

# The “Horizon Effect” and Implications for Rural Human Capital Development<sup>1</sup>

By  
Maureen Kilkenny  
Dept. of Economics  
Iowa State University

**Abstract:** In this paper we pose and test a hypothesis about the dependence of learning on size. A widely-held view is that productivity is low in rural areas, either because people with high value human capital migrate to cities where they can earn a better return, or, because low density areas cannot benefit from the technology spillovers that hypothetically depend on proximity or population density. In contrast, the ‘horizon effect’ hypothesis is that if outcomes depend on effort, and if the effort of members in a group depends positively on the frequency of observations of high performances in the group (their “view of the horizon”); given the statistical fact that larger samples contain more observations of high performance, the average performance of larger groups would be higher. We find evidence consistent with this hypothesis, but with a special twist relevant to rural communities.

In particular, we pose and reject a null hypothesis that the freshman math proficiency rate among high school sophomores is independent of the cross-sectional cohort size, measured by the number of students per grade, and the longitudinal size, measured by the number of grades in the school; controlling for the overall size of the school, spending per pupil on instruction and infrastructure; classroom and teacher characteristics, and local demographics, and spatial autocorrelation. Our spatial econometric analysis of building-level data from Ohio shows that there are robust positive and significant effects of the number of grades in a school. This evidence is consistent with a “horizon effect.” It is also consistent with the one-room-rural-school-house model. The ‘Horizon Effect’ analysis presented in this paper both confirms some antique ideas and may provide hope for human capital development in small town America.

---

<sup>1</sup> Paper for the session “Globalization, Technological Change, Metropolitan Places, and the Hinterlands,” in the symposium “Research Frontiers in Resources and Rural Economics” to celebrate Emery N. Castle; Oregon State University, Corvallis, OR, Oct 5-7, 2006. Many thanks to Monica Haddad, co-author on a closely related paper, for estimating the spatial lag model; Jim LeSage and Nicole Penzler for the data; Peter Orazem, and Justin Tobias for everything else.

## **The “Horizon Effect” and Implications for Rural Human Capital Development**

By Maureen Kilkenny

### ***Motivation***

How can rural areas and small towns compete in the high-tech, global economy in which the majority of the world’s customers reside in metropolitan areas? How will young people who grow up in remote, low population density places that are dependent upon natural resource industries (Castle, 1991) learn how to anticipate the next ‘new-new’ thing-- the goods or services that the rest of the world will want to buy? This paper is about how we learn and how learning in a place depends on population density or proximity to variety. There has been a preoccupation in the literature with ideas that city people are more productive and innovative, that productivity depends positively on population density (Ciccone and Hall, 1996). Does that mean rural areas are doomed to lag?

Not necessarily! A much more compelling explanation for the higher measured productivity in cities has recently been provided by Syverson , 2004 (see also Jovanovic, 1982; Breshnahan and Reiss, 1991; Aghion and Howitt, 1992; Nickell, 1996; Olley and Pakes, 1996; Campbell and Hopenhayn, 2002; and Raith, 2003). The essence of the mechanism elaborated by Syverson is that heightened competition in denser markets makes it harder for inefficient producers to profitably operate. This truncates the lower end of the productivity distribution, so that the measured average is higher and measured variance is lower. Syverson’s careful empirical work using data on the concrete-mixing industry provides robust statistical support of this hypothesis.

Syverson’s theory offers plenty of hope for rural areas. There is no particular reason why a remote, low density place can’t be an innovative place. And to prosper, a place must be innovative. As we know, the way to accumulate capital in a competitive market is to be a first mover, to operate in the early stage of the product cycle (Vernon, 1966; Duranton and Puga, 2001). After that, competition, first local and later global, drives prices down to marginal cost.

The basic stylization of the product cycle, a theory that claims that “you can tell where a product is being produced if all you know is its age” (Magee, 1980) implies that rural areas aren’t likely to be an ideal business location until after a product has established a market niche and the industry needs the relatively abundant low cost labor and space of rural locations. Innovations are, however, the applications of existing inventions to new purposes. Variety, as Jane Jacobs (1984) points out, provides wider exposure. Duranton and Puga (2001) show how larger places that can support more firms and thus more variety may be more innovative. Rural development specialists must ask if there is any way for small places to mimic the breadth of knowledge they need to be innovative.

These are the questions that motivated our research on the dependence of learning on size. We look at schools because schools vary in size and configuration and location, among many other things, and because data on student performance is widely available (e.g., Hanushek, 2002). In this paper we begin to investigate hypotheses related to the question “Is bigger better?” We consider two ways to measure “size”: cross-section and longitudinal. By “cross-section” size we mean, for example, population at any point in time. In a school, cross-section size is the average number of students per grade cohort in the building. This corresponds to population density in a place. By “longitudinal” we mean, for example, the same of many small populations over time. In a school, the longitudinal size is the number of grades in the same building. This corresponds to variety in a place. Does either type of size of one’s learning environment matter? If we can answer that question, it may shed light on ways that rural people can enhance their learning environments and achieve greater global competitiveness.

We test a hypothesis we call the “horizon effect.” Briefly, we assume that the average performance of learners may depend on how much effort each learner exerts (and, in the classroom context) how much their teachers expect and challenge them. We then hypothesize that student effort and teacher expectations may depend positively on the frequency of top performances observed by the students and teachers. Given that there is a larger number of

observations of top performances in a larger group, this leads to an inference that larger groups display higher averages.

To address the question empirically we analyze data about the percent of students proficient in Math in hundreds of Ohio high school buildings of different sizes and types. We find significant evidence of the longitudinal version of the “horizon effect.” When we also control for local demographics and spatial autocorrelation, we find that our hypotheses about horizon effects are robust, and that previously undetected class size effects are apparent. And, like most of the modern analyses of student educational outcomes, we have difficulty identifying the effect of spending (higher salaries to teachers, for example) on learner outcomes. The most significant predictors of learning performance are neighborhood demographics, parental education (positive), size of the school (negative), and the number of grades in the school (positive).

The paper proceeds as follows. After a brief review of the economics literature on class and school size, and ‘peer effects’ in school outcomes, we develop our new hypotheses about ‘horizon’ effects. Next we describe the data, our analytical procedures, and our findings. We conclude with a discussion of the implications for rural communities.

### ***Class Size***

Class size is but one of many potentially relevant measures of the inputs into learning outcomes.

Quoting Hanushek, 2002:

“Studies of educational performance include a variety of different measures of resources devoted to schools. Commonly employed measures include 1) the real resources of the classroom (teacher education, teacher experience, and class size or teacher-pupil ratios); 2) financial aggregates of resources (expenditure per student and teacher salary); and 3) measures of other resources in schools (specific teacher characteristics, administrative inputs, and facilities).

The real resource category receives the bulk of analytical attention. First, these best summarize variations in resources at the classroom level. Teacher education and teacher experience are the primary determinants of teacher salaries. When combined with teachers per pupil, these variables describe the most significant variations in the instructional resources across classrooms. Second, these measures are readily available and well-measured. Third, they relate to the largest changes in schools over the past

three decades... with pupil–teacher ratios falling steadily, teacher experience increasing, and the percent of teachers with a masters’ degree actually doubling between 1960 and 1990.” Page 2073, Hanushek (2002)

When the outcomes of individuals in a group – in a workplace, neighborhood, town, school, or anywhere – are better the larger the size of the group, the individuals may be benefiting from external increasing returns to scale. The urban economics literature is crammed with research on the external increasing returns to scale phenomena called agglomeration economies of scale. Much less attention has been paid to the negative aspects of size or congestion effects. Activities in which bigger is not unambiguously better are primary and secondary education.

Class size is typically measured as the number of students a teacher faces during a period of instruction, which is the typical teacher's pupil load, or the pupil:teacher ratio. Average class sizes range across the US states from a low of 19 in Wyoming or Vermont to a high of 30 in Utah. Small classes may allow students to receive more personal attention from their teachers. Large classes, however, can be less expensive and do not necessarily hinder instruction. Depending on teaching style, student behavior, and other factors to be discussed, large classes may be just as effective as small classes.

Indeed, the empirical evidence about the effect of class size on student performance is at best mixed. Almost three quarters (72%) of the 376 statistical analyses cannot reject the null hypothesis that student outcomes are unrelated to the teacher-pupil ratio in schools (Hanushek, 2002). Hanushek’s summary of these studies is reproduced in Table 1. That table documents the preponderance of null findings as well as the disparate findings about optimal class size and other public education expenditure issues. Note that there is no consensus that smaller class sizes support better student performance. The strongest evidence against the null is for teacher test scores and teacher experience. Student performance appears to depend positively on teacher performance and experience (but not, apparently, on the education of the teacher).

As Hanushek points out, even strong evidence against the null must be interpreted carefully. It is not clear what the true direction of causality is. School spending and input use is

determined through complicated political and behavioral choices by schools and parents. In many states, spending is systematically allocated in a compensatory manner. Extra resources are spent on low-achieving kids to provide for more personalized attention, remedial instruction, and additional technology. It is not clear that either achievement or spending can be assumed exogenous. The identification problem results in an observed relationship that is often *negative*: high spending is often systematically correlated with low achievement. By the same token, class sizes could be small precisely in schools where policy-makers have noticed that kids need more help. Hanushek suggests that one way to overcome these limitations is to conduct controlled experiments.

Percentage distribution of estimated effect of key resources on student performance, based on 376 studies<sup>a</sup>

Resources	Number of estimates	Statistically significant		Statistically insignificant
		Positive	Negative	
<b>Real classroom resources</b>				
Teacher–pupil ratio	276	14%	14%	72%
Teacher education	170	9	5	86
Teacher experience	206	29	5	66
<b>Financial aggregates</b>				
Teacher salary	118	20	7	73%
Expenditure per pupil	163	27	7	66
<b>Other</b>				
Facilities	91	9	5	86
Administration	75	12	5	83
Teacher test scores	41	37	10	53

**Table 1.** Source: Table 6, page 2076; Hanushek, 2002.

Indeed, highly influential research concluding that students appear to do better in smaller classes came from a large experimental study by the State of Tennessee. “Project STAR” was a class size reduction experiment conducted in elementary schools in the 1980s. Classes were comprised of varying fractions of students who had been exposed to a “Small class treatment,” creating groups of students with varying experimentally-induced quality. Their conclusion that students in smaller classes do better has motivated major education policy changes in other states.

On the basis of the STAR results, in 1996 the State of California enacted a policy to provide significant additional funds to all schools to lower class sizes in primary grades to state-prescribed levels. The demand for additional teachers to staff all the new class rooms outpaced supply. The outcome was a dramatic reduction in student performance, most likely due to the assignment of teachers who lacked qualifications or experience to the new class rooms (Stecher and Bohrnstedt, 1999). One implication is that it is clearly not easy to raise student performance by raising teacher-pupil ratios in the short run because it is too difficult to do so without sacrificing teacher quality. And as per Table 1, teacher quality is likely to be more important than the pupil:teacher ratio.

An interesting approach to identifying the effect of class size on student performance was applied by Angrist and Lavy (1999). They used a Talmudic rule that has been used to determine the division of enrollment cohorts into classes in Israeli public schools to identify the effects of class size on the scholastic achievement of Israeli grammar school pupils. The rule caps the pupil:teacher ratio at 40. The rule is an exogenous source of variation that can be used to construct instrumental variables estimates of class-size effects using a regression-discontinuity design approach. The data display an overall positive correlation between scores and the number of students per grade, which turns out to be attributable to the fact that larger schools in Israel are more likely to be located in relatively prosperous big cities. After controlling for this positive association between test scores urbanization, they find a significantly negative association between class size (as instrumented) and student performance.

It is also possible, however, that students can do better in larger classes or among larger cohorts because a larger class contains a wider array of students. As Jesse Levin wrote:

“In this paper the controversial education topic of class size reduction is addressed. Controlling for a large number of observable characteristics and potential endogeneity in the class size variable, an educational production function is estimated using a quantile regression technique. The "conventional wisdom" that class size reduction is a viable means to increase scholastic achievement is discounted. Rather, the results point towards a far stronger *peer effect* through which class size reduction may play an important role. Due to heterogeneity in the newly identified peer effect, class size reduction is shown to be a potentially regressive policy measure.” (abstract, Levin, 2001)

In order for students to be able to learn from each other there must be some dispersion among the abilities in the group. This leads to our interest in what are called ‘peer effects.’

### ***Peer Effects***

If the performance of a group -- in a workplace, a town, a school, anywhere -- depends on the composition of group, the individuals in the group may be experiencing a type of club good externality known as a ‘peer effect.’ Peer effects have been studied in many contexts, most notably in schools, workplaces, among teenagers, and in neighborhoods (e.g., Plotnick and Hoffman, 1999; Evans, Oats and Schwab, 1992). Research on ‘peer effects’ in schools appears to have been first published by urban/regional economists Richard Arnott and John Rowse. Their 1987 paper focused on the influence of a student's classmates on his educational attainment. Their findings that the better a student’s classmates do, the better a student does put in question the efficacy of tracking students by ability, and suggested that mixing students of different abilities might be preferable.

More recent research by Hanushek et al (2003) points out that the analyses of peer effects on student achievement can be plagued by the difficulties of separating peer effects from other confounding influences. Hanushek controls for the most important determinants of achievement such as observable family and school characteristics, which have been shown to confound estimates of peer effects. He still finds that peer achievement has a positive effect on achievement growth and that students throughout the school test score distribution appear to benefit from the presence of higher achieving schoolmates.

It is also important to note that peer effects appear to be asymmetric<sup>2</sup>: positive effects result from increases in the numbers of higher performing peers, but there are no discernable negative effects from increases in the numbers of lower-performing peers in a group. For example, Angrist and Lang (2002) studied Boston Public School's **Metco** integration program that sends mostly poor black students out of the Boston public school district to attend schools in more affluent suburban districts. The transfer doubled the proportion of students with substantially lower test scores in some schools. While clearly this resulted in an overall decline in scores in the receiving schools, the authors found that the decline was not a peer effect, but a composition effect. They found no impact on the average scores in the sample of all non-transfer students. On the other hand, they did find that some of the transfer students mildly benefited. In sum, the existing evidence suggests that lower-performers enjoy a positive benefit of mixing with better students, but there is no discernable negative impact on better students of adding lower performers (see also Zimmer and Toma, 2000; and Williams and Zimmerman, 2003).

### ***Horizon Effects***

We now formalize a hypothesis that students in schools with more potential role models perform better:

#### *Horizon Effect Hypothesis:*

*If the performance of a group depends on the frequency of observations of maximum performances in the group, because the frequency of maximum observations is higher the larger is a sample, then larger size groups will ultimately display higher overall performance.*

The first part of the horizon hypothesis assumes there are asymmetric peer group effects. The second part of the hypothesis is a statistical fact. The larger the sample drawn from a given distribution, the more frequently the maximum is observed in that sample (Poirier, 1995). Our hypothesis is based on the fact that the more students there are per grade, the more often one can observe a high performance (indeed, the more often one observes any level performance). If

---

<sup>2</sup> or 'heterogeneous' (Levin, 2001)

because top performances are more frequent (and low performances are irrelevant), everyone in the school, teachers as well as students, have higher expectations, are inspired to exert more effort or challenge themselves more, then larger cohorts will ultimately display higher averages.

Formally, we assume that individual performance  $y_{is}$  is a positive function of individual effort  $E_i$  of the  $i^{\text{th}}$  student at school  $s$ , and of the level of the material ( $M_s$ ) presented by teachers, among other things:

$$y_{is} = e_{is}(E_{is}, M_s, \dots) \quad (1)$$

Let each performance be drawn independently from a distribution or probability density function (pdf)  $f(y)$ , with a cumulative density function (cdf)  $F(y)$ . Let  $t = \max\{y_{is}\}_i^C$  be the maximum observation of  $y$  (e.g.,  $t$  for “top score” like 100%) in a sample of size  $C$ .

The pdf or distribution of  $t$ ,  $g(t)$ , is:

$$g(t) = C[F(t)^{C-1}]f(t) \quad (2)$$

page 144, Poirier (1995); where  $F(t) = F(y)$  evaluated at  $t$ , etc.

Clearly, the frequency of the maximum observation  $t$  is positively related to the size of the sample,  $C$ , i.e.,  $g(t)$  is a function of  $C$ . (It is also true that the minimum observations are also more frequent the larger is the sample size.) A higher frequency of maximum performances --and minimum performances—in itself does not lead to higher average scores in larger groups. We are assuming a single set of moments: the means and variances do not differ across groups of different sizes per se. But the means and variances may evolve over time. There may be an asymmetric peer-group effect or other mode of feedback from the frequency of observations of the maximum performance, such as the treatment of the group by teachers, or the demonstration effect on effort. The feedback may generate a shift in learning outcomes over time.

In particular, our horizon effect hypothesis is that both individual effort and treatment by teachers depend positively on the frequency that the maximum performance is observed, which is directly proportional to the size of the cohort in the school,  $C_s$ :

$$e_{is}(E_{is}(g(t)), M_s(g(t)), \dots) = z_{is}(C_s) \quad (3)$$

Given the statistical fact about how the frequency of a maximum observation increases with the size of the sample, we can instrument or proxy an unobserved frequency of the maximum performance within a group by the observed size of the group.

Another way in which students and teachers form expectations about what kind of performance is possible is by observing a wider range of students across grades or over time. In the same way that a larger cross-sectional sample includes more observations of a maximum than a smaller cross-section, a larger time-series or panel also contains more than a simple cross-section. Thus, if by school tradition or policy the operative cohort is longitudinal, including students of all grades and ages in the school, then we expect to find a positive influence of the number of grades in the school on average performance. We measure the number of grades in a school ( $G$ ) by subtracting the lowest grade in the school from the highest grade (plus 1), and call this a longitudinal measure of cohort size, which should be positively correlated with the “horizon.”

The cross-section measure of horizon ( $X$ ) is the number of students in a whole school building ( $S$ ) divided by the number of grades ( $G$ ). Clearly,  $G \cdot X = S$ ; which means that a school size may be highly correlated with horizon (see Table 2 and Appendix Table I).

There may also be internal increasing returns or size-related externalities to school size. The internal increasing returns are due, for example, to the fixed costs of school buildings. Benefits internal to the school but external to the individual student arising from school size may be positive, such as the ‘spillovers’ or returns to variety arising from interactions with many other students or teachers. Conversely, there may be scale diseconomies or congestion effects. In particular, because people tend to do good things unilaterally (to internalize the rewards) but do bad things in groups (to defuse responsibility or disperse the punishment), it is possible that student disciplinary problems may depend on reaching a ‘critical mass’ of miscreants. There is also a larger frequency of minimum performances in larger samples. The likelihood of reaching a

critical mass of miscreants clearly increases with the overall size of the school. In this case, average proficiency rates may depend negatively on the size of a school.

We now summarize a school performance function that explicitly includes cross-section and longitudinal measures of size: the number of students per grade ( $X=S/G$ ) and the number of grades ( $G$ ), as well as class size ( $R$ ) measured by the pupil:teacher ratio, and school size instrumented by the amount spent on the building ( $B$ )<sup>3</sup>:

$$y = y(X, G, R, B, \mathbf{F}, \mathbf{T}, \mathbf{D}; \boldsymbol{\beta}) + \varepsilon \quad (4)$$

controlling for:

$\mathbf{F}$  = measures of financial resources devoted to instruction, staff support, etc;

$\mathbf{T}$  = indicators of teacher quality (experience, etc);

$\mathbf{D}$  = student/family/neighborhood characteristics (income, education, density; etc).

### ***Empirical Approach***

To empirically test our hypotheses, we will first estimate a learner outcome function (4) summarized as:

$$\mathbf{Y}_s = \boldsymbol{\beta}\mathbf{X}_s + e_s$$

where  $\mathbf{Y}_s$  is the dependent variable vector, the percent of students in school  $s$  that are proficient in a subject; and  $\mathbf{X}_s$  is a vector of explanatory variables related to school  $s$ , including two ways to measure the view of the horizon: cross-sectionally and longitudinally. The “horizon” is measured both by the number of students in the same grade ( $X$ ) and by the number of grades ( $G$ ) in each building. We also control for class size, measured as the ratio of students to teachers; and school size instrumented by spending on the building. We estimated various functional forms to allow for non-linear relationships.

We focus on schools rather than individuals because we have no data on individual educational outcomes. Furthermore, we focus on school *buildings* rather than school districts for

---

<sup>3</sup> See Appendix Table 1 for correlations;  $B = \$21,036 + 1,352.6 \cdot S$ ; with an  $R^2$  of 72%.

two reasons. First, because of the policy relevance. The federal “No Child Left Behind” Act of 2002 gives students’ parents the choice of school *buildings* within school districts. Thus, a very policy-relevant unit of analysis is the school building. Another good reason to focus on school buildings is the aggregation problem noted by Plenzler (2004):

“Studies that attempt to establish the relationship between economic resources and student performance are often carried out using sample data aggregated to the school district level ... This is primarily due to data limitations. A problem with this approach is that large urban school districts exhibit tremendous variation within the district, so use of a district average student performance score and district average spending, student and teacher characteristics as well as socioeconomic population characteristics obscures this within district variation. A further problem is that smaller *rural school districts often contain only one or two elementary school buildings, so the cross-section of observations aggregated to the district level represent very different spatial scales.*” Page 1, Plenzler, 2004 (*emphasis added*)

The educational outcome and school variables for school year 2001-2002 were generously shared with us by our colleague Jim LeSage and his advisee Nicole Plenzer (2004). They obtained the data from the Ohio Department of Education. We downloaded all other variables characterizing households and neighborhoods at the zip code level from the US Census Bureau (year 2000).

Given our focus on school buildings, the first set of control variables relate to financial resources per school. We include a real average teacher salary, normalized by local average housing price, per teaching FTE expenditure on staff support, total expenditure per pupil, and the percentage of total spending on instruction. The next control variable measures teacher quality: average years of teaching experience. The third set concerns student/family demographics in the zip code area where the school is located: the percentage of the local population in poverty, the percentage of the local population that is non white, and the percentage of households containing someone with a Master’s or PhD degree; per capita income, and population density; Table 2.

**Table 2: Summary Statistics of All Variables**

Variable	label	mean	max	min	stdev
% 10 <sup>th</sup> grade proficient (@ 9 <sup>th</sup> math)	Y	87	100	32	11
Size (Total enrollment)	S	785	2913	56	478
Cross section size (S/G)	X	191	728	14	124
Number of grades	G	4	6	3	0.7
Class size (S/teaching FTE)	R	17	35	8	3.1
Building Expenditures	B	\$1,083,866	\$5,441,190	\$7,392	\$762,927
Total Expenditures Per Pupil	TEXP_P	\$7,437	\$14,864	\$4,127	\$1,584
% of Total Spending on Instruction	ISHARE	58%	83%	40%	6.2%
Average Teacher Experience (yrs)	EXPERNC	14	27	4	2.9
Spending on Staff Support/FTE	SS_FTE	\$1,634	\$8,021	\$17	\$1,445
Real Salary (Salary/House Price)	SAL_H	45%	159%	16%	15%
% of local population in poverty	PERPOV	9%	47%	0.9%	7%
% of local population non-white	PERNOWHI	10%	99%	0.0%	17%
% with Masters or PhD degree	PERGRA	3%	18%	0.0%	2%
Local average per capita income	INC_CAP	\$20,199	\$54,344	\$9,739	\$5,625
Local population density	POPENSITY	1088	10967	17	1737

The percent of 10<sup>th</sup> grade students proficient in 9<sup>th</sup> grade Math is the dependent variable measuring performance. The explanatory variables are described above and in Table 2. The first step is to fit a non-spatial regression model using Ordinary Least Squares:

$$y_s = \beta_0 + \beta_1 x_{1s} + \beta_2 x_{2s} + \dots + \varepsilon_s \quad s = 1, \dots, 633 \quad (5)$$

where  $y_s$  is the percentage of 10<sup>th</sup> grade students proficient in 9<sup>th</sup> grade math at 633 schools for the year 2001-2002;  $X_s$  are the explanatory and control variables; the elements of the vector  $\beta$  are the unknown parameters to be estimated; and  $\varepsilon$  is the error; the difference between the observed and the estimated  $y_s$  given observed explanatory variables  $X$  and the estimated  $\beta$ s.

We also used confirmatory spatial data analysis to assess the dependence of high school math proficiency on our hypothesized explanatory variables in a spatial setting. The data on the percentages of students in the 10<sup>th</sup> grade that passed in the Math test for 9<sup>th</sup> graders across public schools in the State of Ohio displays global and local spatial autocorrelation as well as spatial heterogeneity.

As a consequence of these findings, the second step is to identify which spatial model is recommended: (i) the spatial lag model; (ii) the spatial cross-regressive model; or (iii) the spatial error model. We anticipate that the spatial lag model may be most appropriate for the same reasons that a temporal lag or the first-differencing of time series observations can control for time-invariant unobservables to minimize missing variable bias.

The spatial weights used to test for spatial autocorrelation given the model are ones for contiguous neighboring zip areas, and zeros for non-contiguous zip areas, using a binary queen contiguity matrix. Six spatial autocorrelation tests were computed using this spatial weight matrix: the Moran's *I* test; the Lagrange multiplier test for residual spatial autocorrelation, and its robust version; the Lagrange multiplier test for spatially lagged endogenous variables, and its robust version (Anselin, 1996). The results of these tests are in Table 3.

**Table 3: Diagnostic Tests for Spatial Dependence**

<b>Test</b>	<b>statistic</b>	<b>p-value</b>
Moran's I (error)	0.183	0.855
Lagrange Multiplier (lag)	3.331	0.068
Robust LM (lag)	9.923	0.002
Lagrange Multiplier (error)	0.006	0.938
Robust LM (error)	6.598	0.010
Lagrange Multiplier (SARMA)	9.929	0.007

Because the statistic for the Robust LM (lag) test is the most significantly different from zero (has the lowest p-value), we can conclude that for this dataset, the spatial lag model is the recommended approach to control for the apparently positive spatial dependence. If spatial autocorrelation is ignored, the results are similar to the consequences of omitting a significant explanatory variable from the regression model, i.e., the estimators are biased.

The spatial lag approach is to include a function of the dependent variable in contiguous locations, called a spatial lag, as an additional covariate. According to Anselin (1998), “a spatial lag is constructed as a weighted average (using the weights in the spatial weighted matrix) of the values observed for the neighbors of a given location” (p.260). The magnitude of that influence

is measured by the estimated coefficient  $\rho$ . The spatial lag model is estimated using Maximum Likelihood (ML) using GeoDa 0.9.5-i5 (Anselin *et al.*, 2004). The null hypothesis of no spatial autocorrelation is  $H_0: \rho = 0$  (see below).

Our spatial model of high school math proficiency rates by school is thus:

$$y = \rho W y + B_0 + B_1 x_1 + B_2 x_2 + \dots + B_{17} x_{17} + u \quad (6)$$

where  $y$  is the vector of the dependent variable;  $\rho$  is the spatial autoregressive parameter to be estimated, the vector  $B$  ( $B_0, B_1, B_2, \dots, B_{17}$ ) includes the unknown coefficients on our explanatory variables to be estimated; and  $W$  is the queen contiguity spatial weight matrix.  $W y$  contains the average percent of students proficient in nearest neighbor high schools. As before,  $X$  is the matrix of hypothesized explanatory and control variables, and  $u$  is the vector of error terms.

Table 4 presents the estimated non-spatial and spatial lag models. In Table 4, we can observe that the coefficient  $\rho$  on the spatial lag of the dependent variable is statistically significantly larger than zero (p-value = 0.05) indicating the presence of positive spatial dependence. And, our hypotheses about “horizon effects” are robust (the coefficients are stable between the non-spatial and the spatial lag models).

In particular, we reject the null hypothesis that performance is independent of the longitudinal measure of learning environment size or ‘horizon.’ The coefficient on  $\ln G$  is statistically significantly positive with 99 percent confidence. In the spatial lag model, we also reject the null that performance is independent of class size,  $R$ . The effect of class size on high school student math proficiency rates is strictly concave (Figure 2) and about 20 students per pupil appears to be optimal (higher than the mean class size of 17). Finally, plain old “bigger” is *not* better. The effect of overall school size, instrumented using the fixed cost of buildings, on high school math proficiency rates is significantly negative in the spatial lag model. That negative effect is not, however, strong enough to undermine the positive effect of longitudinal horizon (see Figure 3.)

The coefficients on the per FTE spending on staff support and the share of total spending on instruction are negative and significant in both models with 99 percent confidence, consistent with the literature. As noted above, other empirical tests of the efficiency of public expenditure on schools are unable to identify if outcomes are positively related to school spending. Note also that the coefficients on average teacher experience, real teacher salary, and total spending per pupil are insignificant.

It is also apparent that local demographics and neighborhood characteristics matter. The coefficients on the percent of the area's population in poverty and the percent of population that is non white are negative and significant in both models above the 99 percent level. The coefficients on the percent of the population with a Masters or PhD degree are positive and significant in both models with over 99 percent confidence. And proficiency rates are significantly inversely related to population density (in contrast with Israel's case, as per Angrist and Lavy, 1999). City schools do not appear to be better than rural schools.

The  $R^2$  criterion cannot be used to compare the "fit" of non-spatial OLS and spatial models, so the Akaike Information Criterion (AIC) and Schwartz Criterion (SC) are used. The best model has the lowest values of AIC and SC. In our case, the difference is too slight.

Table 4. High School Math Proficiency Model Estimates (n=633)	Non-spatial model		Spatial Lag model (Queen Contiguity Matrix)	
	OLS		Maximum Likelihood	
Variable	coefficient	SE	Coefficient	SE
Constant	84.52***	11.76	74.37***	12.59
$W\gamma$			0.10*	0.05
LnX (cohort size)	1.53	1.20	1.55	1.18
LnG (number of Grades)	8.49***	2.76	8.45***	2.72
R (Class size)	1.006	0.648	1.072*	0.638
R <sup>2</sup>	-0.026	0.016	-0.027*	0.016
B (\$10,000s)	-0.01	0.01	-0.01*	0.01
Total Exp/pupil	-0.000004	0.0004	0.00003	0.0003
% Instruction	-31.81***	6.55	-31.73***	6.44
Teacher Experience	0.10	0.12	0.10	0.12
SS_FTE	-0.0012***	0.0002	-0.0012***	0.0002
SAL_H	-5.32	3.84	-5.28	3.78
% in poverty	-24.15***	8.49	-21.41***	8.38
% non-white	-26.07***	2.90	-25.20***	2.90
% grad degree	103.91***	24.81	103.41***	24.40
income/cap	-0.0001	0.0001	-0.0001	0.0001
Pop. density	-0.0007***	0.0003	-0.0005*	0.0003
R <sup>2</sup> Percent explained	54%			
AIC Aikake Criterion	4414.9		4413.5	
SC Schwartz Criterion	4486.2		4489.2	

Note: \*\*\* 1% significance; \*\* 5% significance, \* 10% significance level

## Implications and Discussion

In this paper we have posed and tested hypotheses about “horizon” effects. The hypothesis is basically that if the performance of a group depends on the frequency of observation of top performances in the group, then larger groups will display higher performances. The empirical issues include measuring the ‘horizon’ and controlling for misspecification and missing variable or spatial bias.

All models estimated using data about Ohio high schools and their neighborhoods reject the null hypotheses that the percent of high school sophomores proficient at freshman math is independent of cross-section or longitudinal measures of horizon. In particular, the proficiency rates where there are more grades in one building are higher. Proficiency rates at larger schools,

however, are lower, suggesting that on net there are congestion effects rather than increasing returns to scale externalities or agglomeration economies, in high schools.

The net effect of increasing the number of grades (and school size, but holding all else equal) is illustrated in Figure 3. Our spatial lag model estimates suggest that increasing the number of grades from 3 to 6 can raise the expected proficiency rate from 84% to 89%. The increase is, however, within the range of standard error of proficiency rates. We can conclude that increasing the number of grades in a school may help, and will not hurt, sophomore performance rates in high school math.

We would like to discuss our findings in light of reduced-form empirical evidence that productivity is higher in higher density states (e.g., Ciccone and Hall, 1996). If we focus just on the positive dependence of math performance on cohort size, we have a finding similar to Ciccone and Hall's. They estimated how average labor productivity in a state depends on the distribution of employment across counties in a state. They assumed that output is a nonlinear function ( $\gamma$ ) of county workers ( $n_c$ ) per acre ( $a_c$ ) times the acres of land in the county, and that statewide output ( $Q_s$ ) is the sum of county output:

$$Q_s = \sum_c [(n_c/a_c)^\gamma \cdot a_c]$$

Statewide output per statewide workers or statewide productivity ( $N_s = \sum_c n_c$ ), is thus:

$$\frac{Q_s}{N_s} = \sum_c \frac{n_c^\gamma a_c^{1-\gamma}}{N_s}$$

Obviously, if  $\gamma < 1$ , productivity is higher when the workforce per acre is lower. If  $\gamma > 1$ , productivity is higher when the workforce is concentrated on a few acres. Their empirical work showed that  $\gamma > 1$ . The average performance of the workforce appears to be higher in states where there are more workers concentrated on fewer acres.

In our school data, we find no statistically significant effect of student density, measured by the cross-section number of students per grade ( $X$ ) on student math performance. Neither Ciccone and Hall nor most others who claim to have identified positive returns to scale or density

explicitly pose a mechanism by which proximity or size affects performance. They have simply asserted some parametric relationship by which productivity depends on population or population density. They call the unidentified processes “knowledge spillovers” or “urbanization external economies of agglomeration.”

We find Syverson’s (2004) explanation much more compelling. There is no mystery or agglomeration externality underlying the fact that more densely populated places can support more establishments. Fixed costs and transport costs obviously conspire to limit the size of the market that can be served by a supplier of a good or service. In a given area, the larger the number of customers in the area, the more suppliers there can be.

Like Syverson’s hypothesized mechanism, our horizon effect hypothesis does not rely on assumptions that students or teachers in bigger schools are intrinsically more able, nor do we assume that the ‘learning production function’ differs across schools of different sizes. In Syverson’s work, average productivity differences arise simply because low productivity (high cost) firms are driven out of larger markets where competition is stiffer. In our work, the difference in proficiency rates between longer and shorter cohorts may arise, as we hypothesize, because higher effort may be inspired in longer cohorts.

In sum, this research suggests that small towns may enhance their learning environments by enhancing intergenerational communication, enhancing global awareness, networking widely, or institutionalizing a ‘long memory.’ “Horizon effects” may also be operative among citizens, public officials, or businesses in towns. In the least, this research suggests that the quintessentially rural “one-room-schoolhouse” model for schools can be very effective.

## References

- Angrist, Joshua D. and Victor Lavy (1999) "Using Maimonides' Rule To Estimate The Effect of Class Size on Scholastic Achievement" Quarterly Journal of Economics, May:533-575.
- \_\_\_\_\_ and Lang, Kevin (2002) "How Important are Classroom Peer Effects? Evidence from Boston's Metco Program" NBER Working Papers: 9263.
- Anselin, Luc (1996) "The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association". In Fischer, M., Scholten, H. J., and Unwin, D. (eds.), Spatial Analytical Perspectives in GIS. Taylor & Francis: London: 111-125.
- (1998) "Interactive Techniques and Exploratory Spatial Data Analysis". In Longley P.A., Goodchild M.F., Maguire D.J., and Wind D.W, (eds.) Geographical Information Systems: Principles, Techniques, Management and Applications. Wiley: New York.
- ; Syabri, Ibnu and Kho, Youngihn (2004) "GeoDa: An Introduction to Spatial data Analysis," Geographical Analysis
- Arnott, Richard; Rowse, John (1987) "Peer Group Effects and Educational Attainment" Journal of Public Economics 32(3): 287-305.
- Breshnahan, Timothy and Peter Reiss (1991) "Entry and Competition in Concentrated Markets." Journal of Political Economy, 99(5):997-1009.
- Campbell, Jeffrey and Hugo Hopenhayn (2002) "Market Size Matters." NBER Working Paper No. 9113 <http://www.nber.org/papers/w9113>
- Castle, Emery (1991) "The Benefits of Space and the Cost of Distance," Kenneth Pigg, Ed., The Future of Rural America: Anticipating Policies for Constructive Change, Westview Press.
- Ciccone, Antonio, and Robert E. Hall (1996) "Productivity and the Density of Economic Activity" American Economic Review 86(1):54-70.
- Duranton, Gilles, and Diego Puga (2001) "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products" American Economic Review 91(5):1454-1477.
- Evans, William, Wallace Oates and Robert Schwab, (1992) "Measuring Peer Group Effects: A Study of Teenage Behavior," Journal of Political Economy, 966-991.
- Hanushek, Eric (2002) "Publicly Provided Education" Chapter 30 in Handbook of Public Economics Volume 4, Edited by A.J. Auerbach and M. Feldstein, Elsevier Science B.V.
- \_\_\_\_\_ Kain, John; Markman, Jacob; and Rivkin, Steven (2003) "Does Peer Ability Affect Student Achievement?" Journal of Applied Econometrics 18(5): 527- 44.
- Jacobs, Jane (1984) Cities and The Wealth of Nations: Principles of Economic Life New York: Random House.
- Jovanovic, Boyan (1982) "Selection and the Evolution of Industry," Econometrica 50(3): 649-70.

- Levin, Jesse (2001) "For Whom the Reductions Count: A Quantile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement," Empirical Economics 26(1): 221-46.
- Plenzler, Nicole (2004) "Student Performance and Educational Resources: A Spatial Econometric Examination" Master's Thesis, Department of Economics, *University of Toledo* (Prof. James LeSage, major advisor); May.
- Plotnick, Robert and Saul Hoffman, (1999) "The Effect of Neighborhood Characteristics on Young Adult Outcomes: Alternative Estimates," Social Science Quarterly, LXXX:1-18.
- Poirier, D.J. (1995) Intermediate Statistics and Econometrics: A Comparative Approach, Cambridge, MA: MIT Press.
- Raith, Michael. (2003) "Competition, Risk, and Managerial Incentives." American Economic Review 93(4):1425-36.
- Stecher, B.M., and G.W. Bohrnstedt (1999) "Class Size Reduction in California: Early Evaluation Findings, 1996-98" *American Institutes for Research*, Palo Alto, CA.
- Syverson, Chad (2004) "Market Structure and Productivity: A Concrete Example" Journal of Political Economy <http://home.uchicago.edu/~syverson/concreteexample3.pdf> (date accessed March, 2004).
- U.S. Bureau of the Census "Census 2000 Summary File 3 (SF 3) - Sample Data" <http://factfinder.census.gov>
- U.S. Dept of Education "Fiscal Year 2001-2005 State Tables for the U.S. Department of Education" <http://www.ed.gov/about/overview/budget/statetables/index.html>
- Vernon, Raymond (1966) International Investment and International Trade in the product Cycle," Quarterly Journal of Economics 80(2):190-207.
- Williams, Gordon and David Zimmerman (2003) "Peer Effects in Higher Education" *National Bureau of Economic Research*, Inc, NBER Working Papers: 9501
- Zimmer, Ron, and Eugenia Toma (2000) "Peer Effects in Private and Public Schools across Countries" Journal of Policy Analysis and Management Winter 19(1):75-92.

Appendix Table I. Correlation Matrix

	Y	S	X	G	R	F	TEXP_P	ISHARE	EXPERNC	SS_FTE	SAL_H	PERPOV	PERNOWHI	PERGRA	INC_CAP	POPDENSITY
Y	1															
S	-0.14	1														
X	-0.14	0.99	1													
G	0.1	-0.25	-0.37	1												
R	0.21	0.04	0.05	-0.05	1											
F	-0.15	0.85	0.83	-0.22	-0.06	1										
TEXP_P	-0.25	0.07	0.07	-0.06	-0.57	0.35	1									
ISHARE	-0.23	0.23	0.22	0.05	-0.15	-0.06	-0.08	1								
EXPERNC	0.02	-0.05	-0.05	0.03	0.04	-0.03	0.09	0.09	1							
SS_FTE	-0.3	0.2	0.2	-0.09	-0.1	0.23	0.25	-0.15	-0.1	1						
SAL_H	-0.51	-0.02	-0.03	0.02	-0.15	0.01	0.15	0.14	0.23	0.16	1					
PERPOV	-0.53	-0.07	-0.08	0.02	-0.18	-0.02	0.18	0.14	0.09	0.17	0.74	1				
PERNOWHI	-0.64	0.22	0.21	-0.06	-0.25	0.27	0.39	0.17	-0	0.3	0.48	0.56	1			
PERGRA	0.16	0.38	0.37	-0.08	-0.1	0.41	0.33	0.05	-0	0.14	-0.4	-0.3	0.06	1		
INC_CAP	0.28	0.34	0.34	-0.11	-0.07	0.35	0.24	-0.01	-0.1	0.05	-0.6	-0.6	-0.1	0.82	1	
POPDENSITY	-0.49	0.44	0.43	-0.1	-0.28	0.49	0.45	0.22	-0	0.35	0.34	0.31	0.65	0.24	0.08	1

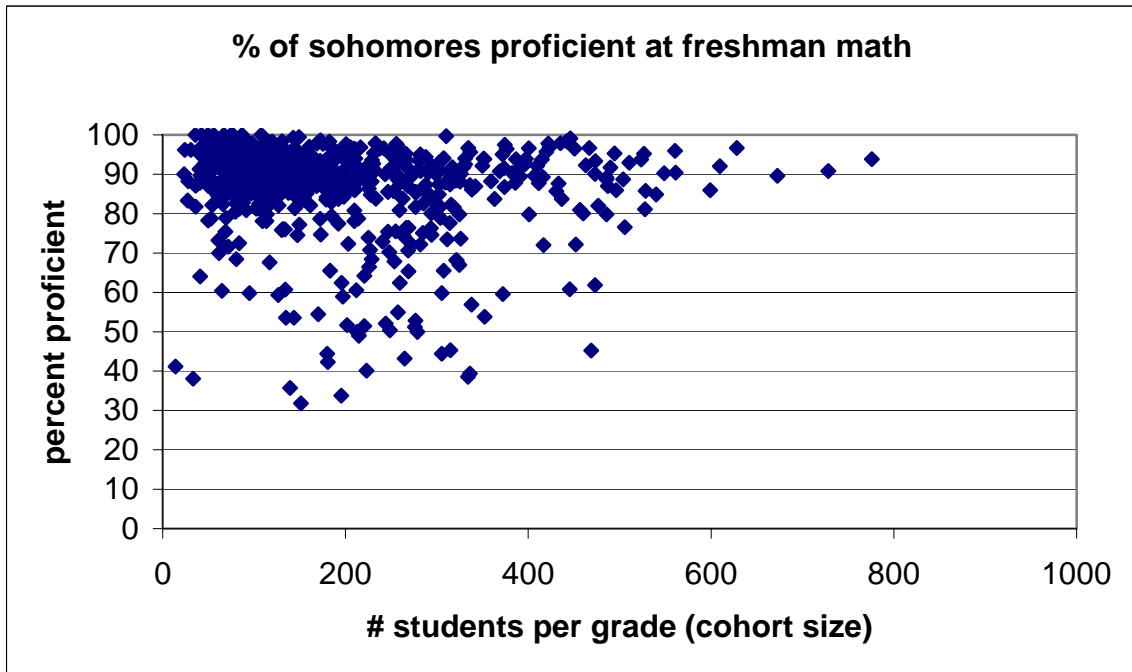
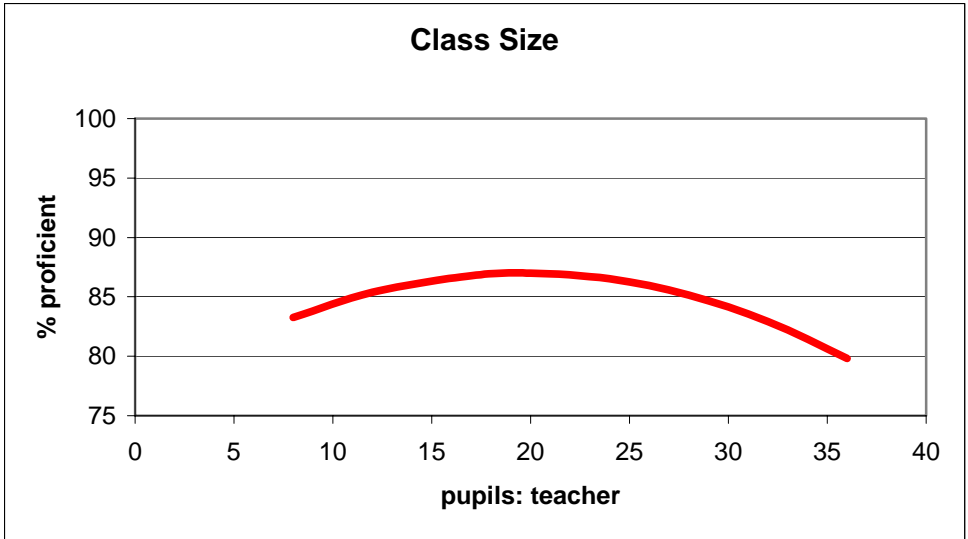
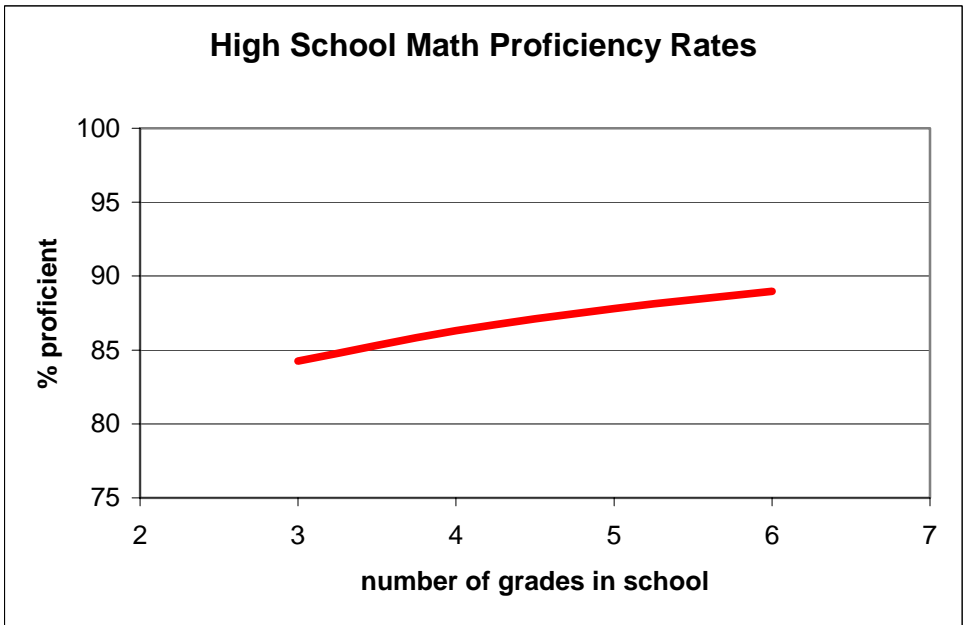


Figure 1: A scatterplot of the Ohio schools data



**Figure 2.** Simulation of the estimated effect of class size on proficiency rates



**Figure 3.** Simulation of the net effect of the number of grades in school on proficiency rates (where total enrollment (spending on building) also rises)