MACS 30200 - Method and Initial Result

Ellen Hsieh

Method

Data

To measure how parent and family involvements, especially time spent with child for cultural activities, influence child's behavior school, the dependent and independent variables in this study are based on data from Parent and Family Involvement (PFI) in Education Survey of the National Household Education Surveys Program of 2016. This PFI survey file contains data regarding to 14,075 children age 20 or younger in kindergarten through 12th grade, including 13,523 students enrolled in public or private school and 552 homeschooled students. The data includes the characteristics of child, parent/guardian and household and other data including school choice, parent and family involvement at school, the child's behavior at school, grade retention, parents' satisfaction with the child's school, family's involvement in school work and activities outside of school, factors affecting family involvement. The survey is completed by the parents or knowledgable adults.

As for this study, we would only focus on the students enrolled in public or private school. Therefore, we conducted data preprocessing to obtain the specific data we need for this study. First, we excluded 552 homeschooled students from the dataset. Then, we kept only the students who has two guardians in the survey. In order to build a appropriate feature selection model, we first selected 84 predictors from 823 variables according to the literature reviews, which are related to child basic information such as race, grade, health condition, native language, school behavior, and parents and family characteristics including socioeconomic

status, race, marital status, family structure, and parental involvement at school and outside school. All binary variables are converted to 0 and 1. In addition, the missing values are imputed using the mean of given variable. The dependent variable is a combination of time contacted by teachers about good behavior and behavior problems, which is the proportion of time contacted about good behavior over the total time contacted by teachers about child's behavior at school. Besides, the interest independent variable in this study, time spent with child for cultural activities, is the accumulation of seven different related variables in the dataset including visited a library in the past month, visited a bookstore in the past month, gone to a play in the past month, visited art gallery in the past month, visited a zoo in the past month, attended a religious event in the past month, attended a sporting event in the past month.

Procedure

The purpose of this study is gain a deeper insight on how spending time with child for cultural activities would influence child's behavior at school. However, the number of remaining independent variables that would affect child's behavior at school is still large after the selection according to previous literature reviews. Therefore, in order to conduct a more effective analysis, we performed a feature selection to obtain ten most important independent variables without including the independent variable that we are interested in, times spent with child for cultural activities. For features selection, we chose the random forest regressor model because tree-based models are more appropriate when confronting mixed types of data. To better tune our random forest regressor model, we first split the data into train and test set with test set size = 0.3. Then, we use randomized search on hyper parameter with 5-fold cross validation to select the best

random forest regressor model with lowest mean square error. Finally, we fit the best model with the whole dataset to increase the opportunity to select important features more accurately.

After obtaining the ten most importance features, we subsequently analyzed the selected variables and our independent variable of interest with dependent variable using ordinary least squares regression. From the regression, we found out the significant variables for our OLS model and studied how our interest variable affect our dependent variable, child's school behavior. However, this regression analysis includes all grades of students from kindergarten, which including full-time and part-time, to twelfth grade to generate a more generalized model, the result might be influenced by the child grade. As a result, we also looked for how grade influence child behavior and its relationship with time spent for cultural activities for child.

Result

From feature selection using random forest regressor, we obtained the top ten important features and their importance regarding to the dependent variable: grade attending (0.059894), first parent/guardian age (0.058115), second parent/guardian age (0.057057), times participated in school meetings (0.044155), expectations for child future education (0.041366), total income (0.035133), first parent/guardian hours worked per week (0.035067), second parent/guardian highest grade level completed (0.034634), second parent/guardian hours worked per week (0.034231), and days eaten the evening meal together in the past week (0.033995). The statistics summary of those ten variables are shown in table 1. In table 1, it also includes statistic summary for the dependent variable and the variable of interest.

Table 1. Statistic Summary for variables selected

Variables	Mean (Standard Deviation)
Dependent Variable	
Time contact about good behavior (%)	0.607 (0.291)
Independent Variable	
Grade attending (%)	
Full-time kindergarten	0.006
Part-time kindergarten	0.054
First grade	0.059
Second grade	0.063
Third grade	0.068
Fourth grade	0.068
Fifth grade	0.070
Sixth grade	0.073
Seventh grade	0.079
Eighth grade	0.081
Ninth grade	0.083
Tenth grade	0.094
Eleventh grade	0.097
Twelfth grade	0.101
Second parent/guardian age	45.70 (8.5)
First parent/guardian age	44.95 (8.25)
Times participated in school meetings	8.45 (9.45)
Expectations for child future education (%)	
Complete less than a high school diploma	0.006
Graduate from high school	0.05
Attend a vocational or technical school	0.056

Attend two or more years of college	0.122
Earn a Bachelor’s degree	0.32
Earn a graduate degree or professional degree	0.44
Total income (%)	
$0 to $10,000	0.021
$10,001 to $20,000	0.04
$20,001 to $30,000	0.059
$30,001 to $40,000	0.063
$40,001 to $50,000	0.064
$50,001 to $60,000	0.065
$60,001 to $75,000	0.09
$75,001 to $100,000	0.165
$100,001 to $150,000	0.205
$150,001 or more	0.228
First parent/guardian hours worked per week	31.45 (19.69)
Second parent/guardian highest grade level completed (%)	
8th grade or less	0.044
High school, but no diploma	0.06
High school diploma or equivalent	0.176
Vocational diploma after high school	0.049
Some college, but no degree	0.155
Associate's degree (AA, AS)	0.088
Bachelor's degree (BA, BS)	0.228
Some graduate/professional education	0.029
Master's degree (MA, MS)	0.118
Doctorate degree (PhD, EdD)	0.019
Professional degree beyond Bachelor’s	0.036

Second parent/guardian hours worked per week	35.34 (18.87)
Days eaten the evening meal together in the past week	4.8 (2.03)
Times spent with child for cultural activities	2.48 (1.69)
N	10003

After selecting ten most important independent variables in the dataset, we constructed an ordinary least square regression model to study the relation between the independent variables and dependent variable. The variable coefficient and standard errors of the variables are shown in Table 2. The most significant variables are grade attending (β = -0.0076, $p < 0.001$), time participated in school meetings (β = 0.0014, $p < 0.001$), expectation for child future education (β = 0.0322, $p < 0.001$), and times spent with child for cultural activities (β = 0.0157, $p < 0.001$). As shown in Figure 1, our variable of interest, time spent with child for cultural activities, has positive significant effect on child's school behavior.

Table 2. Regression Result for Child Behavior at School

Grade attending	-0.0076 ***
	(0.001)
Second parent/guardian age	-0.0003
	(0.000)
First parent/guardian age	0.0004
	(0.000)
Times participated in school meetings	0.0014***
	(0.000)
Expectations for child future education	0.0322***
	(0.003)
Total income	-0.003*
	(0.001)

First parent/guardian hours worked per week	0.0003 (0.000)
Second parent/guardian highest grade level completed	-0.0011 (0.001)
Second parent/guardian hours worked per week	0.000082 (0.000)
Days eaten the evening meal together in the past week	0.0044** (0.001)
Times spent with child for cultural activities	0.0157*** (0.002)

Standard errors are reported in parentheses.

*, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

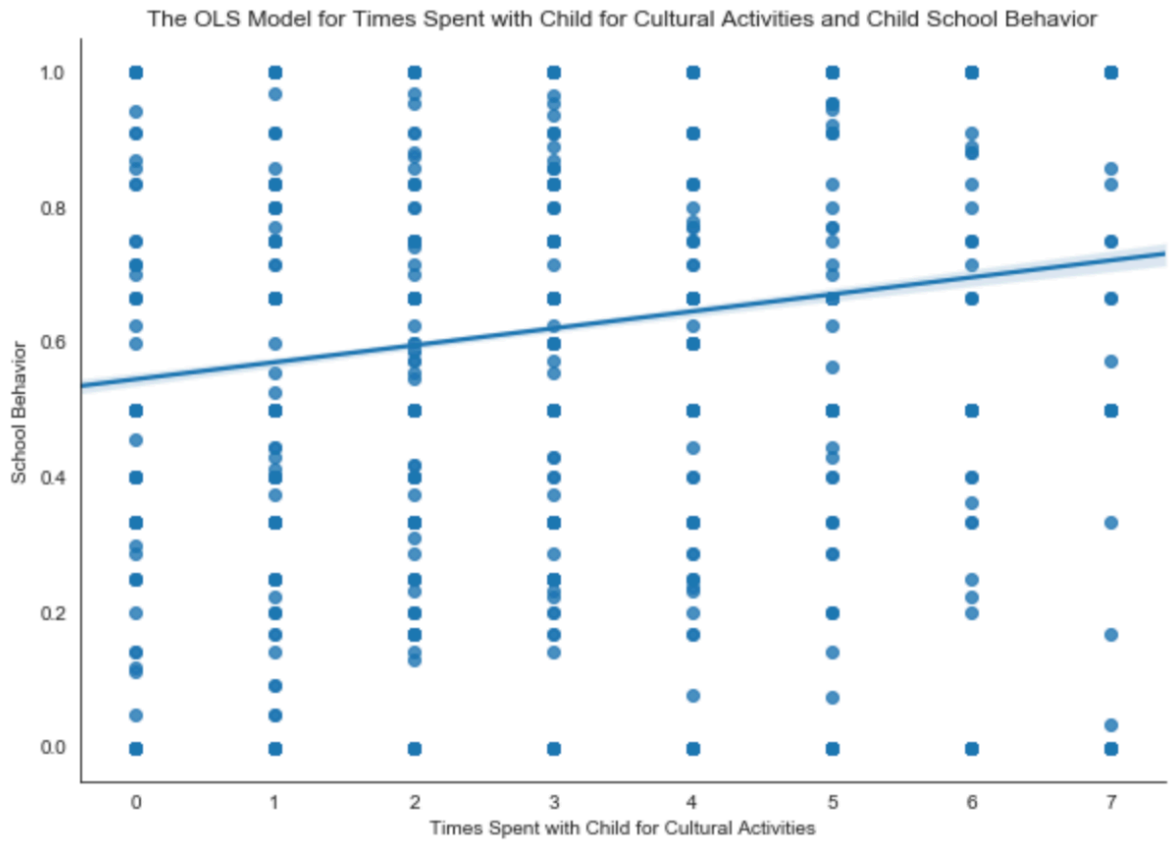

Figure 1. The OLS Model for Time Spent with Child for Cultural Activities and Child's School Behavior

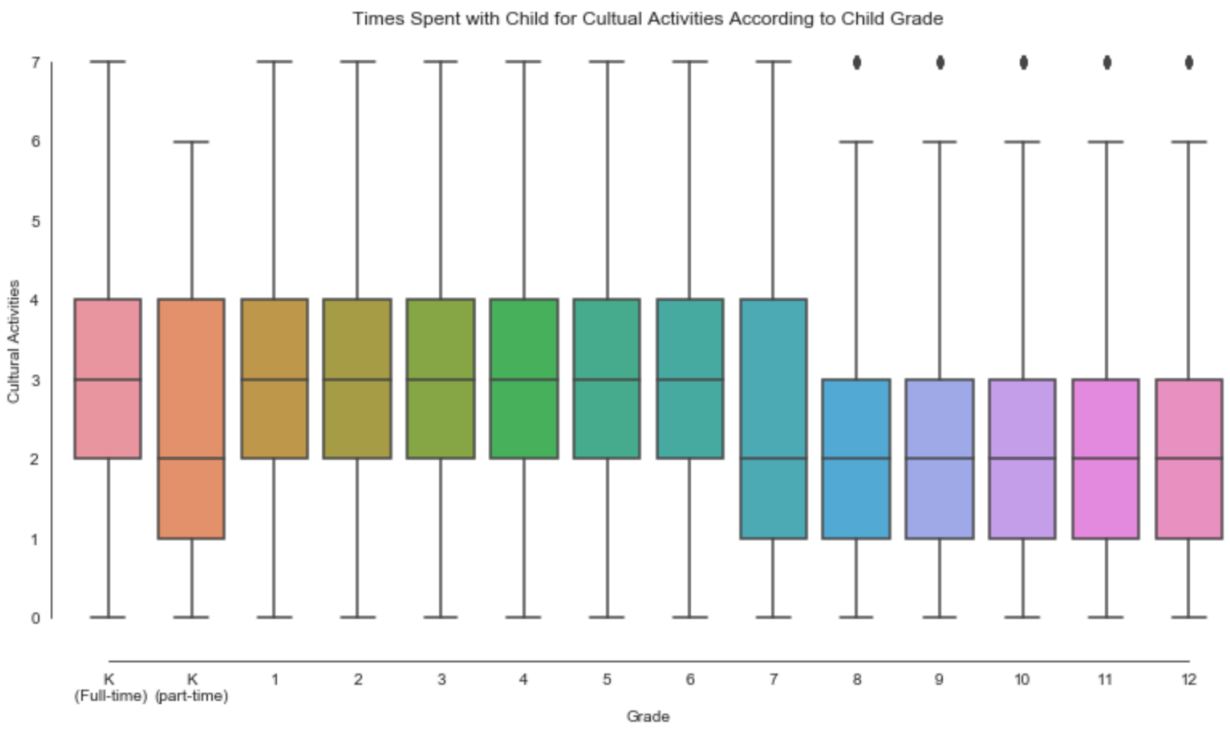

Figure 2. The Relation between Time with Child for Cultural Activities and Child Grade

Apart from analyzing the data using OLS method, we also plotted the relation between independent variable of interest, time spent with child for cultural activities, and child behavior performance at school as shown in figure 2. From the graph, we can see that the number of time spent with child in the past month from kindergarten to sixth grade are similar except for part-time kindergarten. Then, the mean of the number decline after seventh grade and the both bounds for the range of number also tend to become lower.

As we mentioned before, we included all grades of students in the model. In order to take consideration for the effect of grade on our dependent variable, child behavior performance at school. Therefore, we plotted the linear model fit for our dependent variable using child grade as figure 2. It seems that as grade become higher, the behavior performance start to decline.

However, from data distribution we can know that as grade become higher, time contacted about behavior problems or good behavior become less as well.

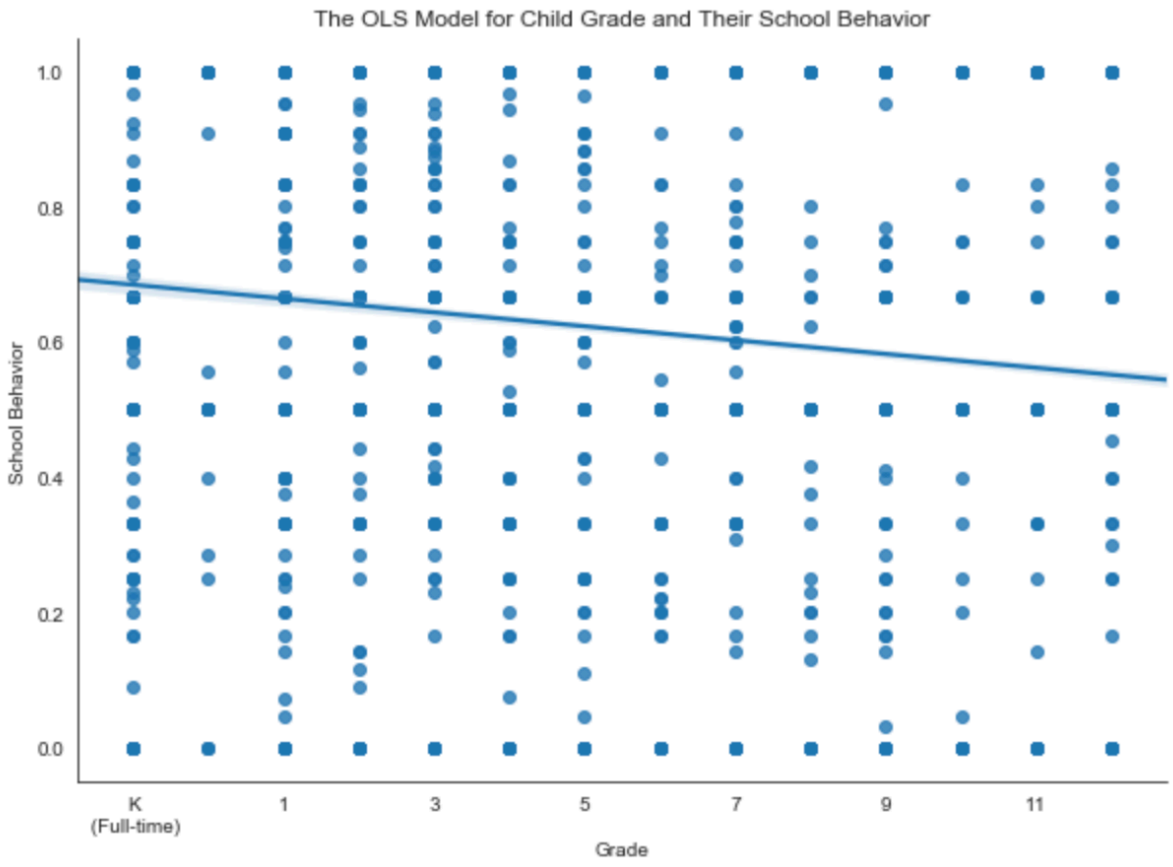

Figure 2. The OLS Model for Child Grade and Child Behavior Performance

Reference

National Center of Education Statistics. (2018). 2016 Early Childhood Program Participation Survey, 2016 Parent and Family Involvement in Education Survey [Data file and code book]. Available from https://nces.ed.gov/nhes/data_files.asp