Applied Econometrics II Project

What is the Relationship Between State Spending and Standardized Test Scores in Idaho?

P. Christopher J. Daigle

30 April 2018

Table of Contents

1. Topic 3
2. Model / Methodology 3
3. Variables 4
4. Data 6
5. Summary Statistics 6
6. Estimates 7
7. Separate Approaches 9
8. Concluding Remarks 10
9. References 12
A. Appendix: Figures 15
B. Appendix: Pertinent Script 19
C. Appendix: Other Script 20

1. Topic

Through the utilization of econometric tools, economic theory, the state of Idaho's education system's standardized testing score data, and previous work by scholars in interdisciplinary fields, I examine the relationship between standardized academic scores and adjustments in school finance. The data utilized will be of cross-sectional nature, subset from data of panel nature, with observations from year 2007 to 2013, containing all scores reported by public school districts, not including charter schools, for the Idaho Standards Achievement Tests (ISAT). I have compiled, cleaned, and structured this data – it is original and of my construction.

Using this data, I will address my central question by asking: does funding have the same effect on the proportion of students in a school district passing the mathematics portion of the ISAT regardless of the year?

2. Model / Methodology

An Ordinary Least Squares (OLS) model is employed on cross sectional data. While a time component exists, the model analyzes individual points in time, not as a time series. The time component is in units of years from 2007 to 2013. This allows me to observe if the variables of interest have the same effect on the dependent variable in each time period. The population of interest is subset by school districts at a particular point in time.

My initial interest was to track school grades and not school districts, but because of the structure of the data and differing levels of aggregation for variables of interest, school districts are more appropriate. Spending, administrative markers (e.g. 4-day school week schedules), demographic markers (e.g. proportion of students in poverty, proportion of the population in poverty, race, gender, etc.) are aggregated at the district level and not grade level. Using district level data instead of grade level data allows me to observe the relationship of funding for the

whole district which is how the money is reported to be spent.

The base estimating model is:

$$(1)\ Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

$$(2)\ Y_{t+1} = \beta_{\alpha 0} + \beta_{\alpha 1} X_{1(t+1)} + \beta_{\alpha 2} X_{2(t+1)} + \beta_{\alpha 3} X_{3(t+1)} + u_{(t+1)}$$

Where (1) is compared to (2) over a succession of increasing periods and the hypothesis is tested:

$$(3)\ H_0 : \beta_1 = \beta_{\alpha 1}$$

$$H_1 : \beta_1 \neq \beta_{\alpha 1}$$

3. Variables

- $t = \{2007, 2008, 2009, \ldots, 2013\}$: indicates the year of observation. No more than a single time period "t" is analyzed in the linear regression. This particular sequence of years was chosen as the ISAT underwent a fundamental change in 2014 and was not administered. When administered in 2015, the ISAT no longer measured the same skills in the same ways as it did from 2007 to 2013.

- $Y_t = AvgAllP$: indicates the average proportion of students in all grades in a district who scored proficient on the mathematics portion of the ISAT. The average for the district as a whole is utilized instead of for the individual school grades as all the other variables available are district level and not grade level (as previously described). From the literature reviewed [Anderson & Walker (2012) and Biddle & Payne (1999) and Klein (1997) and Neymotin (2010) and Papke (2004) and Papke & Wooldridge (2008) and Turner (200)], mathematic scores are known to produce more economically and statistically significant results than reading scores. Therefore, data focused on math scores is utilized for this research.

- $X_{1t} = InfStatePPE$: indicates the if the amount of money spent per-pupil, in a particular school district, from state funds. Other funds available include federal and local. The amount is inflated to the value of a dollar spent in the Idaho State Department of Education's (SDE) fiscal year (FY) 2013 (InfFY13). Per-pupil-expenditure (PPE) is calculated by dividing InfFY13 by the average daily attendance of the district (Mem), as reported by the SDE. The FY inflator was generated from the Bureau of Labor Statistics' (BLS) reported measures of consumer price index (CPI) for the northwest region over the months July to June to coincide with Idaho's State fiscal year calendar. As this analysis is interested in how decreases in state funding and test scores, it is appropriate to observe the state component as opposed to the sum of each source of funding.
- $X_{2t} = FRLProp$: indicates the proportion of students on free-or-reduced lunch (FRL) in a school district. The variable is calculated by dividing the number of students on FRL, as reported by the non-profit organization Idaho Kids Count (IKC), by Mem. FRLProp addresses, potentially, two areas of interest: poverty and nutrition. In the literature, test scores are typically decreasing in the proportion of students in poverty (Payne & Biddle, 1999). Nutrition, especially in periods of development (e.g. as children are growing up), is generally shown to have a positive effect on testing (Card & Krueger, 1998).
- $X_{3t} = 4Day$: indicates if a school district employs a 4-day school-week schedule. This variable is reported from IKC and is an indicator (dummy) variable. A district operating on a 4-day school-week schedule holds classes 4 out of 5 weekdays. Schools have been shown to adopt a four-day school-week schedule during times of financial difficulty and produce varied (positive or negative) impacts on test scores (Klein, 2008) (Hanushek, 1986).
- u_t: illustrates the effect on the proportion of students passing the mathematics portion of the

ISAT for a specific grade in a specific time effected by factors not considered in this model.

4. Data

The data originates in a panel form with clusters such that there is a time component (years), a population component (school districts), made up of grades. It consists of 93 school districts, composed of 10 grades (3 – 12), over 7 time periods (2007 – 2013). I have conditionally subset the data based on school district and year criteria, such that the population analyzed is school districts in particular year. This allows me to create multiple cross sections of the districts in each time period. The data is compiled from sources such as the Idaho State Department of Education, Idaho Kids Count, the Bureau of Labor Statistics, and the US Census Bureau.

5. Summary Statistics

In the way of summary statistics related to the model, it is necessary to observe them by year and by district.

Beginning with the average proportion of students in a district who scored "proficient" on the mathematics portion of the ISAT, AvgAllP, the mean over each year ranges from about 29% to 38% with the lowest percent in 2012 of about 29.023% and the highest percent in 2007 of about 38.466%. See Figure 5.1 in Appendix A.

For Mem, the average daily attendance of students in a district, the mean of Mem ranges from about 1990 to 2110 with the lowest in 2012 of about 1992.38 and the highest in 2008 of about 2112.43. See Figure 5.2 in Appendix A.

For InfStatePPE, the real (FY13 Northwest Regional CPI) amount of state funds spent per-pupil, in a particular school district, from state funds, the mean of InfStatePPE ranges from about $6,340 to $7,940 with the lowest in 2012 of about $6,344 and the highest in 2007 of about $7,942.16. See Figure 5.3 in Appendix A.

For FRLProp, the proportion of students on free-or-reduced lunch (FRL) in a school district, the mean of FRLProp ranges from about 47.7% to 56.5% with the lowest in 2008 of about 47.735% and the highest in 2011 of about 56.421%.

Finally, for 4Day, an indicator variable which shows if a school district employs a 4-day school-week schedule, the mean ranges from about 0.11 to 0.33 with the lowest in 2007 of about 0.116 and the highest in 2013 of about 0.328. Most school districts don't use a 4-day school-week.

For comparisons of these variables overall, in the original form of panel data, see Figure 5.4 in Appendix A.

6. Estimates

Beginning with the base model, Figure 6.1 demonstrates all regressions of interest from 2007 to 2013 for school districts in Idaho – estimating the relationship between the district average of students passing the ISAT mathematics exam (AvgAllP) and real state per-pupil-expenditure (InfStatePPE), proportion of students on free-or-reduced lunch (FRLProp), and if a school district uses a 4-day school-week schedule (4Day).

Only the intercept coefficient (Constant, Beta0) and inflated state per-pupil-expenditure are consistently statistically significant in each time period. However, the values of the coefficients vary in each time period. So, the hypothesis of each coefficient having the same relationship over time is rejected.

For InfStatePPE, the coefficients have a negative relationship with the average proportion of students scoring proficient on the ISAT mathematics test in each year. The most economically meaningful value of InfStatePPE occurs in 2007 such that an increase in one dollar leads to a decrease of about -0.00239 percentage points of average students scoring proficient on the ISAT

mathematics test in a district. Put differently, an increase of approximately $420 of InfStatePPE is related to a one percentage point decrease in AvgAllP.

This seems to make sense as there are two ways to increase per-pupil-expenditure: 1) increase the amount spent in the district; or 2) decrease the number of students in the district. As smaller districts are shown to perform worse, through a quantile analysis based on Mem, then dollar-for-dollar, districts with less students have higher per-pupil-expenditure. So, it makes sense that an increase in per-pupil-expenditure is less associated with larger spending and more associated with smaller districts that perform worse.

For FRLProp, the proportion of students in a district on free-or-reduced lunch, the level of significance varies from being statistically insignificant to highly significant. There is no consistent directional relationship over years. Some relationships are positive and some are negative. In most years, the relationship is negative. Generally, as the proportion of students in a district on free-or-reduced lunch rises, the average number of students in a district passing the ISAT mathematics exam falls.

The largest value occurs in 2010 of -0.1590688, such that, for a one-percentage point increase FRL is associated with about -0.16 percentage points in AvgAllP. Put differently, a 6.29-percentage point increase in the proportion of students on FRL is associated with a one-percentage point decrease in AvgAllP. This seems economically insignificant.

For 4Day, an indicator if a school district uses a 4-day school-week schedule, the level of significance varies from being statistically insignificant to highly significant. There is no consistent directional relationship over years. Some relationships are positive and some are negative. In most years, the relationship is negative. Generally, if a school district employs a 4-day school-week schedule, the average number of students in a district passing the ISAT

mathematics exam falls.

The relationship, while not consistently statistically significant, seems economically significant in that being on a 4-day schedule is associated with no less than a single percentage point change in AvgAllP in any time frame examined. The most economically significant relationship occurred in 2010 where 4Day is associated with a decrease in AvgAllP of almost 9.5 percentage points. As this is an indicator variable, this difference is associated with a shift in intercept, moving the analysis down linearly.

At all points, the intercept term is statistically significant and has a positive association with the proportion of students in a district passing the ISAT mathematics exam. However, the association is as small as about 18.09 and as large as about 59.02 percentage points of average students in a district passing the ISAT mathematics exam. The largest relationship occurs in 2007 of about 59.0193 percentage points of AvgAllP and the smallest relationship occurs in 2013 of about 18.095 percentage points of AvgAllP.

Of all the years, the R-squared of the model is highest in 2007, with 0.2958. The smallest R-squared of the model over each of the years occurs in 2008 with 0.0731. See Figure 6.1 for more details.

From simple observation, I can see that by this method, the coefficients of interest are not the same at every period. The hypothesis tested of the coefficients being the same independent of time frame is false. What follows are a few different approaches to this same question with different models and tests.

7. Separate Approaches

In Figure 7.1, the estimates for the same model are shown, but with robust (White-Huber) standard errors. Still, the initial hypothesis that the coefficients hold the same value

regardless of time period fails. There is no notable difference in the levels of significance between the standard errors and robust standard errors approach. This gives some intuition about the homoskedasticity of the data.

Next, I pool all years together, perform the same regression and test whether the coefficients estimated for each year pooled together are equal to the coefficients estimated for individual years.

$$(4)\ F_{chow} = \frac{[RSS_p - (\sum_{i=2007}^{2013} RSS_i)]/[Df_p - (\sum_{i=2007}^{2013} Df_i)]}{(\sum_{i=2007}^{2013} RSS_i)/Df_p} \sim F_{[Df_p - (\sum_{i=2007}^{2013} RSS_i)],Df_p}$$

$$(5)\ 208.79923 = \frac{[2130765.99 - (1230091.9)]/[6844 - 6820]}{1230091.9/6844} \sim F_{24,6844}$$

$$(6)\ Pvalue = 2.6 * 10^{-24}$$

The hypothesis of whether the coefficients estimated over this pooled group of the data are equal to the coefficients estimated over individual years is rejected. Pooling is not appropriate.

Last, I examine the model applied to students in school grade 3 in 2007 transitioning through time to grade 9 in 2013. The results are seen in Figure 7.2. Note, this evaluates the relationship of district level variables on school grade level variables. The dependent variable has changed from the base model's average proportion of students scoring proficient to the proportion of students scoring proficient in specific school grade.

It's quickly seen that this model yields different results than the base model – nearly none of the variables are statistically or economically significant at any point and the signs of the coefficients (positive or negative) are different from the base model.

8. Concluding Remarks

From the base model, I've shown the relationship between the proportion of students in a school district scoring proficient and real state per-pupil-expenditure, the proportion of students on free-or-reduced lunch, and if the school district employs a 4-day school-week schedule. The

relationship between each of these variables is generally negative. As shown from a Chow Test, the model is not appropriate for estimating all years at the same time. This model is not appropriate for examining school grades instead of school districts.

Further research could be toward better specification of the model for cross sections, applying time series analysis, and applying panel-data analysis. The data is large and unbalanced in many instances. This could also provide for opportunities to apply different techniques to the analysis.

9. References

Anderson, D. Mark & Walker, Mary B. (2012) Does Shortening the School Week Impact Student Performance? Evidence from the Four-Day School Week. *Andrew Young School of Policy Studies Research Paper Series, 12-06*.

Burtless, G. T. (1996). Does money matter?: The effect of school resources on student achievement and adult success. Washington, D.C: Brookings Institution Press.

Card, D & Krueger, A. B. (September 01, 1998). School Resources and Student Outcomes. The Annals of the American Academy of Political and Social Science, 559, 1, 39-53.

Connell, R. W. (September 06, 1994). Poverty and Education. *Harvard Educational Review, 64,* 2, 125-49.

Cooper, B. S., Sarrel, R., Darvas, P., Alfano, F., Meier, E., Samuels, J., & Heinbuch, S. (July 01, 1994). Making Money Matter in Education: A Micro-Financial Model for Determining School-Level Allocations, Efficiency, and Productivity. Journal of Education Finance, 20, 1, 66-87.

Dolan, Robert C. & Schmidt, Robert M. (1987) Assessing the Impact of Expenditure on Achievement: Some Methodological and Policy Considerations. *Economics of Education Review, 6, 3,* 285-299.

Elliott, M. (July 01, 1998). School Finance and Opportunities to Learn: Does Money Well Spent Enhance Students' Achievement? Sociology of Education, 71, 3, 223-45.

Flaherty, S. (March 01, 2013). Does money matter in Pennsylvania school district spending and student proficiency since no child left behind. Eastern Economic Journal, 39, 2, 145-171.

Greenwald, R., Hedges, L. V., & Laine, R. D. (December 07, 1996). The Effect of School Resources on Student Achievement. Review of Educational Research, 66, 3, 361-96.

Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature, 24, 3*, 1141-1171.

Hanushek, E. A. (September 06, 1997). Assessing the Effects of School Resources on Student Performance: An Update. Educational Evaluation and Policy Analysis, 19, 2, 141-64.

Herndon, J. S., & Bembenutty, H. (October 01, 2014). In-school and social factors influencing learning among students enrolled in a disciplinary alternative school. *Learning and Individual Differences, 35,* 49-55.

Heyns, B. (July 01, 1997). Book Review: Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success. Contemporary Sociology, 26, 4, 488-489.

Klein, C. C. (September 06, 2008). Intradistrict Public School Funding Equity, Community Resources, and Performance in Nashville, Tennessee. Journal of Education Finance, 34, 1, 1-14.

Knoeppel, R. C., Verstegen, D. A., & Rinehart, J. S. (December 07, 2007). What Is the Relationship between Resources and Student Achievement? A Canonical Analysis. Journal of Education Finance, 33, 2, 183-202.

Lopus, Jane S. (1990) Do Additional Expenditures Increase Achievement in the High School Economics Class? *The Journal of Economic education, 21, 3,* 277-286.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research.* New York, NY: Cambridge University Press.

Neymotin, F. (June 01, 2010). The relationship between school funding and student achievement in Kansas public schools. Journal of Education Finance, 36, 1, 88-108.

Nyhan, R. C., & Alkadry, M. G. (October 01, 1999). The Impact of School Resources on Student Achievement Test Scores. Journal of Education Finance, 25, 2, 211-227.

Papke, L. (2004). The effects of spending on test pass rates: evidence from Michigan, *Journal of Public Economics,* 89, 821-839.

Papke, L., & Wooldridge, J. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics, 145*, 121–133.

Payne, K. J., & Biddle, B. J. (August 01, 1999). Poor School Funding, Child Poverty, and Mathematics Achievement. Educational Researcher, 28, 6, 4-13.

Pike, G. R., Smart, J. C., Kuh, G. D., & Hayek, J. C. (November 01, 2006). Educational expenditures and student engagement: When does money matter?. Research in Higher Education : Journal of the Association for Institutional Research, 47, 7, 847-872.

Turner, S. E. (June 01, 2000). A Comment on "Poor School Funding, Child Poverty, and Mathematics Achievement.". Educational Researcher, 29, 5, 15-18.

Verstegen, D. A., & King, R. A. (December 07, 1998). The Relationship Between School Spending and Student Achievement: A Review and Analysis of 35 Years of Production Function Research. Journal of Education Finance, 24, 2, 243-62.

Wenglinsky, H. (July 01, 1997). How Money Matters: The Effect of School District Spending on Academic Achievement. Sociology of Education, 70, 3, 221-37.

A. Appendix: Figures

Average Proportion of Students in a District Scoring Proficient on the ISAT Math									
Year	Min.	1-25%	25-50%	50-75%	75-99%	Max.	Mean	% Δ in Mean	No. of Obs
2007	0	37.9	42.619	46.435	60.8	75.5	38.46653	.	980
2008	0	26.4	38.321	42.983	55.05	61.4	31.195	-18.904%	1090
2009	0	15.6	37.499	43.017	55.7	71.4	29.96909	-3.930%	1116
2010	0	29.2	39.411	44	53.743	60	32.63941	8.910%	1278
2011	0	26.33	36.962	41.8	55.733	63.6	30.82826	-5.549%	1298
2012	0	23.4	34.856	39.25	50.513	55	29.02329	-5.855%	1304
2013	0	27.33	37.542	41.935	55.8	90.9	31.71105	9.261%	1284
Figure 5.1, Source: Idaho State Department of Education									

Average Daily Attendance (Number of Students in a District)									
Year	Min.	1-25%	25-50%	50-75%	75-99%	Max.	Mean	% Δ in Mean	No. of Obs
2007	3.29	282.81	750.97	1704.23	23356.13	29900.59	2083.734	.	931
2008	4.4	284.69	732	2011.2	23528.54	30972.16	2112.431	1.377%	1053
2009	9.67	285.4	708.52	2039.4	23322.19	31817.88	2105.429	-0.331%	1074
2010	4.74	270.37	619.86	1727.82	23659.93	32501.54	2063.509	-1.991%	1230
2011	5.13	251.56	593.01	1588.01	23885.83	33383.06	2034.138	-1.423%	1251
2012	5.13	251.56	590.61	1516.98	23885.83	33383.06	1992.376	-2.053%	1285
2013	9.52	253.44	590.41	1597.14	23982.48	31944.91	1998.717	0.318%	1277
Figure 5.2, Source: Idaho State Department of Education									

Inflated FY 2013 State Per-Pupil-Expenditure (USD, $)									
Year	Min.	1-25%	25-50%	50-75%	75-99%	Max.	Mean	% Δ in Mean	No. of Obs
2007	$ 5,364.61	$ 6,150.17	$ 6,983.68	$ 8,472.43	$ 29,696.57	$ 39,798.00	$ 7,942.16	.	931
2008	$ 5,025.34	$ 5,963.82	$ 6,767.91	$ 8,414.93	$ 25,480.77	$ 29,364.71	$ 7,777.72	-2.071%	1053
2009	$ 5,108.01	$ 6,295.27	$ 6,850.81	$ 8,314.85	$ 17,472.58	$ 26,521.56	$ 7,829.49	0.666%	1074
2010	$ 4,011.88	$ 5,157.38	$ 5,811.26	$ 7,256.80	$ 13,302.08	$ 27,653.17	$ 6,573.30	-16.044%	1230
2011	$ 2,534.41	$ 5,467.85	$ 6,074.76	$ 7,573.10	$ 20,197.80	$ 24,171.68	$ 7,028.27	6.921%	1251
2012	$ 3,860.79	$ 4,878.88	$ 5,422.45	$ 6,709.71	$ 20,079.50	$ 23,320.63	$ 6,344.47	-9.729%	1285
2013	$ 3,857.86	$ 4,995.17	$ 5,474.10	$ 6,912.93	$ 16,180.03	$ 23,309.86	$ 6,437.68	1.469%	1277
Figure 5.3, Source: Idaho State Department of Education									

WHAT IS THE RELATIONSHIP BETWEEN STATE SPENDING AND STANDARDIZED 16 TEST SCORES IN IDAHO?

Summary of Relevant Modeled Data							
Variable		Mean	Stand. Dev.	Min	Max	Observations	
AVGAllP	overall	11.88851	18.19027	0	63.2	N	8350
	between		12.67669	0	58.66666	n	2049
	within		14.13391	-28.91461	66.05994	T-bar	4.07516
Mem	overall	2051.717	4137.491	3.29	33383.06	N	8101
	between		4032.694	3.845	32903.68	n	1958
	within		135.7904	-75.77399	3534.009	T-bar	4.13739
InfStatePPE	overall	7066.295	2883.077	2534.409	39798	N	8101
	between		3048.369	3429.716	34581.36	n	1958
	within		923.2097	-2831.232	18469.92	T-bar	4.13739
FRLProp	overall	52.89006	14.49145	0	107.8924	N	6848
	between		14.09663	0	104.8041	n	1577
	within		4.848523	32.14344	78.65206	T-bar	4.34242
4Day	overall	0.2135271	0.4098253	0	1	N	7156
	between		0.37924	0	1	n	1662
	within		0.1908344	-0.6198062	1.07067	T-bar	4.30566

Figure 5.4, Sources: Idaho State Department of Education & Idaho Kids Count

Model 1: AvgAllP							
	Year: 2007	Year: 2008	Year: 2009	Year: 2010	Year: 2011	Year: 2012	Year: 2013
InfStatePPE	***	***	***	***	***	***	***
Beta1	-0.0023901	-0.0010090	-0.0011103	-0.0020283	-0.0015058	-0.0010848	-0.0017485
se	0.0001293	0.0001621	0.0001350	0.0002047	0.0001710	0.0001580	0.0002355
t	-18.49	-6.22	-8.23	-9.91	-8.80	-6.87	-7.43
P > |t|	0.000	0.000	0.000	0.000	0.000	0.000	0.000
FRLProp		***	***	***	***	*	*
Beta2	0.0221024	-0.1182496	-0.1089385	-0.1590688	-0.1095783	-0.0708443	0.0863567
se	0.0243528	0.0293547	0.0245464	0.0360178	0.3152090	0.0311706	0.0347236
t	0.91	-4.03	-4.44	-4.42	-3.48	-2.27	2.49
P > |t|	0.364	0.000	0.000	0.000	0.001	0.023	0.013
4Day		***	***	***		***	***
Beta3	1.5704470	-5.7818830	-5.4651850	-9.4849340	-1.7536880	-6.1161760	-5.6739160
se	1.1129620	1.3171920	1.0965780	1.2351810	1.0218590	0.9488680	1.0921800
t	1.41	-4.39	-4.98	-7.68	-1.72	-6.45	-5.20
P > |t|	0.159	0.000	0.000	0.000	0.086	0.000	0.000
Constant	***	***	***	***	***	***	***
Beta0	59.0193300	19.3210800	19.3675900	32.3277100	25.2959500	20.4155300	18.0950600
se	1.6506200	1.9704900	1.7388850	2.4937660	2.0258650	1.9870380	2.2915440
t	35.76	9.81	11.14	12.96	12.49	10.27	7.90
P > |t|	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. Obs	835	927	933	1048	1042	1044	1019
R^2	0.2958	0.0731	0.1008	0.1522	0.1021	0.1086	0.1064
	F(3, 831) = 116.34	F(3, 104) = 2.63	F(3, 929) = 34.73	F(3, 1044) = 62.46	F(3, 1038) = 39.33	F(3, 1040) = 42.24	F(3, 1019) = 40.27
(P<=0.001) = ***, (P<=0.01) = **, (P<=0.05) = *							
Figure 6.1, Base Model Regression							

Model 2: AvgAllP							
	Year: 2007	Year: 2008	Year: 2009	Year: 2010	Year: 2011	Year: 2012	Year: 2013
InfStatePPE	***	***	***	***	***	***	***
Beta1	-0.0023295	0.0010090	-0.0011103	-0.0020283	-0.0015058	-0.0010848	-0.0017485
robust se	0.0002131	0.0001196	0.0001189	0.0002311	0.0001425	0.0001183	0.0001641
t	-11.21	-8.44	-9.34	-8.78	-10.57	-9.17	-10.66
P > |t|	0.000	0.000	0.000	0.000	0.000	0.000	0.000
FRLProp		***	***	***	***	*	*
Beta2	0.0221024	-0.1182496	-0.1089385	-0.1590688	-0.1095783	-0.0708443	0.0863567
robust se	0.0345744	0.0266444	0.0202922	0.0334193	0.0250350	0.0317408	0.0345284
t	0.64	-4.44	-5.37	-4.76	-4.38	-2.23	2.50
P > |t|	0.523	0.000	0.000	0.000	0.000	0.026	0.013
4Day		***	***	***		***	***
Beta3	1.570447	-5.781883	-5.465185	-9.484934	-1.753688	-6.116176	-5.673916
robust se	1.298304	0.518816	0.481214	0.658888	0.960028	0.767972	0.989605
t	1.21	-11.14	-11.36	-14.40	-1.83	-7.96	-5.73
P > |t|	0.227	0.000	0.000	0.000	0.068	0.000	0.000
Constant	***	***	***	***	***	***	***
Beta0	59.01933	19.32108	19.36759	32.32771	25.29595	20.41553	18.09506
robust se	2.3889530	1.9677410	1.8240000	2.3737040	1.9142980	2.0128360	2.0947670
t	24.71	9.82	10.62	13.62	13.21	10.14	8.64
P > |t|	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. Obs	835	927	933	1048	1042	1044	1019
R^2	0.2958	0.0731	0.1008	0.1522	0.1021	0.1086	0.1064
	F(3, 831) = 45.47	F(3, 923) = 47.09	F(3, 929) = 48.38	F(3, 1044) = 105.5	F(3, 1038) = 52.56	F(3, 1040) = 71.13	F(3, 1015) = 65.13
(P<=0.001) = *** , (P<=0.01) = **, (P<=0.05) = *							
Figure 7.1, Robust SE							

Model 3: AllP							
	Grade: 3, Year: 2007	4, 2008	5, 2009	6, 2010	7, 2011	8, 2012	9, 2013
InfStatePPE							
Beta1	-0.0008410	-0.0004293	0.0014749	-0.0004196	-0.0005514	-0.0007632	0.0003370
robust se	0.0010681	0.0009018	0.0007958	0.0010420	0.0004907	0.0004832	0.0006593
t	-0.79	-0.48	1.85	-0.40	-1.12	-1.58	0.51
P > |t|	0.433	0.635	0.067	0.688	0.264	0.117	0.610
FRLProp	*			*			
Beta2	0.1583717	0.0899424	0.0147950	0.1999410	-0.0079960	0.4600950	-0.1265447
robust se	0.0612901	0.0734372	0.0649830	0.0798938	0.0672941	0.0728162	0.0644439
t	2.58	1.22	0.23	2.50	-0.12	0.63	-1.96
P > |t|	0.011	0.224	0.820	0.014	0.906	0.529	0.052
4Day			**		*		
Beta3	2.7582300	3.0918270	5.3425840	1.9246430	4.9103370	-0.1759084	-0.3438892
robust se	2.2964570	2.5085210	1.9469400	3.2679510	2.3685180	2.6710170	2.0203640
t	1.20	1.23	2.74	0.59	2.07	-0.07	-0.17
P > |t|	0.233	0.221	0.007	0.557	0.041	0.948	0.865
Constant	***	***	***	***	***	***	***
Beta0	36.4501600	45.2849200	32.6935400	30.8870400	45.1685600	45.0686100	43.2568500
robust se	7.1034270	5.8974910	4.8734770	5.1421400	4.5692460	4.4102950	5.3999230
t	5.13	7.68	6.71	6.01	9.89	10.22	8.01
P > |t|	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. Obs	96	100	99	101	102	102	99
R^2	0.061	0.0281	0.1297	0.0673	0.0665	0.0312	0.045
	F(3, 92) = 2.31	F(3, 96) = 1.08	F(3, 95) = 8.85	F(3, 97) = 2.55	F(3, 98) = 2.75	F(3, 98) =0.86	F(3, 95) = 1.56
(P<=0.001) = ***, (P<=0.01) = **, (P<=0.05) = *							

Figure 7.2, Regression with robust errors, following grade 3 in 2007 advancing to grade 9 in 2012, on the proportion of students scoring proficient on the ISAT Mathematics

B. Appendix: Pertinent Script

```
/*----------------------------*/

* Main Analysis - X-sectional

/*----------------------------*/

clear

*iMac

use "/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/App Metrics/Project/ISATMetrics2.dta"

*BigLaptop

*use "/Users/2011home/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/App Metrics/Project/ISATMetrics2.dta"

set more off



forvalues i = 0(1)6 {

	display 2013 - `i'

	gen interest`i' = 0

	replace interest`i' = 1 if Grade == 9 - `i' & Yr == 2013 - `i'

}



* Creating a district average of passing score

gen numAllP = 1

gen avgAllP = 0

sort Yr Dist Grade

forvalues j = 2007(1)2013 {

	bys Yr Dist: replace avgAllP = sum(allp) / sum(numAllP) if Dist == Dist & Yr == `j' & allp != .

}

bys Yr Dist: replace avgAllP = avgAllP[_N]



replace FRLProp = FRLProp *100



sum avgAllP Mem InfStatePPE FRLProp Day

bys Yr: sum avgAllP, detail

bys Yr: sum Mem, detail

bys Yr: sum InfStatePPE, detail

bys Yr: sum FRLProp

bys Yr: sum Day



* Base Model: Regressing district x-sections starting in 2007 going through 2013

forvalues i = 2007(1)2013 {

	dis "year `i' "

	reg avgAllP InfStatePPE FRLProp Day if Yr == `i'

}



* Base Model: including robust se

forvalues i = 2007(1)2013 {

	dis "year `i' "

	reg avgAllP InfStatePPE FRLProp Day if Yr == `i' , r

}



* Chow Test:

*First regress on pooled -

reg avgAllP InfStatePPE FRLProp Day

*Next, perform the F-test for the chow

gen sumResOth = 87247.9645 + 147654.003 + 110011.85 + 235850.349 + 212775.523 + 194540.266 + 242011.896

gen sumDFOth = 831 + 923 + 929 + 1044 + 1038 + 1040 + 1015

display (2130765.99 - (sumResOth))/(6844 - (sumDFOth)) / ((sumResOth) / 6844)

* F = 208.79923

display Ftail(6844,6844 - (sumDFOth),208.79923)

* P = 2.620e-24

* Very close to zero, reject the hypothesis



* Grade Model: including robust se

forvalues i = 3(1)9 {

	dis "Grade `i' in year " 2004 + `i'

	reg allp InfStatePPE FRLProp Day if Grade == `i' & Yr == 2004 + `i' , r

}



xtset cohort Yr

xtsum avgAllP Mem InfStatePPE FRLProp Day



/*----------------------------*/

*	COMPLETE

/*----------------------------*/
```

C. Appendix: Other Script

```
clear

use "/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/App Metrics/Project/ISATMetrics2.dta"

set more off

sort Grade

by Grade: sum Yr

sort Yr

by Yr: sum Grade

* Grade 6 appears most frequently, so I drop all others to so that the proportion passing, the independent to be measured, is appropriate at the district level.

*By only observing one grade per district, we only observe each district once instead of 10 times, once for each grade, where the proportion passing is given at the school district's grade level and the membership is given at the district level (the bases of measurement are different). I select year to be 2013 as that is the base for all other measurements.

keep if Grade == 6

keep if Yr == 2013

*IQR

sort Yr

xtile MemIQR = Mem, n(4)

tab MemIQR, gen(iq)

* Creates 4 variables - iq1, iq2, iq3, iq4 - by innerquartile ranges of about 25%.


* Observe first quartile


keep if iq1 == 1 & Yr == 2013

egen AvgMem=mean(Mem)

gen AbsMem=abs(AvgMem-Mem)

sort AbsMem

keep in 1

browse Dist

* Dist 11 is closest to the mean of the 1'st quartile when selecting only on the 6th Grade in 2013. When desiring to check only IQR1, load data and "keep if Dist == 11"


* Observe second quartile

clear

use "/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/App Metrics/Project/ISATMetrics2.dta"

set more off

keep if Grade == 6

keep if Yr == 2013

*IQR

sort Yr

xtile MemIQR = Mem, n(4)

tab MemIQR, gen(iq)

sum Mem


keep if iq2 == 1 & Yr == 2013

egen AvgMem=mean(Mem)

gen AbsMem=abs(AvgMem-Mem)

sort AbsMem

keep in 1

browse Dist

* Dist 135 is closest to the mean of the 2'nd quartile when selecting only on the 6th Grade in 2013. When desiring to check only IQR3, load data and "keep if Dist == 135"


* Observe third quartile

clear

use "/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/App Metrics/Project/ISATMetrics2.dta"

set more off

keep if Grade == 6

keep if Yr == 2013


*IQR

sort Yr

xtile MemIQR = Mem, n(4)

tab MemIQR, gen(iq)

sum Mem


keep if iq3 == 1 & Yr == 2013

egen AvgMem=mean(Mem)

gen AbsMem=abs(AvgMem-Mem)

sort AbsMem

keep in 1

browse Dist

* Dist 372 is closest to the mean of the 3'rd quartile when selecting only on the 6th Grade in 2013. When desiring to check only IQR3, load data and "keep if Dist == 372"


* Observe fourth quartile

clear

use "/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/App Metrics/Project/ISATMetrics2.dta"

set more off

keep if Grade == 6

keep if Yr == 2013


*IQR

sort Yr

xtile MemIQR = Mem, n(4)

tab MemIQR, gen(iq)

sum Mem


keep if iq4 == 1 & Yr == 2013

egen AvgMem=mean(Mem)

gen AbsMem=abs(AvgMem-Mem)

sort AbsMem

keep in 1

browse Dist

* Dist 132 is closest to the mean of the 4th quartile when selecting only on the 6th Grade in 2013. When desiring to check only IQR3, load data and "keep if Dist == 132"


/*--------------------------*/

*       Time Series

/*--------------------------*/

clear

use "/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/App Metrics/Project/ISATMetrics2.dta"

set more off

keep if Grade == 6


egen AvgMem=mean(Mem)

gen AbsMem=abs(AvgMem-Mem)

sort AbsMem

keep in 1

browse Dist

* District 60 is the one nearest the mean.

clear

use "/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/App Metrics/Project/ISATMetrics2.dta"

set more off

keep if Grade == 6 & Dist == 60

browse

sum Yr allp InfTotPPE InfStatePPE FRLProp Day


clear

*SmallMBP

*use "/Users/Cdaigle/Dropbox/Education/BSU/2015 2016/Spring/Hon 491/ISAT/Districts/STATA/MATHEMATICSISATDistrictMaster03262016.dta"

*iMac

*use "/Users/2011home/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/App Metrics/Project/MATHEMATICSISATDistrictMaster03262016.dta"

*BigLaptop

use "/Users/2011home/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/App Metrics/Project/MATHEMATICSISATDistrictMaster03262016.dta"


set more off


/*--------------------------------------*/

* Generate cohort numbers, starting at 1 in each district

sort dist year grade

gen int cohort = (grade-year) + 2005

replace cohort = 11-cohort if cohort < 1


* Cohorts distinctive numbering

by dist, sort: gen n_dist = 1 if _n == 1

replace n_dist = sum(n_dist)

quietly summ cohort

quietly replace cohort = cohort + (n_dist-1)*r(max)

quietly list, noobs clean


/*--------------------------------------*/

* Rename

rename dist Dist

rename year Yr

rename grade Grade

rename annualexpenditure AnnExp

rename taxes Tax

rename otherlocalsources OthLcl

rename statesources State
```

```
rename federalsources Fed

rename othersources Oth

rename membership Mem

rename totalperpupilexpenditure TotPPE

rename dayweeks Day

rename frlnumber FRL

rename estimatedtotalpopulationofthesch PopEstDist

rename estimatedpopulation517childrenin Pop517EstDist

rename povestimatednumberofrelevantchil PovPop517Dist

rename fyinflationindex InfInd13

rename annualexpenditureadjustedforinfl AnnInfExp

rename totalppeadjustedforinflation InfPPE

rename PovPop517Dist Pov


* Labels

label variable Dist "District number"

label variable Yr "Year of observation (2007-2013)"

label variable Grade "Grade of tested students in a district (3-12)"

label variable AnnExp "Total annual expenditure in a district for a particular year"

label variable Tax "Total amount of tax revenue in a district for a particular year"

label variable OthLcl "Total amount of revenue from local sources not contained in other revenue streams"

label variable State "Total amount of revenue from state sources"

label variable Fed "Total amount of revenue from federal sources"

label variable Oth "Total amount of revenue from sources not contained in any other revenue stream"

label variable Mem "Total number of students in a district on a particular day"

label variable TotPPE "Total per-pupil-expenditure by district"

label variable Day "Binary variable indicating if a school district has 4-day school weeks"

label variable FRL "Total number of students in a district receiving free or reduced lunch"

label variable PopEstDist "Estimate from US Census of the population in a district"

label variable Pop517EstDist "Estimate from US Census of the population aged 5 to 17 years in a district"

label variable Pov "Estimate from US Census of the population aged 5 to 17 years living in poverty in a district"

label variable InfInd13 "Index for Inflation in a fiscal year with fiscal year 2013 as the base year (Jul-Jun)"

label variable AnnInfExp "Annual expenditure Inflated to 2013 dollars"

label variable InfPPE "Per-pupil-expenditure Inflated to 2013 dollars"

label variable allss "Average scaled score for all tested"

label variable allbb "Percent of below basic for all tested"

label variable allb "Percent of basic for all tested"

label variable allp "Percent of proficient for all tested"

label variable alla "Percent of advanced for all tested"

label variable maless "Average scaled score for all Males tested"

label variable maletested "Number of Males tested"

label variable malebb "Percent of below basic for Males tested"

label variable maleb "Percent of basic for Males tested"

label variable malep "Percent of proficient for Males tested"

label variable malea "Percent of advanced for Males tested"

label variable femaless "Average scaled score for all Females tested"

label variable femaletested "Number of Females tested"

label variable femalebb "Percent of below basic for Females tested"

label variable femaleb "Percent of basic for Females tested"

label variable femalep "Percent of proficient for Females tested"

label variable femalea "Percent of advanced for Females tested"

label variable aianss "Average scaled score for all American Indian or Alaskan Native tested"

label variable aiantested "Number of American Indian or Alaskan Native tested"

label variable aianbb "Percent of below basic for American Indian or Alaskan Native tested"

label variable aianb "Percent of basic for American Indian or Alaskan Native tested"

label variable aianp "Percent of proficient for American Indian or Alaskan Native tested"

label variable aiana "Percent of advanced for American Indian or Alaskan Native tested"

label variable asianss "Average scaled score for all Asian or Pacific Islander tested"

label variable asiantested "Number of Asian or Pacific Islander tested"

label variable asianbb "Percent of below basic for Asian or Pacific Islander tested"

label variable asianb "Percent of basic for Asian or Pacific Islander tested"

label variable asianp "Percent of proficient for Asian or Pacific Islander tested"

label variable asiana "Percent of advanced for Asian or Pacific Islander tested"

label variable bafamss "Average scaled score for all Black / African American tested"

label variable bafamtested "Number of Black / African American tested"

label variable bafambb "Percent of below basic for Black / African American tested"

label variable bafamb "Percent of basic for Black / African American tested"

label variable bafamp "Percent of proficient for Black / African American tested"

label variable bafama "Percent of advanced for Black / African American tested"

label variable nhopiss "Average scaled score for all Native Hawaiian / Other Pacific Islander tested"

label variable nhopitested "Number of Native Hawaiian / Other Pacific Islander tested"

label variable nhopibb "Percent of below basic for Native Hawaiian / Other Pacific Islander tested"

label variable nhopib "Percent of basic for Native Hawaiian / Other Pacific Islander tested"

label variable nhopip "Percent of proficient for Native Hawaiian / Other Pacific Islander tested"

label variable nhopia "Percent of advanced for Native Hawaiian / Other Pacific Islander tested"

label variable whitess "Average scaled score for all White tested"

label variable whitetested "Number of White tested"

label variable whitebb "Percent of below basic for White tested"

label variable whiteb "Percent of basic for White tested"

label variable whitep "Percent of proficient for White tested"

label variable whitea "Percent of advanced for White tested"

label variable hisplatss "Average scaled score for all Hispanic or Latino tested"

label variable hisplattested "Number of Hispanic or Latino tested"

label variable hisplatbb "Percent of below basic for Hispanic or Latino tested"

label variable hisplatb "Percent of basic for Hispanic or Latino tested"

label variable hisplatp "Percent of proficient for Hispanic or Latino tested"

label variable hisplata "Percent of advanced for Hispanic or Latino tested"

label variable tworacesss "Average scaled score for all Other/Unknown tested"

label variable tworacestested "Number of Other/Unknown tested"

label variable tworacesbb "Percent of below basic for Other/Unknown tested"

label variable tworacesb "Percent of basic for Other/Unknown tested"

label variable tworacesp "Percent of proficient for Other/Unknown tested"

label variable tworacesa "Percent of advanced for Other/Unknown tested"

label variable frlss "Average scaled score for all Free or Reduced Lunch tested"

label variable frltested "Number of Free or Reduced Lunch tested"

label variable frlbb "Percent of below basic for Free or Reduced Lunch tested"

label variable frlb "Percent of basic for Free or Reduced Lunch tested"

label variable frlp "Percent of proficient for Free or Reduced Lunch tested"

label variable frla "Percent of advanced for Free or Reduced Lunch tested"

label variable lepss "Average scaled score for all Limited English Proficient tested"

label variable leptested "Number of Limited English Proficient tested"

label variable lepbb "Percent of below basic for Limited English Proficient tested"

label variable lepb "Percent of basic for Limited English Proficient tested"

label variable lepp "Percent of proficient for Limited English Proficient tested"

label variable lepa "Percent of advanced for Limited English Proficient tested"

label variable migss "Average scaled score for all Migrant tested"

label variable migtested "Number of Migrant tested"

label variable migbb "Percent of below basic for Migrant tested"

label variable migb "Percent of basic for Migrant tested"

label variable migp "Percent of proficient for Migrant tested"

label variable miga "Percent of advanced for Migrant tested"

label variable spess "Average scaled score for all Special Education tested"

label variable spetested "Number of Special Education tested"

label variable spebb "Percent of below basic for Special Education tested"

label variable speb "Percent of basic for Special Education tested"

label variable spep "Percent of proficient for Special Education tested"

label variable spea "Percent of advanced for Special Education tested"

label variable tiass "Average scaled score for all Title 1 A tested"

label variable tiatested "Number of Title 1 A tested"

label variable tiabb "Percent of below basic for Title 1 A tested"
```

```
label variable tiab "Percent of basic for Title 1 A tested"

label variable tiap "Percent of proficient for Title 1 A tested"

label variable tiaa "Percent of advanced for Title 1 A tested"



/*---------------------------*/

* Set Panel Data

xtset cohort Yr



/*---------------------------*/

* Variable Creation

/*---------------------------*/

*Year Binary

gen Yr07 = 0

replace Yr07 = 1 if Yr == 2007

gen Yr08 = 0

replace Yr08 = 1 if Yr == 2008

gen Yr09 = 0

replace Yr09 = 1 if Yr == 2009

gen Yr10 = 0

replace Yr10 = 1 if Yr == 2010

gen Yr11 = 0

replace Yr11 = 1 if Yr == 2011

gen Yr12 = 0

replace Yr12 = 1 if Yr == 2012

gen Yr13 = 0

replace Yr13 = 1 if Yr == 2013



*Inflation

gen CPIBase =
(231.893+233.001+234.083+234.966+23
3.206+232.029+232.759+234.595+235.5
11+235.488+235.979+236.227)/12

gen Seven =
(206.7+207.5+207.8+207.1+206.3+206.2
+207.79+208.995+210.778+212.036+213
.063+212.68)/12

gen Eight =
(212.542+212.406+212.92+213.917+214
.904+214.733+215.739+216.339+218.53
3+219.437+221.009+223.04)/12

gen Nine =
(223.867+222.823+222.132+221.034+21
7.113+214.685+215.923+217.095+217.3
57+217.91+218.567+219.865)/12

gen Ten =
(219.484+219.884+220.294+220.447+21
9.728+219.307+219.989+220.179+220.8
09+221.202+221.417+221.147)/12

gen Eleven =
(221.331+221.523+221.384+221.708+22
1.671+222.081+223.149+224.431+226.5
58+227.837+228.516+228.075)/12

gen Twelve =
(227.805+228.222+229.147+229.195+22
8.771+228.117+228.98+229.995+232.03
9+232.561+233.053+232.701)/12

gen Thirt =
(231.893+233.001+234.083+234.966+23
3.206+232.029+232.759+234.595+235.5
11+235.488+235.979+236.227)/12

gen CPI07 = CPIBase / Seven

gen CPI08 = CPIBase / Eight

gen CPI09 = CPIBase / Nine

gen CPI10 = CPIBase / Ten

gen CPI11 = CPIBase / Eleven

gen CPI12 = CPIBase / Twelve

gen CPI13 = CPIBase / Thirt

replace InfInd13 = CPI07 if Yr == 2007

replace InfInd13 = CPI08 if Yr == 2008

replace InfInd13 = CPI09 if Yr == 2009

replace InfInd13 = CPI10 if Yr == 2010

replace InfInd13 = CPI11 if Yr == 2011

replace InfInd13 = CPI12 if Yr == 2012

replace InfInd13 = CPI13 if Yr == 2013

gen InfExp = AnnExp*InfInd13

gen LInfExp = log(InfExp)

gen InfTax = Tax*InfInd13

gen LInfTax = log(InfTax)

gen InfOthLcl = OthLcl*InfInd13

gen LInfOthLcl = log(InfOthLcl)

gen InfState = State*InfInd13

gen LInfState = log(InfState)

gen InfFed = Fed*InfInd13

gen LInfFed = log(InfFed)

gen InfOth = Oth*InfInd13

gen LInfOth = log(InfOth)



*IQR

bys Yr: sum Mem, detail



*Mem

gen LMem = log(Mem)



*MemIQR1

gen MemIQR1 = 0

replace MemIQR1 = 1 if Mem <= 282.81 &
Yr07 == 1

replace MemIQR1 = 1 if Mem <=284.69 &
Yr08 == 1

replace MemIQR1 = 1 if Mem <=285.4 &
Yr09 == 1

replace MemIQR1 = 1 if Mem <=270.37 &
Yr10 == 1

replace MemIQR1 = 1 if Mem <=251.56 &
Yr11 == 1

replace MemIQR1 = 1 if Mem <=251.56 &
Yr12 == 1

replace MemIQR1 = 1 if Mem <=253.44 &
Yr13 == 1

replace MemIQR1 = . if MemIQR1 == 0



gen IntMemIQR1 = 1 if MemIQR1 == 1

replace IntMemIQR1 = . if MemIQR1 == 0



gen BMemIQR1 = MemIQR1

replace BMemIQR1 = 0 if
missing(BMemIQR1)



gen LMemIQR1 = log(MemIQR1)



*MemIQR2

gen MemIQR2 = 0

replace MemIQR2 = 1 if Mem > 282.81 &
Mem <= 750.97 & Yr07 == 1

replace MemIQR2 = 1 if Mem > 284.69 &
Mem <= 732 & Yr08 == 1

replace MemIQR2 = 1 if Mem > 270.37 &
Mem <= 708.52 & Yr09 == 1

replace MemIQR2 = 1 if Mem > 270.37 &
Mem <= 619.86 & Yr10 == 1

replace MemIQR2 = 1 if Mem > 251.56 &
Mem <= 593.01 & Yr11 == 1

replace MemIQR2 = 1 if Mem > 251.56 &
Mem <= 590.61 & Yr12 == 1

replace MemIQR2 = 1 if Mem > 253.44 &
Mem <= 590.41 & Yr13 == 1

replace MemIQR2 = . if MemIQR2 == 0



gen IntMemIQR2 = 1 if MemIQR2 == 1

replace IntMemIQR2 = . if MemIQR2 == 0



gen BMemIQR2 = MemIQR2

replace BMemIQR2 = 0 if
missing(BMemIQR2)



gen LMemIQR2 = log(MemIQR2)



*MemIQR3

gen MemIQR3 = 0

replace MemIQR3 = 1 if Mem > 750.97 &
Mem <= 1704.23 & Yr07 == 1

replace MemIQR3 = 1 if Mem > 732 &
Mem <= 2011.2 & Yr08 == 1

replace MemIQR3 = 1 if Mem > 708.52 &
Mem <= 2039.4 & Yr09 == 1

replace MemIQR3 = 1 if Mem > 619.86 &
Mem <= 1727.82 & Yr10 == 1

replace MemIQR3 = 1 if Mem > 593.01 &
Mem <= 1588.01 & Yr11 == 1

replace MemIQR3 = 1 if Mem > 590.61 &
Mem <= 1516.98 & Yr12 == 1

replace MemIQR3 = 1 if Mem > 590.41 &
Mem <= 1597.14 & Yr13 == 1

replace MemIQR3 = . if MemIQR3 == 0



gen IntMemIQR3 = 1 if MemIQR3 == 1

replace IntMemIQR3 = . if MemIQR3 == 0



gen BMemIQR3 = MemIQR3

replace BMemIQR3 = 0 if
missing(BMemIQR3)



gen LMemIQR3 = log(MemIQR3)



*MemIQR4

gen MemIQR4 = 0

replace MemIQR4 = 1 if Mem > 1704.23 &
Yr07 == 1

replace MemIQR4 = 1 if Mem > 2011.2 &
Yr08 == 1

replace MemIQR4 = 1 if Mem > 2039.4 &
Yr09 == 1

replace MemIQR4 = 1 if Mem > 1727.82 &
Yr10 == 1

replace MemIQR4 = 1 if Mem > 1588.01 &
Yr11 == 1

replace MemIQR4 = 1 if Mem > 1516.98 &
Yr12 == 1

replace MemIQR4 = 1 if Mem > 1597.14 &
Yr13 == 1

replace MemIQR4 = . if MemIQR4 == 0



gen IntMemIQR4 = 1 if MemIQR4 == 1

replace IntMemIQR4 = . if MemIQR4 == 0



gen BMemIQR4 = MemIQR4

replace BMemIQR4 = 0 if
missing(BMemIQR4)



gen LMemIQR4 = log(MemIQR4)



*Membership of Dists in IQR1

gen MemIQR107 = Mem if MemIQR1 == 1
& Yr07 == 1

gen MemIQR108 = Mem if MemIQR1 == 1
& Yr08 == 1

gen MemIQR109 = Mem if MemIQR1 == 1
& Yr09 == 1

gen MemIQR110 = Mem if MemIQR1 == 1
& Yr10 == 1

gen MemIQR111 = Mem if MemIQR1 == 1
& Yr11 == 1

gen MemIQR112 = Mem if MemIQR1 == 1
& Yr12 == 1

gen MemIQR113 = Mem if MemIQR1 == 1
& Yr13 == 1
```

```
*Membership of Dists in IQR2

gen MemIQR207 = Mem if MemIQR2 == 1 & Yr07 == 1

gen MemIQR208 = Mem if MemIQR2 == 1 & Yr08 == 1

gen MemIQR209 = Mem if MemIQR2 == 1 & Yr09 == 1

gen MemIQR210 = Mem if MemIQR2 == 1 & Yr10 == 1

gen MemIQR211 = Mem if MemIQR2 == 1 & Yr11 == 1

gen MemIQR212 = Mem if MemIQR2 == 1 & Yr12 == 1

gen MemIQR213 = Mem if MemIQR2 == 1 & Yr13 == 1


*Membership of Dists in IQR3

gen MemIQR307 = Mem if MemIQR3 == 1 & Yr07 == 1

gen MemIQR308 = Mem if MemIQR3 == 1 & Yr08 == 1

gen MemIQR309 = Mem if MemIQR3 == 1 & Yr09 == 1

gen MemIQR310 = Mem if MemIQR3 == 1 & Yr10 == 1

gen MemIQR311 = Mem if MemIQR3 == 1 & Yr11 == 1

gen MemIQR312 = Mem if MemIQR3 == 1 & Yr12 == 1

gen MemIQR313 = Mem if MemIQR3 == 1 & Yr13 == 1


*Membership of Dists in IQR4

gen MemIQR407 = Mem if MemIQR4 == 1 & Yr07 == 1

gen MemIQR408 = Mem if MemIQR4 == 1 & Yr08 == 1

gen MemIQR409 = Mem if MemIQR4 == 1 & Yr09 == 1

gen MemIQR410 = Mem if MemIQR4 == 1 & Yr10 == 1

gen MemIQR411 = Mem if MemIQR4 == 1 & Yr11 == 1

gen MemIQR412 = Mem if MemIQR4 == 1 & Yr12 == 1

gen MemIQR413 = Mem if MemIQR4 == 1 & Yr13 == 1


*LInfState by IQR

gen IQR1LInfStatePPE = LInfState if MemIQR1 == 1

gen IQR2LInfStatePPE = LInfState if MemIQR2 == 1

gen IQR3LInfStatePPE = LInfState if MemIQR3 == 1

gen IQR4LInfStatePPE = LInfState if MemIQR4 == 1


*Sums

gen NonLclExp = Fed + State

gen InfNonLclExp = InfFed + InfState


*PPE

replace TotPPE = AnnExp / Mem

gen LTotPPE = log(TotPPE)

gen InfTotPPE = InfExp / Mem

gen LInfTotPPE = log(InfTotPPE)

gen TaxPPE = Tax / Mem

gen InfTaxPPE = InfTax /Mem

gen LInfTaxPPE = log(InfTaxPPE)

gen OthLclPPE = OthLcl / Mem

gen InfOthLclPPE = InfOthLcl / Mem

gen LInfOthLclPPE = log(InfOthLclPPE)

gen StatePPE = State / Mem

gen InfStatePPE = InfState / Mem

gen LInfStatePPE = log(InfStatePPE)

gen FedPPE = Fed / Mem

gen InfFedPPE = InfFed / Mem

gen LInfFedPPE = log(InfFedPPE)

gen OtherPPE = Oth / Mem

gen InfOthPPE = InfOth / Mem

gen LInfOthPPE = log(InfOthPPE)

gen NonLclPPE = NonLclExp / Mem

gen InfNonLclPPE = InfNonLclExp / Mem

gen LInfNonLclPPE = log(InfNonLclPPE)


*Spending

gen Lcl = AnnExp - (State + Fed)

gen InfLcl = Lcl * InfInd13

gen InfLclPPE = InfLcl / Mem

gen LInfLclPPE = log(InfLclPPE)


*Proporional Poverty

/*By proportion of estimated population in poverty (5-17 years old) in a district

or by membership, corr of membership and pop517 very high, similar*/

gen Pov517Prop = Pov / Mem


*Proportional FRL

gen FRLProp = FRL / Mem


*4Day

bys Dist Yr: gen DCount = 1 if Day == 1 & _n == 1

/*sorts data by district number then by year and assigns the value 1 to DCount

if it's the first observation (_n == 1) of a district in a year having a 4day

(Day == 1) schedule.*/

bys Yr: sum DCount

bys Dist Yr: gen RegCount = 1 if Day == 0 & _n == 1


*Membership by 4Day

gen MemDay = DCount * Mem

bys Yr: sum MemDay

*Membership by 4day and InfStatePPE

gen MemDayPPE = DCount * InfStatePPE

bys Yr: sum MemDayPPE


bys Dist Yr: gen DCount07 = 1 if Day == 1 & Yr07 == 1 & _n == 1

bys Dist Yr: gen DCount08 = 1 if Day == 1 & Yr08 == 1 & _n == 1

bys Dist Yr: gen DCount09 = 1 if Day == 1 & Yr09 == 1 & _n == 1

bys Dist Yr: gen DCount10 = 1 if Day == 1 & Yr10 == 1 & _n == 1

bys Dist Yr: gen DCount11 = 1 if Day == 1 & Yr11 == 1 & _n == 1

bys Dist Yr: gen DCount12 = 1 if Day == 1 & Yr12 == 1 & _n == 1

bys Dist Yr: gen DCount13 = 1 if Day == 1 & Yr13 == 1 & _n == 1


*Mem 4Day by IQR

gen MemDayIQR1 = Mem if DCount == 1 & MemIQR1 == 1

/*gen Day13IQR1 = Mem * DCount13 if MemIQR1 == 1*/

bys Yr: sum MemDayIQR1

gen MemDayIQR2 = Mem if DCount == 1 & MemIQR2 == 1

gen MemDayIQR3 = Mem if DCount == 1 & MemIQR3 == 1

gen MemDayIQR4 = Mem if DCount == 1 & MemIQR4 == 1


*Membership by 4day and IQR1

bys Yr: sum MemDayIQR1

bys Yr: sum MemDayIQR1 if missing(DCount)

*Membership by 4day and IQR2

bys Yr: sum MemDayIQR2

bys Yr: sum MemDayIQR2 if missing(DCount)

*Membership by 4day and IQR3

bys Yr: sum MemDayIQR3

bys Yr: sum MemDayIQR3 if missing(DCount)


*AllPass

gen Pass = alla + allp

*AllFail

gen Fail = allb + allbb


*Ratios

gen FedRatioExp = Fed / AnnExp

gen InfFedRatioExp = InfFed / InfExp

gen StateRatioExp = State / AnnExp

gen InfStateRatioExp = InfState / InfExp

gen NonLclRatioExp = NonLclExp / AnnExp

gen InfNonLclRatioExp = InfNonLclExp / InfExp


*Other Variables

*Interacted Binaries

gen InfState2007 = InfStatePP*Yr07

gen InfState2008 = InfStatePP*Yr08

gen InfState2009 = InfStatePP*Yr09

gen InfState2010 = InfStatePP*Yr10

gen InfState2011 = InfStatePP*Yr11

gen InfState2012 = InfStatePP*Yr12

gen InfState2013 = InfStatePP*Yr13

gen Interact = Pov * FRL
```