Senior Project
Department of Economics

How Much Do Parents Value
School Quality?
An Analysis of Cuyahoga
and Geauga Counties

Stephanie Harford
May 2015

Advisor: *Francesco Renna*

Abstract:

This paper studies how much parents value school quality by examining property values in Geauga and Cuyahoga counties, controlled for open or closed enrollment. Each county has a different open enrollment policy – Geauga County has open enrollment for all districts while Cuyahoga County has closed enrollment for almost all districts. A pooled hedonic model and controlled hedonic model were used to determine if the effect of school quality was higher with open or closed enrollment. The results supported the hypothesis that parents value living in an area with better schools more in a county with closed enrollment.

Table of Contents:

Introduction:

How much do parents value school quality? This is a question that has been studied by many economists over the years because it is important information for parents, school administrators, and policy makers. Proposed school reform policies can be better evaluated if there is a better understanding of the benefits to weigh against the costs. Because it is very difficult to study the direct effects of public school quality on lifetime outcomes, one approximation is to measure how much parents are willing to pay to live in an area with better schools through property values.

However, it's also important to look at factors that may influence how parents value school quality. One important factor may be whether an area has a policy of inter-district open enrollment, which is allowing students from one school district to attend school in another district. If a parent has the option to send their child to a school outside of the school district in which they live, they may value living in a district with better schools less.

Inter-district open enrollment has been somewhat controversial, particularly in the Geauga County area. In recent years, all Geauga County school districts have accepted open enrolled students from other districts, but the West Geauga County school district is currently voting whether or not they should continue this policy. On the other hand, Cuyahoga County, which is adjacent to Geauga County, does not allow inter-district open enrollment in most of their school districts. Because of the controversy, open enrollment is an issue that has been in the media spotlight, so most parents are likely familiar with whether or not the areas they live in have a similar policy. Therefore, I examine whether or not open enrollment policies change how much parents value school quality.

Literature Review:

Before examining the data, it is important to understand past studies on this topic. Prior to 1999, most economists used pooled hedonic regressions to estimate the effect of school quality on housing prices. Then Sandra Black, in her paper "Do better schools matter? Parental valuation of elementary education" studied the effect of elementary school quality on housing prices by specifically looking at houses that bordered attendance district boundaries in the suburbs of Boston, MA. This controlled for potential omitted variable bias from differences in neighborhoods so that the only differences should be which school was attended.

The results were consistent with previous studies in that school quality (measured through standardized test scores) and housing prices were positively correlated. However, the results differed in that the intensity of the effect was only about half the amount as previous studies reported, implying that those studies were heavily biased due to unobserved neighborhood characteristics. Specifically, with the boundary model, a 5% increase in test scores was associated with a 2.1% increase in housing prices, whereas the traditional method produced a result that associated a 5% increase in test scores with a 4.9% increase in housing prices.

Kane, Riegg, & Staiger (2006) studied the relationship between school quality and housing values in Mecklenburg County, NC during 1994-2001. This particular time period was significant for the county as there was a desegregation order and school boundaries were being redrawn. The authors used the same strategy as in Black (1999) and focused on houses near school boundaries and then also looked at the housing values that were affected by redistricting. The results suggest that changing the demographics of students assigned to a school will have an impact on housing prices dependent on how the district lines are restructured, but with a lag of several years. And the overall impact of schools on housing values is shown to be more indirect

than previously thought. When neighborhoods were reassigned to different schools, the population there responded, leading to residential sorting.

Seo & Simons (2009) focused on determining which school quality variables have the most influence on housing prices. Although most of the previous literature has come to the same conclusion that school quality does affect property values, there is much debate on what measures of school quality are most appropriate to use. The data used for this study came from Cuyahoga County, OH for 2000 and 2005, which includes the period when the No Child Left Behind policy was introduced in 2001. A series of hedonic models were run using different school quality measures, including input factors (teacher characteristics and expenditures per pupil); output factors (percentage of students at and above proficiency levels, performance index, and school district report card designations); value-added of year-to-year progress in output; and parent and peer characteristics. The results indicated that after the No Child Left Behind Act, test scores were replaced by the performance index and school district rating as the most important measures due to the fact that the information is more publicly available and comprehensive.

Chiodo, Hernandez-Murillo, & Owyang (2010) studied the effects of school quality on housing prices, but assumed that the effect is nonlinear. The paper built on the foundation of the Black (1999) paper by also only using data for houses bordering attendance district boundaries, but using a nonlinear fixed effects model. The authors argued that there are three reasons why a non-linear model is necessary. First, parents who value school quality a lot will focus their search for housing in the areas that have the highest school quality. This will increase demand while supply is mostly inelastic, increasing prices further. Second, a constant premium across all ranges of school quality is impractical given the availability of private schools or magnet schools. Third, school quality can be considered a luxury good, so will result in a higher

housing-price-to-quality ratio in richer neighborhoods. The data was obtained for St. Louis, MO during the 1998-2001 period. The results did reflect the expected outcome that there was a positive correlation for school quality and housing prices, but further showed that the linear model overestimated the effect in cases of lower quality schools and underestimated the effect in cases of higher quality schools.

Dhar & Ross (2012) developed a new approach for determining the effect of school quality on property values on the district level. The authors acknowledged that traditional models that use cross section data will overestimate the effect. They instead use district boundaries within a panel data set to control for "time invariant neighborhood quality fixed effects on either side of each boundary". The data was obtained for the state of Connecticut from 1994-2004 and included a large number variables on characteristics, market conditions, property taxes, and per-pupil spending. The results were also consistent with previous studies and showed that without proper controls, there is substantial bias from omitted variables. So this approach would be preferred if school district information that varies over time is available and attendance zone information is not appropriate or readily available.

In the conclusion of Black (1999), she indicates that it would be interesting to examine school choice policies when looking at school quality and housing prices. Because there were no other studies that focused on this, I am adding to the existing literature by focusing on it here.

Theoretical Model

When choosing where to live, current and prospective parents will generally take into consideration the quality of schools in the area. So in effect, by choosing where to live, they are "voting" for their preferred school district. This Economic theory comes from Tiebout (1956)

which describes how voters will sort themselves according to preferences for urban amenities by "voting with their feet". Assuming free mobility, people with a certain preference for amenities will group together with others with the same preferences and will collectively vote for policies accordingly. This shows that parents who value higher quality schools will likely try to live in an area with other parents who value higher quality schools, and also highlights why it is important to control for neighborhood effects.

As more parents choose to locate in areas with better school quality, the demand for housing in that area will increase, bidding up the price. This theory comes from Glaser (2008), which explains that the utility of all home owners/renters in a city must achieve spatial equilibrium. Spatial equilibrium is achieved when utility is constant across space. This shows that if two people own houses with similar characteristics but one house is in an area with higher amenity levels such as higher school quality or is closer to the city center, the amount they pay for that home must be proportionately higher to stay in equilibrium. Households must be compensated through lower housing prices in order to accept lower school quality.

There is also the assumption that households make decisions in order to maximize their utility levels. This theory is referenced in Harrison and Rubinfeld (1976), where they explain the connection between housing prices and the demand for clean air. Individual households maximize their utility function:

$$U(x, h)$$

Subject to the budget constraint:

$$y = x + p(h) + T$$

Where:

x = quantity of composite private goods, whose price is set equal to one

h = (h(h_1, ..., h_n) is a bundle of housing attributes

y = annual income

p(h) = housing price function

T = cost of transportation

Methodology

Previous economists have studied the effects of school quality on property values. I have added to this research by determining if open enrollment will change those effects by studying all school districts in Geauga County (representing open enrollment) and most school districts Cuyahoga County (representing no open enrollment).

The effects of school quality on property values were estimated using the following hedonic model for both Geauga and Cuyahoga counties separately:

$$saleamt_j = \beta_0 + \beta_1 score_i + \beta_2 yrbuilt_j + \beta_3 area_j + \beta_4 bedrooms_j + \beta_5 baths_j + \beta_6 halfbath + \beta_7 enroll_i$$
$$+ \beta_8 expend_i + \beta_9 income_i + \beta_{10} pop_i + \beta_{11} minoritypct_i + \epsilon$$

Score, enroll, expend, income, pop, and *minoritypct* are aggregate variables for each school district *i*. *Score* is a performance index, based on an average of standardized test scores; *enroll* is the total number of enrolled students; *expend* is the average expenditure per pupil, *income* is the median income, *pop* is the total population in thousands; and *minoritypct* is the percent of the population that is a minority.

Yrbuilt, area, bedrooms, baths, and *halfbath* are property characteristic variables for each parcel *j*. *Yrbuilt* is the year the home was built, *area* is the total square footage in hundreds of

feet, *bedrooms* is the number of bedrooms, *baths* and *halfbath* are the number of bathrooms and half bathrooms respectively.

Data was obtained for all real property sales that took place in 2013. During this timeframe there were three districts in Cuyahoga County that allowed for open enrollment: Cleveland Municipal, East Cleveland City, and Cuyahoga Valley JVS. Cleveland Municipal was the only district that was open to students in any other district, and was excluded from the sample due to its large size and extreme variance in demographics. East Cleveland City and Cuyahoga Valley JVS were both only open to students in adjacent districts. Therefore, these districts were excluded from the sample and Cleveland-University Heights City was also excluded because it is adjacent to the Cleveland and East Cleveland districts. All districts in Geauga County allowed for open enrollment, so all were included in the sample.

The school district aggregate data was obtained for the 2011-2012 school year from the Ohio Department of Education website. This would most likely have been the most recent data available for parents purchasing a home in 2013.

As a robustness check, the effects of school quality on property values were estimated again using the following controlled hedonic model for both Geauga and Cuyahoga counties separately. This controlled model included only districts on each side of the county boundary line (Kenston and West Geauga in Geauga County, and Orange, Mayfield, and Solon in Cuyahoga County). Because these districts are so close to one another, they share very similar characteristics so that theoretically, the only difference between them is the open enrollment policy. Therefore, in this model, the only school district aggregate variable that was included is score:

$$saleamt_j = \beta_0 + \beta_1 score_i + \beta_2 yrbuilt_j + \beta_3 area_j + \beta_4 bedrooms_j + \beta_5 baths_j + \beta_6 halfbath + \epsilon$$

My hypothesis is that the effect of school quality on property values will be positive in both counties but lower in Geauga County, where there is open enrollment, than in Cuyahoga County, where there is closed enrollment. The lack of school choice should cause the demand for housing in better school districts to increase since parents do not have the option to send their children elsewhere.

<u>Data & Variables:</u>

The map of Northeastern Ohio to the right shows the location of the two counties, adjacent to one another, with similar property tax rates along the boundary line.

Property tax rates

Cuyahoga County property tax rates continue to be the highest in the region. Here's a look at the rates that will be used for tax bills due in 2012.

Tax range per $100,000 of home value

▉ $3,000 or more		☐ $1,800 to $2,199	
▉ $2,600 to $2,999		☐ $1,400 to $1,799	
☐ $2,200 to $2,599		▉ Up to $1,399	
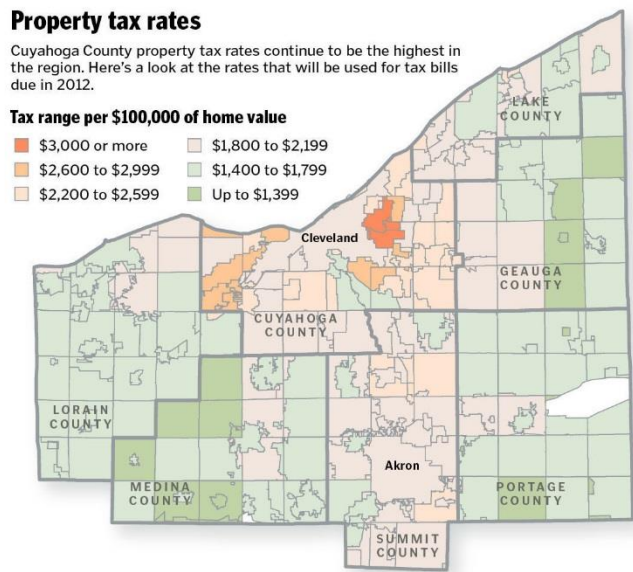
SOURCE: Ohio Department of Taxation   RICH EXNER AND KEN MARSHALL | THE PLAIN DEALER

Tables of the summary statistics and descriptions of the variables can be found in Appendix A. As shown in the scatter plots on Page 11, the general shape of the sale amount based on school district PI score is roughly the same in each county. The sale amount tends to increase overall as the PI score increases for both counties. The clearest difference is the number of PI scores reported. Since the school district variables are aggregated, there were relatively fewer PI scores reported.
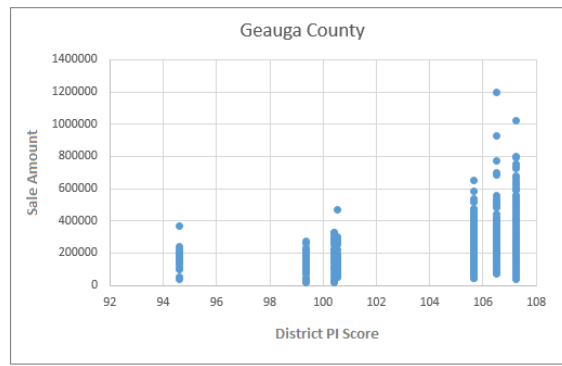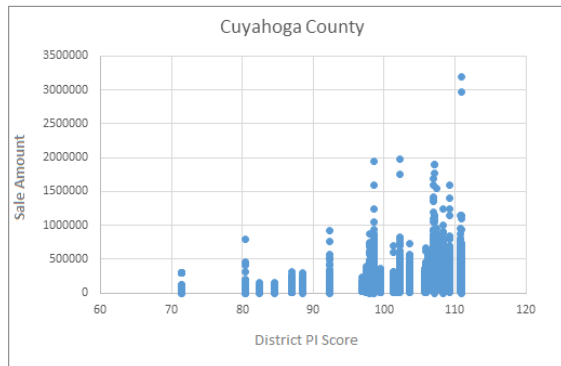
In addition, the first table of summary statistics in Appendix A shows a side-by-side comparison of the mean values for each variable by county. On average, the school districts in Geauga County tend to have families with higher incomes and larger, higher-value houses, and lower population and percentage minority. However, it appears that the two counties are overall similar.

The log of the property sale amount (*saleamt*) is the dependent variable for each model. The 2013 sales data for each county was matched by the parcel ID to the property characteristics taken from the 2010 census data on Neo Cando and the Geauga County Auditor's office. The independent variable of interest was a school district quality indicator. The Ohio Department of Education has available both the Performance Index (PI) Score and the Performance Index Ranking for each school district in the state. The Performance Index Score is based on student performance on the Ohio Achievement Assessments and Ohio Graduate Tests at the 3rd through 10th grade levels. The PI Score (*score*) was chosen over the PI Ranking (*rank*) because it more clearly indicated the performance differences between two districts. Two districts may be far apart in overall ranking, but only a few points away in actual score.

Pooled OLS Results:

The results for each county are shown the Pooled OLS table in Appendix A. All statistically significant variables have signs that are intuitive and consistent with previous research, although score, total district enrollment, expenditure per pupil, population, and percentage minority are not statistically significant for Geauga County. All variables are statistically significant at least at the 95% confidence level for Cuyahoga County.

In Cuyahoga County, which represents closed enrollment, a 1 point increase in the PI Score will increase the sale price by about 1.75%. This result is statistically significant at the 99% confidence level. In Geauga County, which represents an area with open enrollment, a 1 point increase in the PI Score will decrease the sale price by about 0.074%. However, this result is not statistically significant, possibly due to the lower sample size of only 757 observations. Therefore, the PI Score has zero effect on property values in Geauga County.

While the hypothesis did assume that school quality would have a positive effect on housing prices in both Geauga and Cuyahoga Counties, the hypothesis is supported in that the effect in Cuyahoga County is positive and statistically significant. This result is higher than the zero effect in Geauga County as predicted.

Controlled OLS Results:

The results for each county are shown the Controlled OLS table in Appendix A. Again, all statistically significant variables have signs that are intuitive and consistent with previous research, although score, year built, and number of half bathrooms are not statistically significant for Geauga County. All variables are statistically significant at least at the 95% confidence level

for Cuyahoga County, except for number of bedrooms. This could be due to the strong correlation between *bedrooms* and *area*.

This model shows that in Cuyahoga County, a 1 point increase in the PI Score will increase the sale price by about 1.912%. This result is statistically significant at the 99% confidence level. It also very similar to the result in the pooled model. In Geauga County, a 1 point increase in the PI Score will increase the sale price by about 0.645%. However, this result is again not statistically significant. Therefore, the PI Score has zero effect on property values in Geauga County in the controlled model as well.

These results also support my hypothesis and are also in agreement with the pooled results, showing the general robustness of the model.

Conclusion:

The results of this study imply that if parents have the choice to send their children to a different school district, they may not value living in a better school district as much. This policy implication is important because it gives insight to school boards and other policy makers when voting to decide whether or not have a policy of open or closed enrollment. If a district has a high PI score, there may be a negative impact to the property values in that area over time if they vote for a policy of open enrollment. This is important to weigh against the possible benefits of open enrollment policies.

One limitation to this study was the sample size in Geauga County. There were only about 1/10th the number of observations in Geauga County as in Cuyahoga County. This appears to be mostly due to the population difference between the counties. It's possible that I would

have gotten more statistically significant results in Geauga County if a larger sample size was used, perhaps by capturing sales data from multiple years.

It would have been ideal to use a boundary model more similar to the one in Black (1999), looking only at individual properties on each side of a district boundary line. However, this would have required GIS skills that, due to time constraints, I was not able to obtain. The hedonic model that was used is most likely affected to some degree by omitted variable bias, since there can be many different amenities in a given neighborhood. However, school district boundary lines are generally aligned with city or village boundary lines, so differences in amenities on the city level are controlled for.

For the future, it would be interesting if someone with the necessary GIS skills could replicate this study using a boundary model as outlined above. It would also be interesting to re-run this study with a difference-in-differences model in a few years if the West Geauga school district does vote to discontinue their current open enrollment policy. This would highlight the effect of changing an open enrollment policy in a district.

References:

Literature:

Black, Sandra (1999). "Do better schools matter? Parental valuation of elementary education". *Quarterly Journal of Economics, 114*(2), 577-599.

Chiodo, Abbigail J., Hernandez-Murillo, Ruben, & Owyang, Michael T. (2010). "Nonlinear effects of school quality on house prices". *Federal Reserve Bank of St. Louis Review*, 92(3), 185-204.

Dhar, Paramita, & Ross, Stephen L. (2012). "School district quality and property values: Examining differences along school district boundaries". *Journal of Urban Economics*, 71(1), 18-25.

Kane, Thomas J., Riegg, Stephanie K., & Staiger, Douglas O. (2006). "School quality, neighborhoods, and housing prices". *American Law and Economics Review*, 8(2), 183-212.

Seo, Youngme, & Simons, Robert A. (2009). "The effect of school quality on residential sales price". *Journal of Real Estate Research*, 31(3), 307-327.

Theoretical Model:

Glaeser, Edward L. (2008). "The economic approach to cities". Harvard University, John F. Kennedy School of Government. *Faculty Research Working Paper Series*.

Tiebout, Charles M. (1956). "A Pure Theory of Local Expenditures". *The Journal of Political Economy,* 64(5), 416-424.

Data Sources:
"Neo Cando". *NEO CANDO*. Case Western Reserve University, 2012.
http://neocando.case.edu/

"Ohio School Report Cards". *Ohio Department of Education*.
http://reportcard.education.ohio.gov/Pages/District-Search.aspx

"Geauga REALink". Geauga County Auditor's Office, 2010.
http://geaugarealink.co.geauga.oh.us/realink/

Appendix A: Tables

Data Descriptions		
Variable	**Description**	**Source**
saleamt	Property sale amount in 2013	Neo Cando & Geauga County Auditor
score	Performance Index Score	Ohio Department of Education
rank	Performance Index Ranking	Ohio Department of Education
area	Total usable square footage	Neo Cando & Geauga County Auditor
yrbuilt	Year house was built	Neo Cando & Geauga County Auditor
bedrooms	Number of bedrooms	Neo Cando & Geauga County Auditor
baths	Number of full bathrooms	Neo Cando & Geauga County Auditor
enroll	District enrollment in 2011	Ohio Department of Education
expend	Expenditure per pupil in 2011	Ohio Department of Education
income	Median income (from 2010 tax year)	Ohio Department of Education
pop	Population within district (from 2010 census)	Ohio Department of Education
minoritypct	Percentage of population that is minority in 2011-2012 school year	Ohio Department of Education

Means Compared		
Variable	**Cuyahoga**	**Geauga**
Saleamt	136194.86	240106.77
Score	97.831155	104.6640245
Rank	341.1246678	123.6930295
Area	1725.82	2280.36
Yrbuilt	1958.52	1974.47
Bedrooms	3.0718245	3.3967828
Baths	1.4621517	1.9718499
Halfbath	0.4891049	0.6420912
Enroll	5223.15	2435.7
Expend	12819.55	10750.83
Pop	43507.37	16220.37
Income	37265.99	42733.3
Minoritypct	32.8557133	4.9785121

Geauga County Summary Statistics					
Variable	N	Mean	Std. Deviation	Min	Max
Saleamt	746	240106.77	146083.21	15000	1200000
Score	746	104.6640245	3.2920269	94.5921	107.2266
Rank	746	123.6930295	123.6999005	35	544
Area	746	2280.36	879.9832531	386	6040
Yrbuilt	746	1974.47	31.402175	1818	2013
Bedrooms	746	3.3967828	0.7811129	1	7
Baths	746	1.9718499	0.8309523	0	5
Halfbath	746	0.6420912	0.5692788	0	3
Enroll	746	2435.7	887.155253	580	3144
Expend	746	10750.83	845.5691324	9258.59	12198.4
Pop	746	16220.37	4503.13	4208	21201
Income	746	42733.3	8136.38	28993	53394
Minoritypct	746	4.9785121	1.8194931	2.71	7.46

Cuyahoga County Summary Statistics					
Variable	N	Mean	Std. Deviation	Min	Max
Saleamt	13171	136194.86	143308.44	1	3200000
Score	13171	97.831155	9.2038701	71.4899	110.8793
Rank	13171	341.1246678	251.0542726	6	823
Area	13171	1725.82	839.3605626	242	11043
Yrbuilt	13171	1958.52	23.7504003	1818	2012
Bedrooms	13171	3.0718245	0.8678014	1	9
Baths	13171	1.4621517	0.6527718	1	8
Halfbath	13171	0.4891049	0.5578995	0	4
Enroll	13171	5223.15	2727.72	889	11251
Expend	13171	12819.55	2182.44	8971.08	21463.28
Pop	13171	43507.37	29719.2	4215	114098
Income	13171	37265.99	8723.94	25020	65754
Minoritypct	13171	32.8557133	31.0714443	1.06	99.42

Pooled OLS Results		
Variable	Cuyahoga	Geauga
Intercept	-2.52957**	-4.26934
	(-3.59)	(-0.96)
School Score	0.0175**	-0.00073683
	(7.05)	(-0.04)
Area (in hundreds of feet)	0.04322**	0.033985**
	(37.11)	(12.98)
Year Built	0.00359**	0.00273**
	(13.54)	(5.46)
Bedrooms	0.10033**	0.0386**
	(12.44)	(1.82)
Bathrooms	0.03026**	0.06557**
	(2.34)	(2.69)
Half Bathrooms	0.15767**	0.10733**
	(13.73)	(3.76)
Total District Enrollment	-0.00011055**	-0.00018396
	(-12.43)	(-1.03)
Expenditure Per Pupil	0.00002495**	0.00004148
	(7.77)	(0.82)
District Population (in thousands)	0.00967**	0.03362
	(10.93)	(1.22)
Log Income	0.38852**	0.90716**
	(6.33)	(1.96)
Percent Minority	-0.00677**	-0.02569
	(-11.45)	(-1.50)
Adjusted R-Squared	0.6243	0.5797
Number of Observations	13171	746
**Significant at least at the 95% confidence level (t-values in parentheses) *Significant at the 90% confidence level		

Controlled OLS Results		
Variable	**Cuyahoga**	**Geauga**
Intercept	5.97053**	8.12983
	(4.77)	(1.44)
School District Score	0.01912**	0.00645
	(5.57)	(0.12)
Area (in hundreds of feet)	0.033776**	0.035131**
	(16.42)	(10.49)
Year Built	0.00149**	0.00115
	(2.31)	(1.37)
Bedrooms	0.03168	0.05793**
	(1.61)	(1.99)
Bathrooms	0.15863**	0.09626**
	(6.40)	(2.93)
Half Bathrooms	0.09827**	0.05550
	(3.86)	(1.44)
Adjusted R-Squared	0.678	0.6326
Number of Observations	1200	369
**Significant at least at the 95% confidence level		

Merge code for individual Cuyahoga County city data sets:

```
%macro MultImp(dir=,out=);

%let rc=%str(%'dir %")&dir.%str(\%" /A-D/B/ON%');
filename myfiles pipe %unquote(&rc);

data list;
length fname $256.;
infile myfiles truncover;
input myfiles $100.;

fname=quote(upcase(cats("&dir",'\',myfiles)));
out="&out";
drop myfiles;
call execute('
  proc import dbms=xlsx out= _test
           datafile= '||fname||' replace ;
  run;
  proc append data=_test base='||out||' force; run;
  proc delete data=_test; run;
');
run;
filename myfiles clear;

%mend;

%MultImp(dir=E:\SAS\Data\cuyahoga\Excel,out=merged);
```

Regression codes for Cuyahoga County:

```
Libname Econ1 "E:\Econ1";

Proc Import Datafile= "E:\Final Version\cuyahoga_merged_data.csv"
      DBMS=CSV
      OUT=Econ1.cuyahoga
      Replace;
      run;

Proc Import Datafile= "E:\Final Version\ranking.csv"
      DBMS=CSV
      OUT=Econ1.ranking
      Replace;
      run;

Proc Import Datafile= "E:\Final Version\demographics.csv"
      DBMS=CSV
      OUT=Econ1.demographics
      Replace;
      run;

Proc Import Datafile= "E:\Final Version\cuyahoga_sale.csv"
      DBMS=CSV
      OUT=Econ1.sale
      Replace;
      run;

Data One;
```

```
        Set Econ1.cuyahoga;
        IF PLACENME="Bay Village city" THEN SDNAME="Bay Village City";
        IF PLACENME="Beachwood city" THEN SDNAME="Beachwood City";
        IF PLACENME="Bedford city" THEN SDNAME="Bedford City";
        IF PLACENME="Bedford Heights" THEN SDNAME="Bedford City";
        IF PLACENME="Walton Hills vil" THEN SDNAME="Bedford City";
        IF PLACENME="Berea city" OR "Brook Park city" OR "Middleburg Heigh" THEN
SDNAME="Berea City";
        IF PLACENME="Brook Park city" THEN SDNAME="Berea City";
        IF PLACENME="Middleburg Heigh" THEN SDNAME="Berea City";
        IF PLACENME="Brecksville city" THEN SDNAME="Brecksville-Broadview Heights
City";
        IF PLACENME="Broadview Height" THEN SDNAME="Brecksville-Broadview Heights
City";
        IF PLACENME="Brooklyn city" THEN SDNAME="Brooklyn City";
        IF PLACENME="Bentleyville vil" THEN SDNAME="Chagrin Falls Exempted Village";
        IF PLACENME="Chagrin Falls vi" THEN SDNAME="Chagrin Falls Exempted Village";
        IF PLACENME="Brooklyn Heights" THEN SDNAME="Cuyahoga Heights Local";
        IF PLACENME="Cuyahoga Heights" THEN SDNAME="Cuyahoga Heights Local";
        IF PLACENME="Valley View vill" THEN SDNAME="Cuyahoga Heights Local";
        IF PLACENME="Euclid city" THEN SDNAME="Euclid City";
        IF PLACENME="Fairview Park ci" THEN SDNAME="Fairview Park City";
        IF PLACENME="Garfield Heights" THEN SDNAME="Garfield Heights City Schools";
        IF PLACENME="Independence cit" THEN SDNAME="Independence Local";
        IF PLACENME="Lakewood city" THEN SDNAME="Lakewood City";
        IF PLACENME="Maple Heights ci" THEN SDNAME="Maple Heights City";
        IF PLACENME="Gates Mills vill" THEN SDNAME="Mayfield City";
        IF PLACENME="Highland Heights" THEN SDNAME="Mayfield City";
        IF PLACENME="Mayfield Heights" THEN SDNAME="Mayfield City";
        IF PLACENME="Mayfield village" THEN SDNAME="Mayfield City";
        IF PLACENME="North Olmsted ci" THEN SDNAME="North Olmsted City";
        IF PLACENME="North Royalton c" THEN SDNAME="North Royalton City";
        IF PLACENME="Olmsted Falls ci" THEN SDNAME="Olmsted Falls City";
        IF PLACENME="Hunting Valley v" OR "Moreland Hills v" OR "Orange village" OR
"Pepper Pike city" THEN SDNAME="Orange City";
        IF PLACENME="Moreland Hills v" THEN SDNAME="Orange City";
        IF PLACENME="Orange village" THEN SDNAME="Orange City";
        IF PLACENME="Pepper Pike city" THEN SDNAME="Orange City";
        IF PLACENME="Parma city" OR "Parma Heights ci" OR "Seven Hills city" THEN
SDNAME="Parma City";
        IF PLACENME="Parma Heights ci" THEN SDNAME="Parma City";
        IF PLACENME="Seven Hills city" THEN SDNAME="Parma City";
        IF PLACENME="Richmond Heights" THEN SDNAME="Richmond Heights Local";
        IF PLACENME="Rocky River city" THEN SDNAME="Rocky River City";
        IF PLACENME="Shaker Heights c" THEN SDNAME="Shaker Heights City";
        IF PLACENME="Glenwillow villa" THEN SDNAME="Solon City";
        IF PLACENME="Solon city" THEN SDNAME="Solon City";
        IF PLACENME="Lyndhurst city" THEN SDNAME="South Euclid-Lyndhurst City";
        IF PLACENME="South Euclid cit" THEN SDNAME="South Euclid-Lyndhurst City";
        IF PLACENME="Strongsville cit" THEN SDNAME="Strongsville City";
        IF PLACENME="Highland Hills v" THEN SDNAME="Warrensville Heights City";
        IF PLACENME="North Randall vi" THEN SDNAME="Warrensville Heights City";
        IF PLACENME="Warrensville Hei" THEN SDNAME="Warrensville Heights City";
        IF PLACENME="Westlake city" THEN SDNAME="Westlake City";
        DROP PCLASS;
        DROP LATITUDE;
        DROP LONGITUDE;
        DROP ADDRESS;
        DROP NUMBER;
        DROP STREET;
        DROP TSTAMP1;
        DROP FILTER;
        DROP FILTER1;
```

```sas
        IF AREA = 0 THEN DELETE;
        IF AREA = "." THEN DELETE;
        IF BEDROOMS = 0 THEN DELETE;
        IF BEDROOMS = "." THEN DELETE;
        IF YRBUILT < 1800 THEN DELETE;
        IF MKTVAL = 0 THEN DELETE;
        IF MKTVAL = "." THEN DELETE;
        run;

Data Two;
        Set Econ1.ranking;
        RENAME DISTRICT_NAME = SDNAME;
        RENAME _2011_Enrollment = ENROLL;
        RENAME _2011_PI_RANK = RANK;
        RENAME _2011_EXPENDITURES_PER_PUPIL = EXPEND;
        RENAME _2011_PI_SCORE = SCORE;
        DROP DISTRICT_IRN;
        DROP DISTRICT_TYPE;
        DROP _2011_GRADE_SPAN;
        DROP REASON_NO_PI_SCORE;
        DROP _2011_LRC_RATING;
        DROP _2011_AYP_STATUS;
        DROP _2011_PCT_STANDARDS_MET;
        DROP _2011_NUMBER_OF_STANDARDS_THAT_A;
        DROP _2011_VALUE_ADDED_COMPOSITE;
        IF COUNTY ^= "Cuyahoga" THEN delete;
        run;

DATA Three;
        Set Econ1.demographics;
        RENAME District_Name = SDNAME;
        RENAME Median_Income = INCOME;
        RENAME Population_Within_District = POP;
        RENAME Percent_Minority = MINORITYPCT;
        DROP IRN;
        DROP Enrollment__ADM_;
        DROP Log_of_Enrollment__ADM_;
        DROP Student_Poverty__Pct__Econ__Disa;
        DROP Median_Income__Standardized_;
        DROP Pct__With_Professional_Occupatio;
        DROP Pct__Professional__Standardized_;
        DROP Pct__of_Pop__with_at_Least_a_Bac;
        DROP Pct__with_Bachelors_____Standard;
        DROP Socioeconomic_Status__SES__Compo;
        DROP Pop__Density;
        DROP Pop__Density_Capped;
        DROP Pop__Den__Cap____Standardized_;
        DROP Pct__of_Non_Agricultural_Propert;
        DROP Pct__of_Non_Ag__Property_Value__;
        DROP Population__Standardized_;
        DROP City_Dummy_Variable;
        DROP Location_Composite;
        DROP Log_of_Pct__Minority;
        DROP Non_Agriculturate___Non_Resident;
        DROP Log_of_Non_Ag____Non_Res__Tax_Ca;
        DROP _2007_Typology_Code;
        DROP _2013_Typology_Code;
        DROP _2007_Typology_Description;
        IF District_County ^= "Cuyahoga" THEN DELETE;
        run;

DATA Four;
        Set Econ1.sale;
```

```
        RENAME CONVAMT = SALEAMT;
        DROP CONVTYP;
        DROP ADDRESS;
        DROP NUMBER;
        DROP STREET;
        DROP CNTYNME;
        DROP TSTAMP1;
        DROP FILTER;
        DROP FILTER1;
        IF SALEAMT = "." OR 0 THEN DELETE;
        run;

PROC SORT DATA=work.one;
        BY SDNAME;
        RUN;

PROC SORT DATA=work.two;
        BY SDNAME;
        RUN;

PROC SORT DATA=work.three;
        BY SDNAME;
        RUN;

DATA cuyahoga1;
    MERGE work.one(IN=fromonex) work.two(IN=fromtwox);
    BY sdname;
RUN;

DATA cuyahoga2;
    MERGE work.cuyahoga1(IN=fromcuy1x) work.three(IN=fromthreex);
    BY sdname;
RUN;

PROC SORT DATA=work.cuyahoga2;
        BY PARCEL;
        run;

PROC SORT DATA=work.four;
        BY PARCEL;
        run;

DATA Econ1.cuyahoga_final_data;
        MERGE work.cuyahoga2(IN=fromcuy2x) work.four(IN=fromfourx);
        BY PARCEL;
        RUN;

DATA Five;
        SET Econ1.cuyahoga_final_data;
        IF SALEAMT = "." THEN DELETE;
        IF SALEAMT = 0 THEN DELETE;
        IF RANK = "." THEN DELETE;
        LSALEAMT = log(saleamt);
        LINCOME = log(income);
        run;

proc reg data= work.five;
        model LSALEAMT= SCORE AREA YRBUILT BEDROOMS BATHS HALFBATH ENROLL EXPEND POP
LINCOME MINORITYPCT;
        run;

DATA Six;
        SET work.five;
```

```
        IF SDNAME = "Bay Village City" THEN DELETE;
        IF SDNAME = "Beachwood City" THEN DELETE;
        IF SDNAME = "Bedford City" THEN DELETE;
        IF SDNAME = "Berea City" THEN DELETE;
        IF SDNAME = "Brecksville-Broa" THEN DELETE;
        IF SDNAME = "Brooklyn City" THEN DELETE;
        IF SDNAME = "Chagrin Falls Ex" THEN DELETE;
        IF SDNAME = "Cuyahoga Heights" THEN DELETE;
        IF SDNAME = "Euclid City" THEN DELETE;
        IF SDNAME = "Fairview Park Ci" THEN DELETE;
        IF SDNAME = "Garfield Heights" THEN DELETE;
        IF SDNAME = "Independence Loc" THEN DELETE;
        IF SDNAME = "Lakewood City" THEN DELETE;
        IF SDNAME = "Maple Heights Ci" THEN DELETE;
        IF SDNAME = "North Olmsted Ci" THEN DELETE;
        IF SDNAME = "North Royalton C" THEN DELETE;
        IF SDNAME = "Olmsted Falls Ci" THEN DELETE;
        IF SDNAME = "Parma City" THEN DELETE;
        IF SDNAME = "Richmond Heights" THEN DELETE;
        IF SDNAME = "Rocky River City" THEN DELETE;
        IF SDNAME = "Shaker Heights C" THEN DELETE;
        IF SDNAME = "South Euclid-Lyn" THEN DELETE;
        IF SDNAME = "Strongsville Cit" THEN DELETE;
        IF SDNAME = "Warrensville Hei" THEN DELETE;
        IF SDNAME = "Westlake City" THEN DELETE;
        run;


proc reg data= work.six;
        model LSALEAMT= SCORE AREA YRBUILT BEDROOMS BATHS HALFBATH;
        run;

proc means data= work.five;
        var SALEAMT SCORE RANK AREA YRBUILT BEDROOMS BATHS HALFBATH ENROLL EXPEND POP
INCOME MINORITYPCT;
        RUN;

proc means data= work.six;
        var SALEAMT SCORE RANK AREA YRBUILT BEDROOMS BATHS HALFBATH ENROLL EXPEND POP
INCOME MINORITYPCT;
        RUN;
```

Regression codes for Geauga County:

```
Libname Econ1 "E:\Econ1";

Proc Import Datafile= "E:\Final Version\geauga_sale.csv"
        DBMS=CSV
        OUT=Econ1.geauga
        Replace;
        run;

Proc Import Datafile= "E:\Final Version\ranking.csv"
        DBMS=CSV
        OUT=Econ1.ranking
        Replace;
        run;

Proc Import Datafile= "E:\Final Version\demographics.csv"
        DBMS=CSV
        OUT=Econ1.demographics
```

```sas
        Replace;
        run;


Data One;
        Set Econ1.geauga;
        IF SchoolDistrictName="BERKSHIRE LSD" THEN SDNAME="Berkshire Local";
        IF SchoolDistrictName="CARDINAL LSD" THEN SDNAME="Cardinal Local";
        IF SchoolDistrictName="CHARDON LSD" THEN SDNAME="Chardon Local";
        IF SchoolDistrictName="KENSTON LSD" THEN SDNAME="Kenston Local";
        IF SchoolDistrictName="LEDGEMONT LSD" THEN SDNAME="Ledgemont Local";
        IF SchoolDistrictName="NEWBURY LSD" THEN SDNAME="Newbury Local";
        IF SchoolDistrictName="WEST GEAUGA LSD" THEN SDNAME="West Geauga Local";
        IF SchoolDistrictName = "CHAGRIN FALLS EVSD" THEN DELETE;
        IF SchoolDistrictName = "KIRTLAND LSD" THEN DELETE;
        IF SchoolDistrictName = "MENTOR EVSD" THEN DELETE;
        IF SchoolDistrictName = "RIVERSIDE LSD" THEN DELETE;
        RENAME ParcelId = PARCEL;
        RENAME SaleAmount = SALEAMT;
        RENAME BedroomCount = BEDROOMS;
        RENAME FullBathCount = BATHS;
        RENAME HalfBathCount = HALFBATH;
        RENAME YearBuilt = YRBUILT;
        RENAME FinLivingArea = AREA;
        RENAME MktTotal = MKTVAL;
        DROP PropClass;
        IF AREA = 0 THEN DELETE;
        IF BEDROOMS = 0 THEN DELETE;
        IF BATHS = 0 THEN DELETE;
        IF SALEAMT = 0 THEN DELETE;
        run;


Data Two;
        Set Econ1.ranking;
        RENAME DISTRICT_NAME = SDNAME;
        RENAME _2011_Enrollment = ENROLL;
        RENAME _2011_PI_RANK = RANK;
        RENAME _2011_EXPENDITURES_PER_PUPIL = EXPEND;
        RENAME _2011_PI_SCORE = SCORE;
        DROP DISTRICT_IRN;
        DROP DISTRICT_TYPE;
        DROP _2011_GRADE_SPAN;
        DROP REASON_NO_PI_SCORE;
        DROP _2011_LRC_RATING;
        DROP _2011_AYP_STATUS;
        DROP _2011_PCT_STANDARDS_MET;
        DROP _2011_NUMBER_OF_STANDARDS_THAT_A;
        DROP _2011_VALUE_ADDED_COMPOSITE;
        IF COUNTY ^= "Geauga" THEN delete;
        run;


DATA Three;
        Set Econ1.demographics;
        RENAME District_Name = SDNAME;
        RENAME Median_Income = INCOME;
        RENAME Population_Within_District = POP;
        RENAME Percent_Minority = MINORITYPCT;
        DROP IRN;
        DROP Enrollment__ADM_;
        DROP Log_of_Enrollment__ADM_;
        DROP Student_Poverty__Pct__Econ__Disa;
        DROP Median_Income__Standardized_;
        DROP Pct__With_Professional_Occupatio;
        DROP Pct__Professional__Standardized_;
```

```sas
        DROP Pct__of_Pop__with_at_Least_a_Bac;
        DROP Pct__with_Bachelors_____Standard;
        DROP Socioeconomic_Status__SES__Compo;
        DROP Pop__Density;
        DROP Pop__Density_Capped;
        DROP Pop__Den__Cap____Standardized_;
        DROP Pct__of_Non_Agricultural_Propert;
        DROP Pct__of_Non_Ag__Property_Value__;
        DROP Population__Standardized_;
        DROP City_Dummy_Variable;
        DROP Location_Composite;
        DROP Log_of_Pct__Minority;
        DROP Non_Agriculturate___Non_Resident;
        DROP Log_of_Non_Ag____Non_Res__Tax_Ca;
        DROP _2007_Typology_Code;
        DROP _2013_Typology_Code;
        DROP _2007_Typology_Description;
        IF District_County ^= "Geauga" THEN DELETE;
        run;

PROC SORT DATA=work.one;
        BY SDNAME;
        RUN;

PROC SORT DATA=work.two;
        BY SDNAME;
        RUN;

PROC SORT DATA=work.three;
        BY SDNAME;
        RUN;

DATA geauga1;
    MERGE work.one(IN=fromonex) work.two(IN=fromtwox);
    BY sdname;
RUN;

DATA geauga2;
    MERGE work.geauga1(IN=fromgeauga1x) work.three(IN=fromthreex);
    BY sdname;
RUN;

DATA Four;
        SET work.geauga2;
        IF SALEAMT = "." THEN DELETE;
        IF SALEAMT = 0 THEN DELETE;
        IF RANK = "." THEN DELETE;
        LSALEAMT = log(saleamt);
        LINCOME = log(income);
        run;

proc reg data= work.four;
        model LSALEAMT= SCORE AREA YRBUILT BEDROOMS BATHS HALFBATH ENROLL EXPEND POP
LINCOME MINORITYPCT;
        run;

DATA Five;
        SET work.four;
        IF SDNAME="Berkshire Local" THEN DELETE;
        IF SDNAME="Cardinal Local" THEN DELETE;
        IF SDNAME="Chardon Local" THEN DELETE;
        IF SDNAME="Ledgemont Local" THEN DELETE;
        IF SDNAME="Newbury Local" THEN DELETE;
```

```
        run;


proc reg data= work.five;
        model LSALEAMT= SCORE AREA YRBUILT BEDROOMS BATHS HALFBATH;
        run;

proc means data= work.four;
        var SALEAMT SCORE RANK AREA YRBUILT BEDROOMS BATHS HALFBATH ENROLL EXPEND POP
INCOME MINORITYPCT;
        RUN;

proc means data= work.five;
        var SALEAMT SCORE RANK AREA YRBUILT BEDROOMS BATHS HALFBATH ENROLL EXPEND POP
INCOME MINORITYPCT;
        RUN;
```