all land is equivalent.

The share of land in income paid to labor and land in the private sector without agriculture and mining required to construct adjusted wages is calculated as 0.5 percent from the *Flow of Funds Accounts of the United States, 1982-1990* prepared by the *Board of Governors of the Federal Reserve System* (1997). Varying the value between 0 and 1.5 percent changes the estimates of all other parameters very little.

The climate variables for counties are 24 variables for average daily temperature and rainfall by month. The averages are calculated over the period 1890 to 1980 and are available upon request from the *National Climatic Data Center*.

[7] This gives rise to measurement error, another reason to use instruments for employment.

Counties in the sample cover Washington DC and all continental states except Alaska, Louisiana, West Virginia, Wyoming, and Virginia. Virginia has been excluded because the data on average wages and employment could not be matched to the data on the area of counties. The other states have economies that depend heavily on the mining sector: mining contributes on average 21 percent to value added in these states, compared to 1 percent in all other states (Ciccone and Hall, 1996). The total number of counties in the sample is 2826. Two counties are defined as neighbors if they have a common border. Counties have 5 neighbors on average.

Spatial differences in wages are large. The states with the lowest average wages are South Dakota, 33 percent below the national average; Nebraska, 29 percent below the national average; Missouri, 27 percent below the national average; and North Dakota, 26 percent below the national average. New Jersey and Connecticut are the states with highest average wages, 24 and 19 percent respectively above the national average. Differences in wages across CMSAs are also large. The CMSA around Detroit has average wages that are 35 percent higher than average wages in Michigan and 24 percent higher than the national average. Average wages in the CMSA around New York City are 31 percent higher than average wages in New York State and 26 percent higher than the national average. The frequency distribution of the logarithm of average wages across counties in the sample is plotted in the Appendix.

5 Results

A useful starting point is estimation of (3) assuming that TFP in each county is unrelated to TFP or schooling in neighboring counties. Table 1 presents GLS-estimates of the key parameters. All parameters in the table are significant at the one-percent level. The three groups of control variables (not in the table) for TFP-differences across states and Washington DC, for TFP-differences across CMSAs, and for TFP-differences driven by climate are each jointly significant at the 0.1-percent level. The MSA dummy is significant at the one-percent level. The Ocean-

Great Lakes dummy is insignificant at the 1-percent level. Combining the estimates of the effect of employment-density and employment on adjusted wages suggests that doubling employment in a county increases average wages by 7.2 percent. This estimate may overstate the true effect of employment on wages however as employment is probably positively related to exogenous TFP. It is for this reason that I re-estimate the equation of Table 1 using G2SLS with area and average area of neighbors as instruments for density and employment. Table 2 indicates that the effect of density becomes very small and is no longer significant when instruments are used.[8] Employment remains significant but is only 60 percent of the value estimated in Table 1. These estimates suggest that the GLS-estimates in Table 1 suffered from endogeneity bias. It is interesting to note that the estimated effect of schooling on adjusted wages increases using G2SLS. The effect of one additional year of schooling on average wages is estimated to be 11 percent. This is somewhat higher than the effect of schooling estimated in individual wage-regressions (Card, 1999) and may therefore suggest schooling externalities. The control variables for states plus Washington DC, CMSAs, and climate remain significant at the 0.1-percent level. The MSA dummy remains significant at the five-percent level. The Ocean-Great Lakes dummy stays insignificant at the 1-percent level.

Table 3 re-estimates the equation of Table 2 without density as a determinant of TFP, but allows for technology diffusion across neighboring counties, as well as schooling externalities across neighboring counties. The method used is G2SLS and the full set of control variables is included. The results show that all key variables are significant at the five-percent level. It is interesting to note that average schooling of neighbors decreases adjusted wages as predicted by the model with technology diffusion. Technology diffusion predicts a negative effect because, holding average wages of neighbors constant, a higher average level of schooling in neighboring counties implies a lower level of TFP in neighboring counties.

[8] The R^2 of the regressions of the endogenous right-hand side variables on the set of instruments is never smaller than 60 percent (see the Tables in the Appendix for details).

However, the results in Table 3 also indicate that parameters are estimated rather imprecisely. It turns out that this is due to the inclusion of CMSA and MSA dummies in the estimation. The coefficients on these dummies are small and highly insignificant. For example, the MSA dummy is estimated to be 2.2 percent with a standard error of 2.5 percent. The employment-weighted absolute value of the estimated CMSA dummies is 5 percent and the hypothesis that all CMSA dummies are zero has a P-Value of 0.85. For comparison, in the model without technology diffusion the G2SLS-estimate of the MSA dummy is 5.8 percent with a standard error of 2.1 percent; the employment-weighted absolute value of the estimated CMSA dummies is 22 percent; and the hypothesis that all CMSA dummies are zero has a P-Value of 0.001. The comparison suggests that technology diffusion explains much of the variation in average wages that would be attributed to CMSA and MSA dummies in a model without technology diffusion.

Re-estimating the equation of Table 3 without the insignificant CMSA and MSA dummies using G2SLS yields the results in Table 4. All parameters are now estimated more precisely. The point-estimates of the effect of employment and schooling fall and the point-estimates of the effect of average adjusted wages of neighbors and average schooling of neighbors are also lower than in Table 3. Average schooling of neighbors continues to have a negative effect on adjusted wages as predicted by the technology diffusion model. The Ocean-Great Lakes dummy stays insignificant at the 1-percent level and will therefore also be dropped.

There is a simple way to test for the significance of schooling externalities in the equation of Table 3. Denote the coefficient on ***SCHOOL*** with *B* and the coefficient on ***NSCHOOL*** with *C*. The equation of Table 4 implies that $s\,B + C = 0$ is equivalent to $f + s\,h = 0$. We should therefore be unable to reject $s\,B + C = 0$ if there are no schooling externalities (neither from schooling in the county nor from schooling in neighboring counties). The point estimate of $s\,B + C$ implied by Table 4 is -0.016. The null-hypothesis $s\,B + C = 0$ has a P-value of 0.28 and can therefore not be rejected at conventional significance levels.

Table 5 re-estimates the equation of Table 4 under the assumption that $C = -s B$. The resulting estimate of the effect of the aggregate scale of production on wages is 0.036. Doubling employment in a county increases wages by 3.6 percent.[9] The estimate of the long-run degree of technology diffusion between neighboring counties is 0.62. This implies that a 10 percent increase in the level of TFP of a county's neighbors eventually increases wages in the county by 6.2 percent. The effect of one additional year of schooling in a county on average wages is 6.9 percent. It may be worthwhile to point out that this estimate is similar to the return to schooling estimated in individual, Mincerian wage-regressions (Card, 1999). The equation of Table 5 can be re-estimated without schooling as an instrument once the effect of schooling on wages is restricted to be equal to the estimate in individual wage-regressions. Setting $n = 0.08$ (the middle of the range estimated in individual wage-regressions) and re-estimating does not change the effect of employment or average TFP of neighbors on adjusted wages of counties.

The equation estimated in Table 5 explains 49 percent of the large variation in wages across US counties. The hypothesis that all climate variables are equal to zero has a P-value of 11.6. The hypothesis that all state dummies are equal to zero has a P-value of 9.6. The employment-weighted average of the absolute values of state dummies is 4 percent. It is interesting to compare the last two results to the estimates and significance of the state dummies in the model without technology diffusion. There, the hypothesis that all state dummies are zero has a P-value of 0.001 and the employment-weighted absolute value of the state dummies is 15 percent (four times greater). Technology diffusion therefore appears to explain much of the variation in average wages that would be attributed to state dummies in a model without technology diffusion.

Table 6 re-estimates the equation of Table 5 with employment-density as an additional determinant of wages. The result is an estimate of the effect of density on

[9] The effect of employment on wages increases when state dummies are dropped. Re-estimating the equation of Table 4 without state dummies yields an estimate of 0.052.

wages of –0.001 with a standard error of 0.008 percent. All other estimates are basically unchanged. These results differ from those obtained by Ciccone and Hall (1996). They found density at the county-level to be key in explaining differences in average labor productivity across states.[10] The discrepancy is mainly due to the fact that the results in Table 6 control for differences in TFP across states while Ciccone and Hall's approach did not allow them to do so. Another difference with their results is that schooling at the county-level is significant in Table 5 while they find schooling to be an insignificant determinant of average labor productivity across states.

The robustness of the results in Table 5 can be checked by including additional regressors in the estimating equation. One possibility is to include average adjusted wages of counties that are not neighbors but neighbors of neighbors. Including this variable makes it possible to check to what extent technology diffusion works through neighbors. It turns out that average adjusted wages of counties that are not neighbors but neighbors of neighbors enter with a small, negative coefficient and are insignificant at the two-percent level. All other coefficients remain significant at the two-percent level. Another possibility is to include total wages of neighbors. This makes it possible to check whether spillovers between neighboring counties are linked to the size (employment) of neighbors. The total wages of neighbors enter with a small, positive coefficient but are insignificant at the three-percent level. All other coefficients remain significant at the four-percent level.

The residuals of the first step of the G2SLS-estimation of the equation in Table 5 can be used to test for heteroskedasticity and spatial correlation in exogenous TFP across counties. There is evidence for heteroskedasticity of the residuals associated with presence in different states. The estimated variance of the residuals by state is in fact used in the second step of the G2SLS-estimation. There is however no evidence of any significant correlation between the residuals of counties and the

[10] The correlation between average wages and average labor productivity across states is 0.92.

average residuals of their neighbors. Regressing the residual of counties on the average residuals of their neighbors yields a small, positive estimate that is insignificant at the one-percent level.

The equation of Table 5 is overidentified. This can be used to test the specification using two standard tests. The test of overidentifying restrictions described by Gallant and Jorgenson (1979) and the test of exogeneity of instruments described by Hausman (1983). The test of overidentifying restrictions yields that the overidentifying restrictions cannot be rejected at the 35 percent significance level. The test of exogeneity yields that exogeneity of the instruments cannot be rejected at the 27 percent significance level.

I also explore two alternative specifications for the matrix $\boldsymbol{N}$ by weighting the TFP of neighbors by their share in employment and wages of all neighbors. The results change very little. In particular, density remains insignificant and there are still increasing returns to employment.

6 Summary

The paper tackles two main issues. The first is whether total factor productivity (TFP) at the county-level is driven by the aggregate scale of production, the average level of schooling, or the density of economic activity. The second issue is whether the distribution of wages across US counties provides evidence for spatial technology diffusion. The empirical findings in the paper suggest that the spatial distribution of TFP is driven by the aggregate scale of production. Furthermore, there is evidence of a substantial degree of spatial technology diffusion. One of the findings supporting this view is that average wages in a county decrease with the average level of schooling of neighboring counties once employment in the county and average wages in neighboring counties are controlled for. This finding is robust and consistent with theories of spatial technology diffusion as higher average levels of schooling in neighboring counties are equivalent to lower levels of TFP when neighbors' average wages are held constant. Point-estimates suggest that a 10 percent increase in the TFP of a county's neighbors increases wages in the county by 6 percent. Doubling employment in the county increases wages by 3.6 percent. The model with externalities driven by aggregate employment at the county-level and technology diffusion across neighboring counties developed in the paper explains half of the large variation of wages across counties.

References

Arrow, Kenneth (1962), "The Economic Implications of Learning by Doing," *Review of Economic Studies* 29, pp. 155-173.

Board of Governors of the Federal Reserve System (1997), *Flow of Funds Accounts of the United States*, *1982-1990*, Washington DC.

Card, David (1999), "Education in the Labor Market," in Orley Ashenfelter and David Card, *Handbook of Labor Economics*, North Holland, Amsterdam.

Ciccone, Antonio and Robert E. Hall (1996), "Productivity and the Density of Economic Activity," *American Economic Review* 86, pp. 54-70.

Coy, Owen (1917), *California County Boundaries*, Valley Publishers, Fresno, California.

Encyclopedia Americana (1918), The Encyclopedia Americana Corporation, New York.

Fujita, Masahisa (1989), *Urban Economic Theory*. Cambridge: Cambridge University Press.

Gallant, Ronald, and Dale Jorgenson (1979), "Statistical Inference for a System of Simultaneous, Non-linear, Implicit Equations in the Context of Instrumental Variable Estimation," *Journal of Econometrics* 11, pp. 275-303.

Glaeser, Edward, Heidi Kallal, Jose Scheinkman, and Andrei Shleifer (1992), "Growth in Cities," *Journal of Political Economy*, Volume 100, pp. 142-163.

Glaeser, Edward and David Mare (1994), "Cities and Skills," Hoover Institution Working Paper Series.

Hanson, Gordon (1998), "Market Potential, Increasing Returns, and Geographic Concentration," NBER WP 6429, Cambridge, Massachusetts.

Hausman, Jerry (1983), "Specification and Estimation of Simultaneous Equation Models," in *Handbook of Econometrics* edited by Zvi Griliches and Michael Intriligator, North Holland, Amsterdam.

Henderson, Vernon (1986), "Efficiency of Resource Usage and City Size," *Journal of Urban Economics* 18, pp. 23-56.

Henderson, Vernon, Ari Kuncoro, and Matt Turner (1995), "Industrial Development in Cities," *Journal of Political Economy* 103, pp. 1067-1081.

Jacobs, Jane (1969), *The Economy of Cities*, Vintage Books, New York.

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

Lucas, Robert E. (1988), "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22, pp. 3-41.

Lucas, Robert E. (1993), "Making A Miracle," *Econometrica* 61, pp. 251-272.

Lucas, Robert E. (1999), "Externalities and Cities," mimeo, Department of Economics, University of Chicago.

Marshall, Alfred (1890), *Principles of Economics*, Macmillan, London.

Mincer, Jacob (1974), *Schooling, Experience, and Earnings*, Columbia University Press, New York.

Moomaw, Ronald (1981), "Productivity and City Size: A Critique of the Evidence," *Quarterly Journal of Economics* 95, pp. 194-223.

Moomaw, Ronald (1985), "Firm Location and City Size," *Regional Science and Urban Economics* 18, pp. 45-78.

Porter, Michael (1990), *The Competitive Advantage of Nations*, Free Press, New York.

Rauch, James (1993), "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities," *Journal of Urban Economics* 34, pp. 380-400.

Romer, Paul (1986), "Increasing Returns and Long Run Growth," *Journal of Political Economy* 98, pp. S71-S102.

Segal, David (1976), "Are there Returns to City Size?" *Review of Economics and Statistics* 58, pp. 234-251.

Sveikauskas, Leo (1975), "The Productivity of Cities," *Quarterly Journal of Economics* 89, pp. 393-413.

US Bureau of Census, *Data Users Services Division* (1992), *Census of Population and Housing, 1990: Summary Tape File 3A*, Washington, DC.

US Bureau of Economic Analysis, Regional Economic Measurement Division (1993), *REIS: Regional Economic Information System,* CD-ROM, Washington, DC.

Appendix

Appendix A: Results

Table 1: GLS-estimates of the model without technology diffusion

	Coefficient	Std. Error
DEN	0.026	0.006
EMP	0.046	0.006
SCHOOL	0.028	0.005

Notes: Results of estimating $\boldsymbol{AdjW} = \boldsymbol{CNTR} + g\,\boldsymbol{EMP} + q\,\boldsymbol{DEN} + (n + h)\boldsymbol{SCHOOL}$ where $\boldsymbol{AdjW} = \boldsymbol{w} + \hat{b}\,\boldsymbol{DEN}$ and $\hat{b} = 0.005$. There are 2826 observations. The method used is GLS. CNRT variables are 24 variables for climate, 44 state dummies and a dummy for Washington DC, 17 CMSA dummies, a MSA dummy, and an Ocean-Great Lakes dummy.

Table 2: G2SLS-estimates of the model without technology diffusion

	Coefficient	Std. Error
DEN	0.0002	0.008
EMP	0.027	0.009
SCHOOL	0.11	0.017

Notes: Results of estimating $\boldsymbol{AdjW} = \boldsymbol{CNTR} + g\,\boldsymbol{EMP} + q\,\boldsymbol{DEN} + (n + h)\boldsymbol{SCHOOL}$ where $\boldsymbol{AdjW} = \boldsymbol{w} + \hat{b}\,\boldsymbol{DEN}$ and $\hat{b} = 0.005$. There are 2826 observations. The method used is G2SLS. CNRT variables are 24 variables for climate, 44 state dummies and a dummy for Washington DC, 17 CMSA dummies, a MSA dummy, and an Ocean-Great Lakes dummy. The instruments used are the CNTR variables, area of each county, the average area of counties' neighbors, and schooling. The R^2 of the regression of ***EMP*** on the instruments is 62 percent. The R^2 of the regression of ***DEN*** on the instruments is 70 percent.

Table 3: A first model with technology diffusion

	Coefficient	Std. Error
EMP	0.044	0.012
SCHOOL	0.069	0.035
NadjW	1.25	0.51
NSCHOOL	-0.11	0.05

Notes: Results of estimating $\boldsymbol{AdjW} = \boldsymbol{CNTR} + g\,\boldsymbol{EMP} + (n + h)\boldsymbol{SCHOOL} + s\,\boldsymbol{NAdjW} + (f - sn)\boldsymbol{NSCHOOL}$ where $\boldsymbol{AdjW} = \boldsymbol{w} + \hat{b}\,\boldsymbol{DEN}$ and $\hat{b} = 0.005$. There are 2826 observations. The method used is G2SLS. CNRT variables are 24 variables for climate, 44 state dummies and a dummy for Washington DC, 17 CMSA dummies, a MSA dummy, and an Ocean-Great Lakes dummy. The instruments used are the CNTR variables, area of each county, the average area of counties' neighbors, schooling, and average schooling of counties' neighbors. The R^2 of the regression of ***NAdjW*** on the instruments is 65 percent.

Table 4: The model with technology diffusion without CMSA, MSA dummies

	Coefficient	**Std. Error**
EMP	0.035	0.007
SCHOOL	0.064	0.022
NadjW	0.65	0.12
NSCHOOL	-0.057	0.018

Notes: Results of estimating $\boldsymbol{AdjW} = \boldsymbol{CNTR} + g\,\boldsymbol{EMP} + (n + h)\boldsymbol{SCHOOL} + s\,\boldsymbol{NAdjW} + (f - sn)\boldsymbol{NSCHOOL}$ where $\boldsymbol{AdjW} = \boldsymbol{w} + \hat{b}\,\boldsymbol{DEN}$ and $\hat{b} = 0.005$. There are 2826 observations. The method used is G2SLS. CNRT variables are 24 variables for climate, an Ocean Great-Lakes dummy, and 44 state dummies plus a dummy for Washington DC. The instruments used are the CNTR variables, area of each county, the average area of counties' neighbors, schooling, and average schooling of counties' neighbors.

Table 5: The model with technology diffusion only

	Coefficient	**Std. Error**
EMP	0.036	0.007
SCHOOL	0.069	0.021
NadjW	0.62	0.09

Notes: Results of estimating $\boldsymbol{AdjW} = \boldsymbol{CNTR} + g\,\boldsymbol{EMP} + n\,\boldsymbol{SCHOOL} + s\,\boldsymbol{NAdjW} - sn\,\boldsymbol{NSCHOOL}$ where $\boldsymbol{AdjW} = \boldsymbol{w} + \hat{b}\,\boldsymbol{DEN}$ and $\hat{b} = 0.005$. There are 2826 observations. The method used is G2SLS. CNRT variables are 24 variables for climate, and 44 state dummies plus a dummy for Washington DC. The instruments used are the CNTR variables, area of each county, the average area of counties' neighbors, schooling, and average schooling of counties' neighbors.

Table 6: The model with technology diffusion and density

	Coefficient	**Std. Error**
EMP	0.038	0.009
DEN	-0.001	0.008
SCHOOL	0.071	0.022
NadjW	0.62	0.1

Notes: Results of estimating $\boldsymbol{AdjW} = \boldsymbol{CNTR} + g\,\boldsymbol{EMP} + q\,\boldsymbol{DEN} + n\,\boldsymbol{SCHOOL} + s\,\boldsymbol{NAdjW} - sn\,\boldsymbol{NSCHOOL}$ where $\boldsymbol{AdjW} = \boldsymbol{w} + \hat{b}\,\boldsymbol{DEN}$ and $\hat{b} = 0.005$. There are 2826 observations. The method used is G2SLS. CNRT variables are 24 variables for climate and 44 state dummies plus a dummy for Washington DC. The instruments used are the CNTR variables, area of each county, the average area of counties' neighbors, schooling, and average schooling of counties' neighbors.

Appendix B: Derivation of Equation (5)

The capital demand equation is

$$K_{fc} = abY_{fc} / r\ , \tag{A1}$$

where r denotes the national rental cost of capital. The land demand equation is $M_{fc} = (1-a)Y_{fc} / r_{cN}$ where r_{cN} is the rental cost of land in county c. Aggregation implies

$$M_{fc} = M_c Y_{fc} / Y_c\ . \tag{A2}$$

Making use of the capital demand equation (A1) and the land demand equation (A2) in the production function and aggregating yields

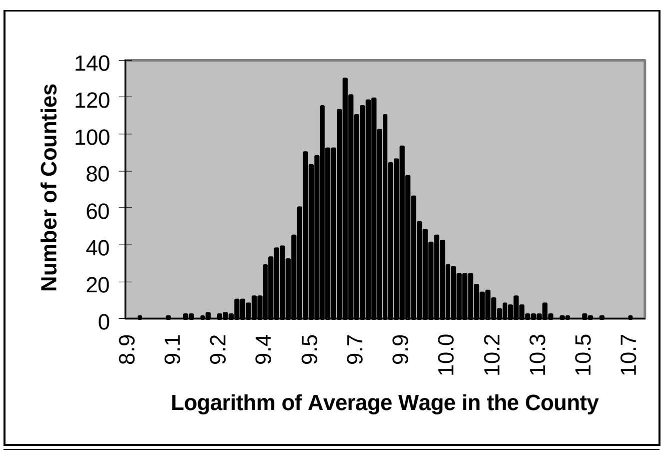
$$\frac{Y_c}{L_c} = (A_c)^{\frac{1}{1-ab}} (H_c)^{\frac{a(1-b)}{1-ab}} \left(\frac{ab}{r}\right)^{\frac{b}{1-b}} \left(\frac{L_c}{M_c}\right)^{-\frac{1-a}{1-ab}}\ . \tag{A3}$$

(A1), (A2), and constant returns to scale of the production function imply $w_c L_{fc} = a(1-b)Y_{fc}$ where w_c denotes average wages in county c. Aggregation yields

$$w_c = a(1-b)\frac{Y_c}{L_c}\ . \tag{A4}$$

(A3) and (A4) combined yield (5) in the main text.

Appendix C: Frequency of Average Wages Across 2826 US Counties in 1990



Notes: In 1990 US$.