Peer Effects & Differential Attrition: Evidence from Tennessee's Project STAR

Sanjay Satish

03/09/2022

Professor Robert Garlick, Faculty Advisor

Professor Michelle Connolly, Faculty Advisor

Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University

Durham, North Carolina

2022

Abstract

This paper explores the effects of attrition on student development in early education. Utilizing data from Tennessee's Project STAR experiment, this paper aims to expand upon the literature of peer effects as well as competition between public and private schools. It first reproduces the attrition profile of Project STAR first outlined in Rohlfs & Zilora's (2014) to determine the nature of attrition in the experiment. It departs from other papers in the literature by utilizing survival analysis to determine which characteristics of students prolonged participation in the experiment. This paper then uses these findings to estimate the peer effects of attrition on students who remained in the experiment through the change in test scores of students between the 1st and 2nd grades. Such peer effects are subsequently decomposed using clustering to estimate the peer effects from students who may have left to private schools.

This paper aims to provide evidence that student sorting across different types of schools has educational impacts on the students they leave behind. At the same time, it analyzes attrition in Randomized Controlled Trials and may provide evidence of subsequent confounding factors or spillover effects on the broader population related to such developments.

JEL Classification: I, I21, I26, H4, J13.

Keywords: Attrition, Economics of Education, Peer Effects, Private Schools, Project STAR.

1 Introduction

Imagine yourself as an elementary school teacher trying to manage an unruly group of students. Do you seat them across the classroom from each other, hoping that once separated they stop their disruptive behavior? Or does this choice mean that they'll simply find someone else to talk to, resulting in further disruption? Answering these questions may prove instrumental for the success of your students. There exists extensive empirical literature examining the effect of exogenous policy shifts on student achievement. A considerable portion on the current economics of education literature, however, is now centered around the estimation of peer effects: the unobserved, peer-to-peer spillovers from students both inside and outside the classroom. Studying these mechanisms through which student outcomes may be maximized or diminished has significant consequences for policymakers.

It is well documented in the literature that student-to-student peer effects do in fact exist and that such effects are, in part, influenced significantly by peer quality. It is also well known that both private schools provide a channel through which wealthier, high-ability students are able to leave the public school system. Many private schools have entrance exam requirements which deliberately select for high-achieving students. Given the existence of such rationing, it is perhaps the case that the act of exiting a school may have unobserved effects on peers in the classroom. That is, if the distribution of students that exit public schools for private schools consists primarily of high-achieving students, perhaps there exist negative or reduced positive peer effects from such withdrawal. My paper aims to analyze the effects of this attrition, especially in relation to the varying degree of these effects based on student ability.

Within the context of United States public education, private schools create both exclusionary and rivalrous conditions. In particular, Private schools create rivalrous conditions by introducing rationing through two distinct channels: tuition rationing and cream-skimming. Tuition rationing is the practice through which private schools select for students with high-income backgrounds by charging for seats – disallowing lower-income families from con-

This paragraph, particularly the last sentence, should be a bit more precise about the relationship between high ir

sumption without some form of subsidization. Cream-skimming is the practice through which private schools select for high-ability students by introducing entrance exam cutoffs as a requirement for admission. In other words, the highest performing students in public schools (i.e. the "cream) is "skimmed" into private schools. Although, in some cases, private schools may "cross-subsidize" (i.e. high-income, low-ability students' tuition is used to provide scholarships for low-income high-ability students) it remains that the plurality of students that enter private-schools are likely to be of high-ability.

Utilizing evidence from Tennessee's Student/Teacher Achievement Ratio Project (Project STAR) – an experiment in which elementary school students and teachers were randomly assigned to different class sizes – my paper aims to estimate the peer effects of student attrition on academic achievement between the 1st and the 3rd grade. Given that students were randomized into class types, it is likely that measures of peer effects from Project STAR are not biased due to sorting. Additionally, as part of the experiment, a robust set of student, teacher, and school-level characteristics were collected and students were tracked individually throughout the experiment. My analysis draws upon subsequent data on graduation and dropout rates of the students involved in STAR as well as the identification of students that left the public school system in order to classify students who potentially left the experiment for private schools.

First, I examine attrition in Project STAR writ large. Though the identification of the proportion of students who either moved across classrooms or schools within the experiment and those that left the experiment entirely, I paint a picture of differential attrition that occurred.

Careful about language. Rather use "association between attrition and student ability" and a

Second, using this identification, I introduce a model of attrition which measures the effect of student ability, controlling for confounds, on the probability a particular student stayed in the experiment. I compare the results from this model on the schools in Project STAR with attrition data from schools in the control group and compare both. This model motivates the empirical strategy behind peer effects estimation in the paper.

I found this a bit confusing. Why on the changes in lagged rest scores, rather than subsequent test scores?

Third, I consider how attrition across all students who left Project STAR may have impacted students by considering the effects of attrition on changes in their lagged test scores. I estimate these peer effects using a regression analysis to explore both changes in the magnitude of the effect based on class size.

Finally, I perform an empirical analysis aiming to isolate the peer effects from students who left the public school system entirely. I use information on public school testing data both during and after Project STAR to classify students who potentially left the public school system. In order to identify students who potentially left the experiment for private schools, I fit a clustering model aimed at sorting such students into "similar" groups. Using these clusters, I undertake a sensitivity analysis to measure the robustness of my measurements to certain assumptions on attrition. I also undertake a short discussion of the limitations of measuring unobservables and how inconsistencies in randomization within Project STAR may bias my results.

2 Literature Review

Since the publication of the Coleman Report in 1966[1] there has been suspicion that student unobservables have a significant impact on education production.[2] The classic education production function predicts that test scores can be represented as a function of class size and confounds[3]. Subsequent research has shown that such a relationship is likely identifying only a portion of what affects student achievement[4], especially when viewed in conjunction with the documented relationship between peer effects and achievement. Some, as first described by Lazear (2001), theorize that disruption in the classroom presents a tradeoff between learning and class size – a higher proportion of "better behaved students" leads

[1]A study commissioned by the US government and mandated by the Civil Rights Act of 1964 aimed at measuring educational opportunity in the United States.

[2]Dickinson, E. E. (2016).

[3]Lazear, E. P. (2001), p.777; Examples include: student's own ability, teacher experience, parental income, percentage of students in school on free and reduced lunch, demographic characteristics, etc.

[4]Hanushek (2008).

to a larger optimal class size.[5] That is, disruptive students may have a greater negative impact on smaller class-sizes than larger ones. Such negative spillovers from disruption can perhaps be best understood through the lens of peer effects. The channel through which one's own production results in spillovers or externalities on peers achievement is measured by peer effects. Subsequent literature has broadened this model of production to focus both on exogenous inputs to production (e.g. class size, curriculum, etc.) as well as inputs that cannot be controlled by policymakers (e.g. parental income, student's attentiveness, peer effects, etc.).[6]

Why call these inputs "exogeneous"?

Although there exists a vast literature on peer effects, papers analyzing the peer effects of attrition are somewhat limited, especially given the lack of longitudinal data of student-tracking across both the private and public school systems. One notable study is that of Dills (2005) which utilized the entry of a magnet school in Virginia to estimate the effect of the exit of high-ability students on the proportion of students scoring in the top quartile on a nationally-standardized exam. Dills finds that the loss of high-achieving students to the magnet school lowers the achievement of their lower-scoring counterparts who remained in public schools.[7] Key limitations of Dills' paper, though, include an inability to track individual students' test scores and a lack of demographic controls. Randomized experiments offer the best data to track peer effects given the elimination of selection bias due to peer sorting.

Within the context of Project STAR, papers such as Boozer (2001) focus on estimating the difference in peer-effects due to variation in class-size. Researchers have found that, especially for small classes, much of the effect of class-size on achievement is actually captured by the increased intensity of spillovers from peers in these settings.[8] Given missing data concerns in Project STAR, Sojurner (2011) develops a model for estimating peer effects with the caveat of missing peer data. Sojourner utilizes pre-assignment measures of student

[5] Lazear, E. P. (2001), p.779.
[6] Hanushek (2008).
[7] Dills (2005).
[8] Boozer (2001).

ability in conjunction with random assignment to estimate effects, finding that, on average, the effects of bring around high-achieving peers are positive and that peer-ability perhaps matters more for low-achieving students than for high-achieving students.[9] Key limitations of Project STAR include differential attrition and a lack of data on where students who left the experiment attended school after leaving. Rohlfs & Zilora (2014) exploit whether or not a student had public school testing records after the conclusion of Project STAR as a proxy for identifying if a student remained in the public school system. Such developments should assist in the estimation of peer effects of attrition and reduce the limitations the data impose.

3 Data

3.1 Project STAR

Project STAR was class-size randomization experiment across grades K-3 in Tennessee between 1985-1989. 11,600 students across 79 schools voluntarily participated in the study. Students and teachers were randomized into one of three class types at the start of the 1985 school year: small classes (13-17 students), regular-sized classes (22-25 students), and regular-sized classes with a teacher's aide (22-25 students in addition to a full-time teacher's aide). All Tennessee school systems were invited to participate in STAR and costs associated with STAR would be funded by the State. A minimum of 57 students in a school was necessary for participation, allowing for one class of each of the three class types. Most schools stayed in the experiment; however, between grades 1-3 four schools withdrew for several reasons including not being able to maintain the randomization of the expiriment.

After being initially assigned to a class type, students were intended to be kept in that class for the duration of the experiment (i.e. 4 years, or between K-3rd grade, although some students were placed across class types). Among participating schools, each was required to

[9] Sojurner (2011).

have at least one of the three class types and randomization occurred within each school. Study administrators collected a variety of data on students from observable characteristics (e.g. race, age, gender, etc.) on both students, teachers, and schools as well as a variety of standardized test scores, administered at the end of each school year. Students who entered a participating school in the middle of the experiment were also randomly assigned to a class type upon entry. There are some limitations to the design, mainly that at the beginning of 1st grade, students in the regular and regular-aide classes were re-randomized across these types, with approximately one half of students in regular classes being reassigned to regular-aide classes. Secondly, approximately 10 percent of students moved across small and regular-sized classes between grades.[10] Thirdly, between grades 1 and 2 in year two of the experiment (1987) 54 second grade teachers were given additional training before assignment to class types. Researchers determined that there was no significant difference between the scores of students with trained versus untrained teachers. Lastly, there exists evidence of differential attrition from the experiment by class type. I exploit this to determine the severity of the peer effects from attrition based on a student's class type. It is important to note that identification of students who left the experiment, as noted by Rohlfs & Zilora, can only be separated into students who either changed class types, switched to another public school, or left the public school system entirely. Thus, I am unable to accurately identify among the students that left the public school system, which left to a private school. Further information about the experiment is well described in Krueger (1999).

3.2 Test Scores

The primary educational attainment measure used to track students was the Stanford Achievement Test or "SAT". These scores are reported as item-response-theory scores, which allow for comparison across grades. Starting in the first grade, Basic Skills First or "BSF" tests were administered. These tests corresponded to the state standards in math

[10]Kruger (1999), p.499.

and reading. The BSF tests standards vary between grades, thus they are unable to be compared across grades for peer effects or attrition estimation. I utilize the SAT scores as my primary measure of student ability due to this limitation.

3.3 Summary Statistics

Table 1 is a summary table of student, teacher, and school level characteristics for students in grade 1 of Project STAR. As shown, most of the schools at this level are rural, with most students staying in project star for a total of 3 school years. Table 2 is a representation of the same characteristics for students who were in the experiment during Kindergarten but left during first grade. Inclusion criteria for Table 2 was determined by whether or not the student was "flagged" for being in project star in Kindergarten and was subsequently flagged for not being included in the experiment during 1st grade. It is unclear as to whether or not missing data may confound these flags (i.e. if a student may be flagged as not being in the experiment if there was no data collected for that student – even if they were enrolled in a Project STAR classroom). Table 3 is a representation of the students who were in Project STAR in kindergarten and stayed in the experiment in 1st grade. Summary statistics in Table 3 can be compared with Table 2 in order to understand the characteristics of students who left the experiment between grades K and 1. At first glance, it seems that the students who left seem to be of lower ability, lower income, have a higher probability of being in special education, and more concentrated in inner cities

	Small Class	Regular Class	Regular Class with Aide
Experimental Characteristics			
Number of Years in STAR	3.2 ± 1.0	2.9 ± 1.1	3.0 ± 1.0
Number of Years in Small Classes	3.0 ± 1.0	0.2 ± 0.5	0.1 ± 0.3
School Urbanicity			
Inner City	19.8%	22.1%	18.5%
Suburban	23.5%	24.9%	21.2%
Rural	47.3%	45.0%	50.1%
Urban	9.4%	8.0%	10.2%
Teacher Characteristics			
Female	97.5%	100.0%	100.0%
White	81.2%	83.8%	81.6%
Years of Experience	12.2 ± 8.7	10.3 ± 8.7	12.7 ± 9.2
Class Size	15.7 ± 1.6	22.7 ± 2.3	23.4 ± 2.4
Receives Free Lunch	47.8%	51.9%	50.3%
Special Education	0.6%	1.5%	1.5%
Math SAT Scaled Score	538.7 ± 44.1	525.3 ± 41.7	529.6 ± 42.9
Reading SAT Scaled Score	530.0 ± 56.6	513.5 ± 53.5	521.3 ± 54.7
Listening SAT Scaled Score	572.7 ± 34.5	563.8 ± 32.4	567.2 ± 33.9
Word Study Skills SAT Scaled Score	523.0 ± 52.6	506.2 ± 54.0	513.7 ± 51.8

Table 1: Summary table for students in grade 1 of project star.
Note: Some variables expressed as Mean ± Standard Deviation. Per the data, Inner-city schools are defined as those with more than half of students on free or reduced lunch. Schools in outlying areas of metropolitan areas are considered suburban. Urban schools are those in towns with population over 2500. All other schools are considered rural. Test score data is measured at the end of each year, on testing dates specified by the state.

I might combine tables 2 and 3, possibly into a wide table on a landscape

	Small Class (N=453)	Regular Class (N=603)	Regular Class with Aide (N=580)
Experimental Characteristics			
Number of Years in STAR	1.1 ± 0.4	1.1 ± 0.5	1.1 ± 0.5
Number of Years in Small Classes	1.0 ± 0.2	0.0 ± 0.2	0.0 ± 0.1
Kindergarten School Urbanicity			
Inner City	28.3%	33.5%	30.7%
Suburban	32.2%	27.4%	30.7%
Rural	30.0%	31.3%	31.6%
Urban	9.5%	7.8%	7.1%
Kindergarten Teacher Characteristics			
Female	100.0%	100.0%	100.0%
White	84.5%	71.5%	82.4%
Years of Experience	8.5 ± 5.7	8.9 ± 5.9	NA ± NA
Class Size	15.2 ± 1.4	22.6 ± 2.1	22.9 ± 2.3
Receives Free Lunch	54.1%	56.7%	57.8%
Special Education	5.7%	3.8%	4.5%
Math SAT Scaled Score	473.5 ± 51.1	467.8 ± 48.5	469.0 ± 46.8
Reading SAT Scaled Score	431.4 ± 33.0	425.3 ± 29.5	427.3 ± 29.7
Listening SAT Scaled Score	531.7 ± 34.7	527.6 ± 35.3	527.5 ± 33.2
Word Study Skills SAT Scaled Score	427.9 ± 35.7	422.0 ± 33.9	425.0 ± 32.8

How is grade 1 class si

Table 2: Summary table for students who were in Project STAR in Kindergarten but left the experiment before entering grade 1.
Note: some variables expressed as Mean ± Standard Deviation. Values correspond to kindergarten-grade information. Does not account for students who may have re-entered the experiment in later years or distinguish between students by attrition category. Variable definitions are consistent with Table 1.

	Small Class (N=1303)	Regular Class (N=1425)	Regular Class with Aide (N=1490)
Experiment Characteristics			
Number of Years in STAR	3.5 ± 0.8	3.5 ± 0.8	3.5 ± 0.8
Number of Years in Small Classes	3.3 ± 1.0	0.3 ± 0.8	0.3 ± 0.7
Kindergarten School Urbanicity			
Inner City	18.5%	18.0%	20.3%
Suburban	21.0%	17.1%	19.7%
Rural	51.0%	57.1%	49.3%
Urban	9.4%	7.8%	10.7%
Kindergarten Teacher Characteristics			
Female	100.0%	100.0%	100.0%
White	86.6%	83.2%	85.4%
Years of Experience	9.0±5.8	9.1±5.7	NA±NA
Class Size	15.1 ± 1.5	22.2 ± 2.2	22.6 ± 2.3
Recieves Free Lunch	44.3%	43.2%	45.8%
Special Education	3.1%	2.8%	2.2%
Math SAT Scaled Score	497.1 ± 47.5	489.8 ± 45.8	488.4 ± 44.1
Reading SAT Scaled Score	443.7 ± 31.6	438.8 ± 30.7	438.7 ± 31.7
Listening SAT Scaled Score	542.9 ± 32.2	541.6 ± 31.6	538.9 ± 32.2
Word Study Skills SAT Scaled Score	441.7 ± 37.3	436.2 ± 35.9	436.3 ± 37.9

Table 3: Summary table for students who were in Project STAR in Kindergarten and continued in the experiment in grade 1.
Note: some variables expressed as Mean ± Standard Deviation. Values correspond to kindergarten-grade information. Does not account for students who may have re-entered the experiment in later years or distinguish between students by attrition category. Variable definitions are consistent with Table 1.

4 Empirical Framework

You've organized sections 4 and 5 as "methods 1, ..., K, then results

4.1 Identifying Attrition

Since the completion of Project STAR, subsequent studies have been commissioned to analyze the long-run impacts of the experiment. As part of these studies, data is collected on the academic achievement of students who participated in Project STAR in grades 4 through 8. At the same time, students within the experiment were tracked. If a student moved within the experiment to another Project STAR school, that student's movement can be identified through the data. Utilizing this data, I am to reproduce Rohlfs & Zilora's identification of student status in the 3rd grade by their initial class assignment. Rohlfs & Zilora identify five types of attrition within Project STAR:

1. Students who were in a Project STAR school at time $t-1$, but who's school left the experiment at time t.

2. Students who changed class type within a school.

3. Students who left to another Project STAR school.

4. Students who left to another public school.

5. Students who left the public school system (could be characterized by death, moving to a private school, moving out of state, or test absence).

Identifying students who fall into attrition types 1-4 is easily seen through the data. Both students and schools are flagged for leaving the experiment. Though there may be some missing data concerns, subsequent identification of students who fall into categories 1 and 3 can be undertaken by examining whether or not there exists Project STAR testing data for these experiments both within the experiment's timeframe (i.e. grades K-3) or in the subsequent studies tracking the students post-experiment. Rohlfs & Zilora identify students who left the public school system as those who left Project STAR in 3rd grade and test scores do not exist for them in 4th grade. I strengthen this identification by checking if such students do not have public school records for both 4th and 5th grade. It is likely the case that if a student left Project STAR for a private school, they would not re-enter the public school system until after the 5th grade. This is due to the fact that private schools in certain geographic regions may only accept students between grades K-5. Thus, if a student left for a private school, it is likely that they do not have test score data for both grades 4 and 5. Another potential point of identification is to see which of these students perhaps re-entered the public school system in grades 6-8 (typically denoted as ”middle school”). Students with access to only K-5 private schools may have perhaps been forced to re-enter the public school system for middle school. By identifying attrition types 1-5, I garner an understanding of the differences in attrition across class types – providing the opportunity to further identify the characteristics of students, namely ability, that are perhaps correlated with attrition.

4.2 Quantifying the Effect of Ability on Attrition

To quantify the effect of student ability on attrition, I turn to survival analysis. Within the context of Project STAR, attrition can be considered a survival observation of interest. That is, the act of a student leaving the experiment entirely (i.e. they fall into attrition categories 1,4, and 5) can be considered a ”failure.” Survival analysis allows for the identification of factors which may be correlated with the amount of time a stayed in the experiment (in other words, their time to failure). Here, I represent $T_{i,g,c,s}$ as the failure time (i.e. the number of years a student stayed in the experiment until they left) for each student i in grade g in classroom c at school s. It may be the case that due to missing data concerns or students leaving the experiment in one year and coming back in another, $T_{i,g,c,s}$ is unobserved (or, perhaps, not ”completely” observed). Such observations are considered to be censored, with censoring time C_i. I model the data under the assumption that $T_i > C_i$ (i.e. the ”survival time” is longer than the ”censoring time”). This assumption is consistent with empirical evidence of certain students leaving and returning to the experiment. I model the survival time through the following accelerated failure time (AFT) model, which is often robust to omitted covariates:

What does this mean?

$$log(T_i) = \beta_0 + \beta_1(TS_{i,g,c,s}) + \beta_2(CT_{i,g,c,s}) + \beta_4(FL_{i,g,c,s)\beta_5(X_{i,g,c,s})+\epsilon_{i,g,c,s}} \quad (1)$$

Is this a single categorical variable or a vector of two binary variables? You should use the latter approach.

Where $TS_{i,g,c,s}$ is the test score for student i in grade g in classroom c in school s, $CT_{i,g,c,s}$ is a categorical variable for the class type student i was enrolled in upon entry to the experiment (e.g. small, regular-sized, or regular sized with teachers aide), $FL_{i,g,c,s}$ is a categorical variable for the free-lunch status of student i in grade g, and $X_{i,g,c,s}$ is a vector of controls for each student including observable student, teacher, and school data. Examples of this include student attendance, teacher's highest education level, student's race, percentage of school on free and reduced lunch, school geography (e.g. urban, rural, etc.), and others. The coefficients of this model have a multiplicative effect on time. That is, if $\beta_1 = 1$, then $exp(\beta_1) \approx 2.718$. Holding all other variables constant, an individual with $TS_{i,g,c,s}$ one unit greater than another is expected to stay in the Project STAR experiment 2.718 times longer than the other. In other words, the probability this individual "survived" to time $2.718t$ is the same as the probability that another individual has "survived" to time t.

The AFT model is a fully parametric model. Thus, I must specify a probability distribution for $log(T_i)$, or a distribution of attrition probabilities throughout the experiment. Based on the nature of the experiment, attrition was more likely in early years (i.e. Kindergarten and 1st grade) than later on in the experiment as schools who were unable to continue randomization dropped out or students who were unhappy with their class assignment left STAR. Thus it is reasonable to model the attrition probabilities in the experiment as monotonically decreasing with time. The Weibull distribution allows for the estimation of monotonically increasing, decreasing, or constant hazards and has relevant analytical advantages in the presence of right-censoring. Given my assumptions regarding the attrition probabilities over time, I represent the hazard function ($\lambda(t)$) of $T_{i,g,c,s}$ using the Weibull distribution:

$$\lambda(t) = p\lambda^p t^{p-1}$$

Here, p is the scale parameter and a value of p less than 1 indicates that attrition rates decrease over time in STAR. An illustration of a Weibull distribution with p less than 1 (monotonically decreasing) is depicted below:

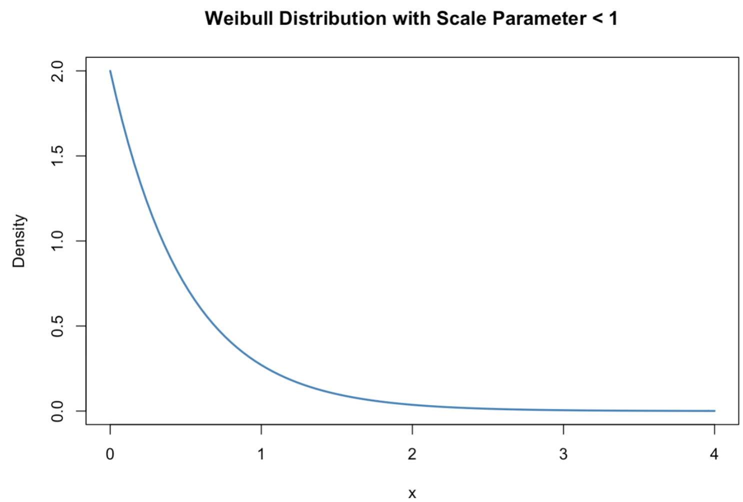


Figure 1: Illustration of monotonically decreasing Weibull Distribution with shape parameter 1 and scale parameter, $p = 0.5$

4.3 Identification of Differential Peer Effects From Attrition Across Class Size

I exploit the panel nature of the Project STAR data to use a fixed effects model to determine the differential peer effect from attrition across all students who left Project STAR based on class size. The dependent variable of interest is a measure of a student's test score, given by the testing data in Project STAR. I make use of lagged covariates to measure the effect of attrition in time $t-1$ on test scores in time t. Key to the estimation of peer effects within the context of education is the assumption that the test scores of student i depend on the background characteristics of student i's peers. For example, student i's learning is likely to be improved if student j's test scores are high. That is, I am primarily concerned with the exogenous peer effects of achievement. I include school fixed effects to account for unobserved differences across schools. There may be school-specific, time-invariant effects which this fixed effect picks up on (e.g. location, funding characteristics, etc.) which should aid in the robustness of my results.

You're not ass

My model is outlined in equation (2):

$$
\begin{aligned}
(2)\qquad TS_{i,g,c,t,s} = \beta_0 &+ \beta_1(L_{i,g,c,t-1,s}) + \beta_2(A_{i,g,c,t-1,s}) + \beta_3(\overline{CA}_{i,g,c,t-1,s}) \\
&+ \beta_4(\overline{PA}_{i,g,c,t-1,s}) + \beta_5(CT_{i,g,c,t-1,s}) + \beta_6(\sigma^L_{i,g,c,t-1,s}) \\
&+ \beta_7(\overline{PA}_{i,g,c,t-1,s} \times L_{i,g,c,t-1,s}) + \beta_8(\sigma^L_{i,g,c,t-1,s} \times L_{i,g,c,t-1,s}) \\
&+ \beta_9(X_{i,g,c,t-1,s}) + \alpha_s + +\epsilon
\end{aligned}
$$

Here, $TS_{i,g,c,t,s}$ is the test score for student i in grade g in classroom c in school s at time t. $L_{i,g,c,t-1,s}$ is the proportion of students in student i's class that left the experiment at time $t-1$. $A_{i,g,c,t-1,s}$ is student i's ability, measured by their test score in time $t-1$. $\overline{CA}_{i,g,c,t-1,s}$ is classmate ability for student i. That is, the average test score of the students in student i's class (not including student i) that don't leave the experiment, measured by their test score in time $t-1$. $\overline{PA}_{i,g,c,t-1,s}$ is the departed peer ability for student i. That is, the average test score of the students in student i's class that left the experiment in $t-1$. This is my main coefficient of interest as I am interested in how the leavers' ability affects students. $CT_{i,g,c,t-1,s}$ is a categorical variable for the class type student i was enrolled in during time $t-1$ (e.g. small, regular-sized, or regular sized with teachers aide). $\sigma^L_{i,g,c,t-1,s}$ is the variance in departed peer ability (i.e. the variance in test scores across all departed peers in student i's class). $\overline{PA}_{i,g,c,t-1,s} \times L_{i,g,c,t-1,s}$ is an interaction term between the departed peer ability and the proportion of students who leave the experiment and $\sigma^L_{i,g,c,t-1,s} \times L_{i,g,c,t-1,s}$ is an interaction between the departed variance in departed peer ability and the proportion of leavers. $X_{i,g,c,t-1,s}$ is a vector of controls for each student including observable student, teacher, and school data. Examples of this include student attendance, teacher's highest education level, student's race, percentage of school on free and reduced lunch, school geography (e.g. urban, rural, etc.), and others. Lastly, α_s is the school fixed effect and ϵ is the regression error term.

4.4 Identification of Attrition Groups

Given that I am unable to directly classify the students who left Project STAR into those who dropped out, left for private schools, or migrated out of state I estimate the potential groupings of departed students through a two part procedure. First, utilizing clustering, I separate students who left the public school system entirely into distinct groups. I subsequently test the sensitivity of these groupings by clustering on different sets of controls, following a procedure similar to that described in Emily Oster's 2016 paper: *Unobservable Selection and Coefficient Stability: Theory and Evidence*. Such methods allow me to see, under the assumptions of my model, how "non-randomly" would the data have to be selected in order to observe the clusters I identify. Different sets of controls may induce different selections for which students may resemble, for example, those likely to leave for private schools, so understanding how sensitive my model is to these assumptions would be beneficial to understand the limitations of this analysis.

4.4.1 Clustering Analysis

To group students who left the public school system into "clusters," I fit a variation of the k-means clustering algorithm. The usefulness of clustering is that it allows me to separate students into groups that "look" alike thus allowing me to perhaps infer or make assumptions about which students went to private school without the data needed to classify these students as private school students. K-means clustering is a technique through which for a given set of observations, these observations are separated into k sets such that some notion of "closeness" within the points in the set is minimized. To explain more rigorously:

Imagine we have a collection of N students who we know left the public school system $x_1, \ldots, x_n$ denoted by the set X. For each of these students there exists a $d \times 1$ vector $\vec{v_i} \in R^D$ where each row of the vector represents the value of each covariate for student x_i (i.e. row one is race, row two is gender, etc.). We denote the set of these vectors as $V = \{\vec{v_1}, \ldots, \vec{v_n}\}$. First, we take k random points in X and generate a set $M = m_1, \ldots, m_k$ of the mean Eculidean distances between all of the points each of the randomly selected initialization points. Then, we assign each x_i to the cluster with the closest mean: that is the cluster with the least squared Eculidean distance.

We end up with k clusters which represent points that are "close" to one another. We then iterate this process multiple times, recalculating the means of the observations assigned to each cluster. We thus end up with k clusters of observations in X which we believe to be "similar" in some way.

4.4.2 Beyond K-Means

Alternatively, can you avoid/minimize this problem by converting the

K-means, although a widely used model, is only applicable for clustering around numeric variables. For ordinal or categorical variables (e.g. free-lunch status) it is hard to discern what the "difference" between two levels of a variable may be. For example, if one observation's Ecludiean distance from another is .5, what does this mean for a variable that takes values of 0 or 1? How is one able to discern what a free-lunch status of 0.5 is? Thus, other techniques are required for mixed numerical and categorical data. I utilize a variation of that outlined by Filaire (2018).

In order to measure distances or, more accurately, dissimilarities between categorical features, I utilize the Gower distance:

$$d(i,j) = \frac{1}{p}\sum_{i=1}^{p} d_{ij}^{(f)} \tag{3}$$

Here, f corresponds to a given variable between two observations i, j and $d(i,j)$ is the dissimilarity between these two observations along that feature. For numerical variables, dissimilarity is observed according to the following formula:

$$d_{i,j}^{(f)} = \frac{\mid x_{if} - x_{jf} \mid}{R_f} \tag{4}$$

This is calculated as the absolute value difference between the two values for the variable f across observations i, j divided by R_f, the maximum range observed in the data between two variables of this given variable.

For categorical variables, differences are either 1 or 0, where 0 indicates two different values and 1 indicates the same value.

4.5 K-Medoids

I recommend explaining this method in a little more detail.

In conjunction with the Gower distance, I utilize the K-Medoids algorithm, a variation of K means. K-Medioids breaks the data into groups around medioids and minimizes, in this case, the Gower distance between observations. Initialization points in K-Medioids, unlike K-Means, are actual data points, rather than random points within the N dimensional space the observations may occupy. This lends for better interpretation regarding the "center" of each cluster. Additionally, K-Medioids when combined with the Gower distance is more robust than K-Means to noise and outliers due to its minimization of dissimilarities rather than Euclidean distances.

4.6 Assessing Optimiality

In order to determine the optimal number of clusters K, I utilize the silhouette coefficient. This value determines how similar an observation is to its own cluster when compared to others. The value ranges from -1 to 1, where higher values indicate more optimal conditions for clustering. To explain more rigorously:

Once the data has been clustered by K-Medoids into K distinct clusters C_k, for any data point $i \in C_k$, we can denote:

$$a(i) = \frac{1}{\mid C_k \mid} \sum_{j \in C_k, i \neq j} d(i,j)$$

as the mean distances between each point i and all other points in the cluster C_k. Here, we subtract 1 from $\mid C_k \mid$ due to the fact that we do not want to account for the distance between i and itself $(d(i,i)$. The smaller $a(i)$ is the closer i is to the points in its cluster. We can thus consider the silhouette value of a data point i to be the following:

$$s(i) = \frac{argmin\{a(i)\} - a(i)}{max\{argmin\{a(i)\}, a(i)\}} \tag{5}$$

where $argmin\{a(i)\}$ represents the smaller man distance between i and all points in clusters for which i is not a member. Thus, a value of $s(i)$ close to 1 indicates that observations are well matched (i.e. the level of dissimilarity $a(i)$ between data points is minimized. We can subsequently plot these silhouette values for different numbers of clusters and determine what the optimal cluster size is based on the maximal interpretive silhouette value.

s(i) forms an N*1 vector. When you plot this, do you avera

5 Results

5.1 Pie Charts

TBD, sorting out issue with authors of Rolhfs and Zilora paper. Results seem to very closely match for the categories I'm able to identify (within 1-2%). I will potentially include these charts in the "Data" Section of the paper, as I believe it flows better given the changes I've made to the results.

I agree.

5.2 Survival Analysis of Attrition

5.2.1 Preliminary Survival Estimates – Kin

I agree with all the interpretation of results in this section, oth

Kaplan Meier estimates of survival probabilities for hort of Project STAR are depicted in the ensuing f dependent variable is "time until attrition" (i.e. Years in star until attrition). Here, students who left are first identified by whether they spent the full four years in STAR (K-3 grade) and subsequently identified by their first exit from the experiment. There are a number of students who leave and come back into the experiment and are re-randomized upon entry. I consider consider these students right censored ($T_i > C_i$). That is, after their first time leaving the experiment, they are considered "left" even though they may have re-entered at a later date. This assumption allows for better parameterization of survival estimates and better represents the monotonically decreasing nature of attrition in STAR. Other papers in the literature make use of "composite duration" and "composite class type" variables which subjectively sort STAR students into different class types and re-calculate years in the experiment. I believe that these measures, although useful for assessing randomization conditions in the experiment, bias estimates of attrition due to the classification of students who re-enter the experiment multiple times as having never been in the experiment. In addition, I remove students who leave STAR due to the fact that their school left the experiment. Thus, Figures 1-4 are illustrations of what I consider "unforced" attrition (i.e. it was the student's or student's guardian's choice to leave STAR)

I agree with th

These initial estimates depict three interesting patterns. First, with regard to attrition based on class type - Figure 4 shows that differential attrition based on class type is likely overstated in the literature. Although there is a statistically significant difference in attrition probabilities across class types at the 5% level, it seems that the effect of being in a regular or regular-aide class does not increase the probability of leaving the experiment drastically compared to being in a small class. It is plausible that this result is due to the fact that students who leave and re-enter the experiment more than once were preferentially re-assigned to smaller classes, thereby providing them an incentive to re-join the experiment. These effects of increased retention in smaller classes are muted in the figure, which displays student's "first" chance to leave the experime

I agree about test scores and free-lunch status. But attrition do

Second, it is clear that lower-income, lower

3) leave the experiment much faster than th

This is shown both by the significant reduction in survival probability for students who are in the bottom 25th percentile of test scores, in Inner City schools, or on free-lunch. This provides evidence that the main channel for attrition in public schools may not be students moving to private schools, but rather drop-outs or significant migration of low-income families. These effects seem independent of class size, potentially leading to the conclusion that such attrition is constant regardless of the random assignment of the student to a class – leading non-experimental forces to be contributing to this attrition.

Third, the attrition probabilities seem to be monotonically decreasing across the course of the experiment. As shown in Figures 1-4, most of the attrition occurs in the first year of the experiment, with less and less students leaving as time goes on. This is consistent with the re-randomization and re-entry that occurred between grades K-1. As well, it seems there is more variation in the students who left in this first year than otherwise. Such results justify the use of a Weibull distribution to parametric the survival estimates, which allows for monotonically decreasing attrition probabilities. This result is elaborated upon in the following section and confirmed through the distributional convergence chart located in the Appendix.

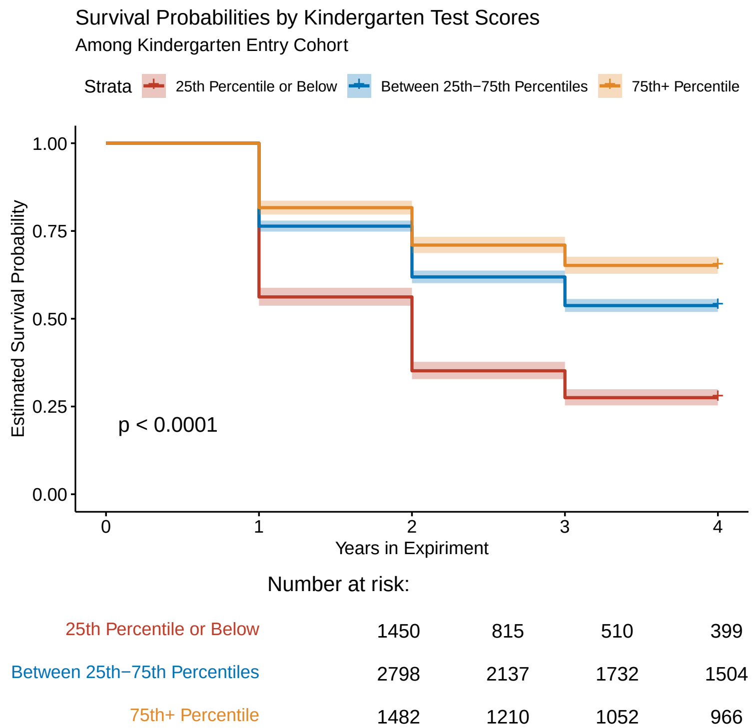


Figure 2: Test scores calculated as the sum of kindergarten math, word skills, listening, and reading SAT scaled scores. Percentiles calculated across all students in the kindergarten entry cohort. Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals.

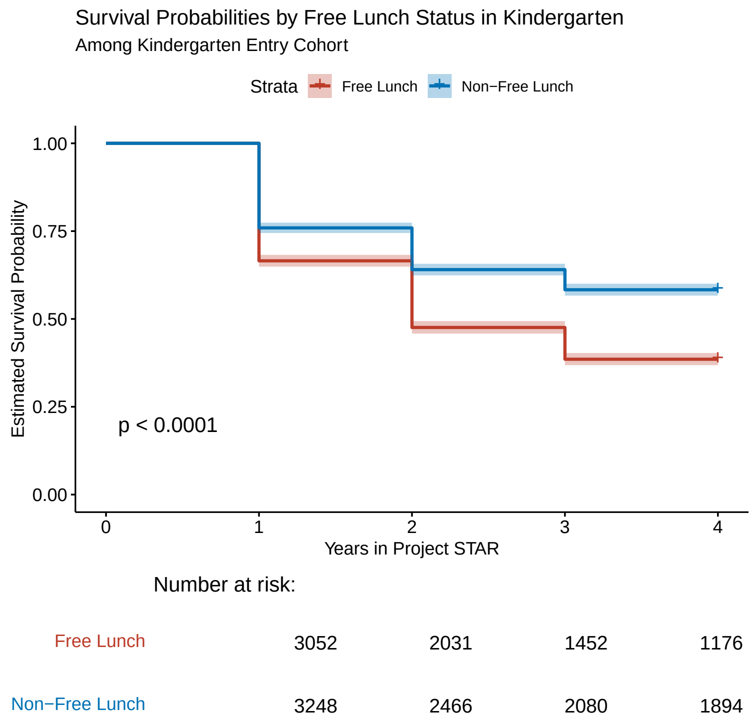


Figure 3: Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals.

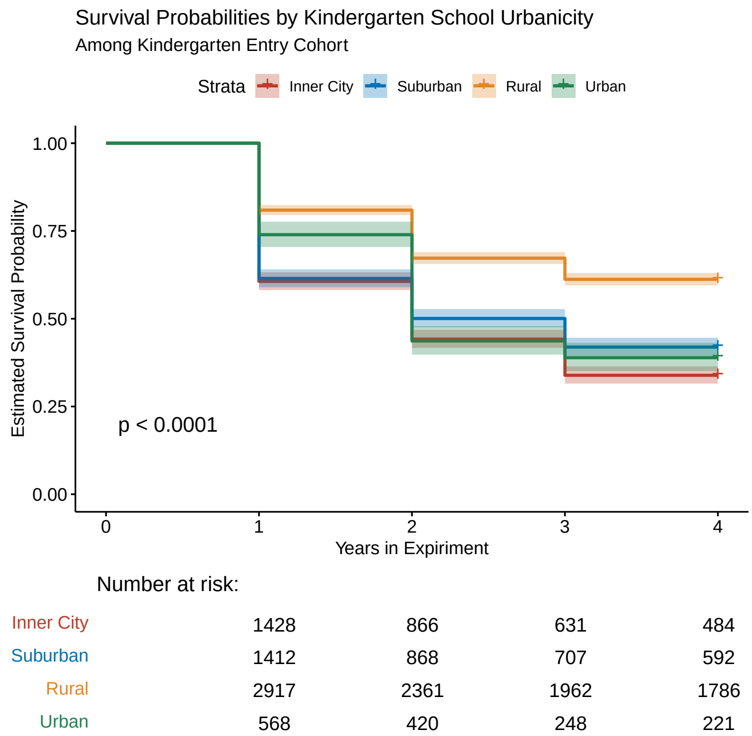


Figure 4: Students may have switched to a different school during their final year in STAR. Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals.

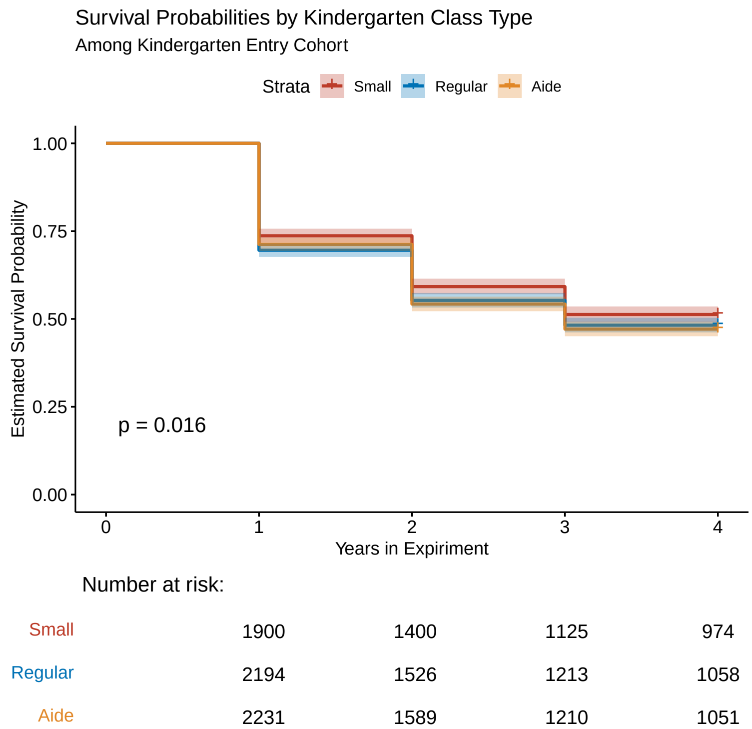


Figure 5: Students may have switched to a different class type during their final year in STAR. Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals.

5.2.2 Survival Analysis of Attrition - Grades K to 1

The below table shows regression results for the AFT model of student attrition in STAR among the Kindergarten Entry Cohort. Specification 1 is fitted on all students in the kindergarten cohort, while specification 2 is fitted on the subset that stayed beyond 1st grade (hence they have the opportunity to switch schools or class types). Specification 1 allows us to see the total effects of attrition across all years in the experiment and specification 2 aims to reduce the effects of attrition that arose from re-randomization or school constraints after the first year of the experiment.

Similar to the survival curves, I decline to fit this model on any students whose schools dropped out of the experiment, once again focusing on "unforced" attrition. Additionally, I drop observations where a student in the kindergarten cohort switched to a different school in 1st grade and came back to the same kindergarten school in 2nd grade. Such students would not be considered "departed" in my experiment; however their school-to-school movement may bias my results. In total, there are 7 students who fall into this category, a very small amount. I estimate that less than .15% of total observations exhibit a similar phenomenon (multiple school switches back to original school) across all years of the experiment.

The results from specification 1 are consistent with those found in the non-parametric estimates in the previous section. Specifically, students who are not on free lunch are expected to survive $e^{0.287} \approx 1.332$ times longer than students who are on free lunch, holding all else constant. Students in rural schools are expected to survive $e^{0..341} \approx 1.406$ times longer than students in inner-city schools and similarly students in urban schools are expected to survive $e^{0.141} \approx 1.151$ times longer than students in inner-city schools, holding all else constant. With regard to ability, I find that students with kindergarten test scores between the 25th and 75th percentiles of all students in the kindergarten cohort are expected to survive $e^{0.506} \approx 1.657$ times longer than those in the 25th percentile or lower, holding all else constant. Similar results are observed for students in the 75th percentile or higher who are expected to survive $e^{0.741} \approx 2.098$ times longer than those in the 25th percentile or lower, holding all else constant.

The results for both income and ability are also observed in specification two, although with lower effect sizes. Across both specifications, these effects are significant at the 1% level. It is also important to note that I do not find significant effects from class size on attrition among this cohort. Diagnostic plots in the appendix show that the assumption of a Weibull distribution to model attrition is well-founded. Attrition appears to monotonically decrease as the experiment goes on and the residuals appear to converge.

Table 4: Coefficient-Level Estimates for AFT Model of Attrition for Kindergarten Entry Cohort

	Dependent variable: Years Until First Exit	
	All Students in K-Cohort	Students Who Stayed Past First Grade
	(1)	(2)
Non-Free Lunch	0.287***	0.249***
	(0.037)	(0.035)
School Suburban	-0.069	-0.066
	(0.079)	(0.077)
School Rural	0.341***	0.103
	(0.082)	(0.080)
School Urban	0.141*	-0.060
	(0.086)	(0.084)
Switched Schools		0.097
		(0.066)
Regular Class	-0.048	-0.016
	(0.036)	(0.038)
Regular Class w/ Aide	-0.030	-0.025
	(0.036)	(0.036)
Switched Class Types		0.024
		(0.031)
Kindergarten Test Score Between 25th-75th Percentile	0.506***	0.307***
	(0.034)	(0.032)
Kindergarten Test Score Above 75th Percentile	0.741***	0.482***
	(0.045)	(0.044)
Not Special Education	-0.022	-0.020
	(0.080)	(0.080)
Not Pulled Out for Special Instruction	0.189***	0.150**
	(0.067)	(0.063)
Constant	0.544***	1.299***
	(0.161)	(0.153)
Controls	Yes	Yes
Parametric Form	Weibull	Weibull
Observations	5,504	3,970
Log Likelihood	-6,816.844	-3,625.752
χ^2	1,163.570*** (df = 26)	456.399*** (df = 27)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Controls include: Student Gender, Student Race, Kindergarten Teacher Experience (yrs.), Days Absent in Kindergarten, Kindergarten School Grade Range,% Students in Kindergarten School Receiving Free Lunch, % Students in Kindergarten School Bused. Student characteristics determined based on kindergarten information unless otherwise specified.

5.2.3 Survival Analysis of Attrition - Grades 1 to 2

The below table shows regression results for the AFT model of student attrition in STAR among all students in 1st grade. This includes new students who entered in the 1st grade cohort as well as those that stayed from kindergarten. Similar to the previous section, specification 1 is fitted on all students in 1st grade, while specification 2 is fitted on those who stayed for at least 1 year. Similar to the previous model, I drop observations where students switch back and forth between schools during the experiment.

Among this cohort, the results are mixed. With regard to ability, the results from specification 1 are consistent with those in the previous section. Students above the 25th percentile are expected to stay in the experiment longer than those below the 25th percentile based on 1st grade test scores. These effects are insignificant among those who stayed past 2nd grade. This is potentially due to the fact that attrition in the final year of the experiment was quite small, with little variation among students that left.

It is important to note that in specification 1, I find that students in inner-city schools are expected to survive longer than those in their suburban, rural, or urban counterparts. This is consistent with the finding that being on free lunch is no longer a significant factor in attrition, with it in specification 2 actually contributing to an increase in survival time. Class size effects are also significant in both specifications, with students in regular-aide classes more likely to survive than their small-class counterparts. It is most likely the case that attrition among the kindergarten students was more pronounced along income lines; however, students with low-ability were less likely to stay in the experiment, regardless of their entry cohort.

Table 5: Coefficient-Level Estimates for AFT Model of Attrition for 1st Grade Entry Cohort

	Dependent variable: Years Until First Exit	
	All Students in 1^{st} Grade	Students Who Stayed Past 2^{nd} Grade
	(1)	(2)
Non-Free Lunch	0.010	-0.004**
	(0.013)	(0.002)
School Suburban	-0.081***	-0.001
	(0.029)	(0.005)
School Rural	-0.067**	-0.003
	(0.031)	(0.005)
School Urban	-0.075**	-0.001
	(0.032)	(0.005)
Switched Schools		0.001
		(0.005)
Regular Class	0.017	0.005**
	(0.013)	(0.002)
Regular Class w/ Aide	0.044***	0.007***
	(0.014)	(0.002)
Switched Class Types		0.004
		(0.003)
1st Grade Test Score Between 25th-75th Percentile	0.187***	0.003
	(0.014)	(0.003)
1st Grade Test Score Above 75th Percentile	0.188***	0.001
	(0.017)	(0.003)
Not Special Education	-0.144	-0.023
	(0.102)	(0.019)
Not Pulled Out for Special Instruction	-0.021	-0.009***
	(0.016)	(0.003)
Constant	1.132***	1.145***
	(0.115)	(0.021)
Controls	Yes	Yes
Parametric Form	Weibull	Weibull
Observations	5,354	4,077
Log Likelihood	-6,478.014	-326.741
χ^2	299.507*** (df = 25)	66.860*** (df = 27)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Controls include: Student Gender, Student Race, 1^{st} Grade Teacher Experience (yrs.), Days Absent in 1^{st} Grade, 1^{st} Grade School Grade Range,% Students in 1^{st} Grade School Receiving Free Lunch, % Students in 1^{st} Grade School Bused. Student characteristics determined based on 1^{st} Grade information unless otherwise specified.

5.3 Peer Effects from Attrition

Let's talk about this in the 18 March meeting. Regressin own first

5.3.1 Grades K-1

Table 6 below depicts the estimates from the peer effects from attrition between grades K-1. The regression is fit over all students that stayed between grades K and 1 and did not switch schools. Specification (2) is that with school fixed effects and is this my preferred specification. As shown, much of a student's test scores are a product of their own test scores as well as the contemporaneous test scores of their peers. As well, students who were not on free lunch had first grade test scores approximately 24.4 points higher than those on free lunch in kindergarten, holding all else constant. This speaks to the relationship between income and ability. Most importantly, we see that for every 1 point increase in the mean test score of leavers in a student's kindergarten class, there is a .15 point drop in a student's first grade test score. This effect is significant at the 5% level. This leads to evidence of a channel of attrition for higher-ability students, or at least the notion that the ability of one's departed peers has an effect on one's test scores in the ensuing year, even after the entrance of new peers into the classroom.

Table 6: Coefficient-Level Estimates for Peer Effects among Kindergarten Entry Cohort

	Dependent variable:			
	Total 1st Grade Test Score		1st Grade Math Score	1st Grade Reading Score
	(1)	(2)	(3)	(4)
Total Kindergarten Test Score	0.874***	0.888***	0.223***	0.274***
	(0.019)	(0.019)	(0.006)	(0.008)
Proportion of Leavers in Kindergarten Class	4.837	2.141	1.330	-0.002
	(4.247)	(4.852)	(1.415)	(1.874)
Mean Test Score of Leavers in Kindergarten Class	-0.093	-0.148**	-0.022	-0.056*
	(0.065)	(0.075)	(0.022)	(0.029)
Variance of Test Score of Leavers in Kindergarten Class	0.0002	0.0003	0.00000	0.0001
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Regular Class Kindergarten	-1.589	-3.635	0.111	-3.562
	(7.926)	(8.399)	(2.450)	(3.244)
Regular-Aide Class Kindergarten	-0.068	-1.687	-0.013	-2.880
	(7.969)	(8.363)	(2.439)	(3.230)
Regular Class 1st Grade	-8.779	-8.501	-1.057	-2.875
	(7.787)	(8.171)	(2.383)	(3.156)
Regular-Aide Class 1st Grade	-7.549	-6.129	-2.403	-0.087
	(7.769)	(8.164)	(2.381)	(3.153)
Peer Average Test Score Kindergarten	-0.544***	-0.417***	-0.104***	-0.137***
	(0.037)	(0.052)	(0.015)	(0.020)
Peer Average Test Score 1st Grade	0.744***	0.638***	0.167***	0.174***
	(0.027)	(0.035)	(0.010)	(0.013)
Non-Free Lunch in 1st Grade	23.353***	24.370***	3.339*	7.921***
	(6.204)	(6.257)	(1.825)	(2.417)
Proportion of Leavers in K : Mean Test Score of Leavers in K	-0.002	-0.001	-0.001	0.0001
	(0.002)	(0.003)	(0.001)	(0.001)
Proportion of Leavers in K : Variance of Test Score of Leavers in K	-0.00002	-0.00002	-0.00000	-0.00001*
	(0.00001)	(0.00001)	(0.00000)	(0.00001)
School Fixed Effects	No	Yes	Yes	Yes
Observations	2,522	2,522	2,522	2,522
R^2	0.693	0.598	0.510	0.493
Adjusted R^2	0.688	0.581	0.489	0.472
Residual Std. Error	94.534 (df = 2479)			
Note:	*p<0.1; **p<0.05; ***p<0.01			

You should also repo

Controls include: Student Gender, Student Race, Kindergarten Teacher Experience (yrs.), Days absent in Kindergarten, Free Lunch, School Grade Range, % Students in School Bused, 1st Grade Teacher Experience (yrs.), Days Absent in Received Special Education in K or 1st Grade, Received Special Instruction in K or 1st Grade, Kindergarten Teacher 1st Grade Teacher Education, 1st Grade Teacher Experience (yrs.), Days Absent in 1st Grade, 1st Grade Teacher Race Urbanicity. Student characteristics determined based on Kindergarten Grade information unless otherwise specified.

5.3.2 Grades 1-2

Table 7 below depicts the estimates from the peer effects from attrition between grades 1-2. The regression is fit over all students that stayed between grades 1 and 2 and did not switch schools. Specification (2) is that with school fixed effects and is this my preferred specification. As shown, similar to the previous table, a student's test scores are a product of their own test scores as well as the contemporaneous test scores of their peers. The income effect is also replicated here, students who were not on free lunch had second grade test scores approximately 13.9 points higher than those on free lunch in first grade, holding all else constant. The effect of leavers is insignificant among this cohort. This potentially speaks to the varying nature of attrition in these grades as evidenced in the previous section.

Table 7: Coefficient-Level Estimates for Peer Effects among all STAR students in 1st Grade

	Dependent variable:			
	Total 2nd Grade Test Score		2nd Grade Math Score	2nd Grade Reading Score
	(1)	(2)	(3)	(4)
Total 1st Grade Test Score	0.806***	0.808***	0.197***	0.227***
	(0.012)	(0.012)	(0.004)	(0.004)
Proportion of Leavers in 1st Grade Class	4.030	1.971	0.413	0.269
	(2.562)	(2.862)	(1.066)	(0.992)
Mean Test Score of Leavers in 1st Grade Class	0.006	-0.006	0.003	-0.011
	(0.032)	(0.037)	(0.014)	(0.013)
Variance of Test Score of Leavers in 1st Grade Class	0.00004	-0.0001	0.00004	-0.0001
	(0.0001)	(0.0002)	(0.0001)	(0.0001)
Regular Class 1st Grade	-14.150**	-16.859**	-1.583	-5.976**
	(7.144)	(7.483)	(2.788)	(2.595)
Regular-Aide Class 1st Grade	-6.589	-10.933	1.167	-5.739*
	(8.316)	(8.684)	(3.235)	(3.011)
Regular Class 2nd Grade	7.043	5.787	-1.574	3.938
	(6.973)	(7.311)	(2.724)	(2.535)
Regular-Aide Class 2nd Grade	6.466	9.273	-2.840	6.353**
	(7.939)	(8.219)	(3.062)	(2.850)
Peer Average Test Score 1st Grade	-0.510***	-0.458***	-0.113***	-0.125***
	(0.025)	(0.029)	(0.011)	(0.010)
Peer Average Test Score 2nd Grade	0.721***	0.592***	0.166***	0.148***
	(0.024)	(0.029)	(0.011)	(0.010)
Non-Free Lunch in 2nd Grade	13.912***	16.414***	2.054	3.411**
	(4.830)	(4.833)	(1.801)	(1.676)
Proportion of Leavers in 1st : Mean Test Score of Leavers in 1st	-0.002	-0.001	-0.0001	-0.0001
	(0.001)	(0.001)	(0.001)	(0.0005)
Proportion of Leavers in 1st : Variance of Test Score of Leavers in 1st	-0.00000	-0.00001	-0.00000	0.00000
	(0.00001)	(0.00001)	(0.00000)	(0.00000)
School Fixed Effects	No	Yes	Yes	Yes
Observations	2,908	2,908	2,908	2,908
R^2	0.771	0.712	0.524	0.615
Adjusted R^2	0.768	0.702	0.508	0.602
Residual Std. Error	73.844 (df = 2868)			
Note:	*p<0.1; **p<0.05; ***p<0.01			

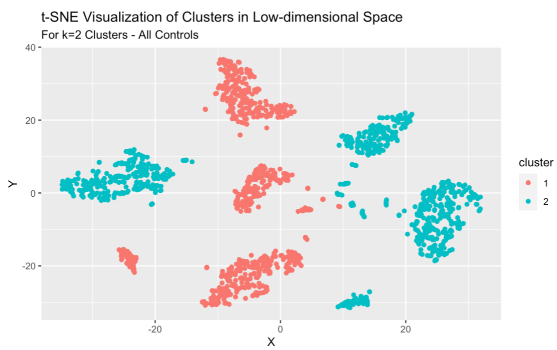
Controls include: Student Gender, Student Race, 2nd Grade Teacher Experience (yrs.), Days absent in 2nd Grade, Receiving Free Lunch, School Grade Range, % Students in School Bused, 1st Grade Teacher Experience (yrs.), Days Absent in 1st Grade, Received Special Education in 2nd or 1st Grade, Received Special Instruction in 2nd or 1st Grade, 2nd Teacher Education, 1st Grade Teacher Education, 1st Grade Teacher Experience (yrs.), Days Absent in 1st Grade, 1st Grade Teacher Race, 2nd Grade Teacher Race. School Urbanicity. Student characteristics determined based on 1st Grade information unless otherwise specified.

5.4 Clustering Analysis of Attrition

5.4.1 Baseline Analysis for Two Groups

(a) Clusters for Data Without Controls for Income

(b) Clusters for Data Without Controls for Achievement



(c) Clusters for Data With all Controls

Figure 6: Visualization of Clusters for K=2

You should explain a little more how

The above figure depicts a visualization of different clustering arrangements for $K = 2$ clusters across datasets with various controls. It appears, at least initially, that these clusters are non-overlapping with clear delineations and high levels of similarity across observations. The dataset shown in Figure 6(a) includes controls for: race; gender; and kindergarten test scores in math, reading, listening, and world skills. The dataset shown in Figure 6(b) includes controls for: race; gender; free-lunch status; special education; special instruction; recommendation to repeat kindergarten; and days absent in kindergarten. The dataset shown in Figure 6(c) includes controls for: race; gender; kindergarten test scores in math, reading, listening, and world skills; free-lunch status; special education; special instruction;

and days absent in kindergarten. A summary table of the clusters is provided below:

	All Controls		No Income Controls		No Test Score Controls	
Cluster:	(1)	(2)	(1)	(2)	(1)	(2)
Number of Observations:	726	836	729	838	725	835
Race						
White	59.2%	59.9%	59.3%	60%	59.2%	60%
Black	39.9%	39.4%	39.7%	39.1%	40%	39.3%
Asian	0.15%	0.36%	0.12%	0.3%	0.13%	0.36%
Hispanic	0.15%	0.11%	0.12%	0.2%	0.13%	0.11%
Native American	0.3%	0%	0.24%	0%	0.27%	0%
Other	0.3%	0.23%	0.24%	0.3%	0.27%	0.23%
Mean Kindergarten Test Scores						
Math SAT Scaled Score	476.3	465.1	475.7	464.7	476.3	465
Reading SAT Scaled Score	431.6	424.7	431.3	424.5	431.6	424.7
Listening SAT Scaled Score	532	525.4	531.9	525.9	532	525.4
Word Study Skills SAT Scaled Score	428.7	422.2	428.1	421.83	428.7	422.2
Male	0%	100%	0%	100%	0%	100%
Free Lunch	57%	56.5%	56.8%	56%	56.7%	56.4%
Special Education	3.3%	5.2%	3.7%	5.2%	3.3%	5.3%
Receives Special Instruction	5%	6.2%	4.4%	5.7%	5%	6.2%
Mean Days Absent	11.3	11.6	11.7	12	11.3	11.6

Table 8: Summary table of characteristics for students in clusters.
Note: all values correspond to kindergarten-grade information unless otherwise specified. All students represented in the table are those that left the experiment. Variable definitions are consistent with Table 1.

5.4.2 Attrition Pathways From Optimal Group Selection

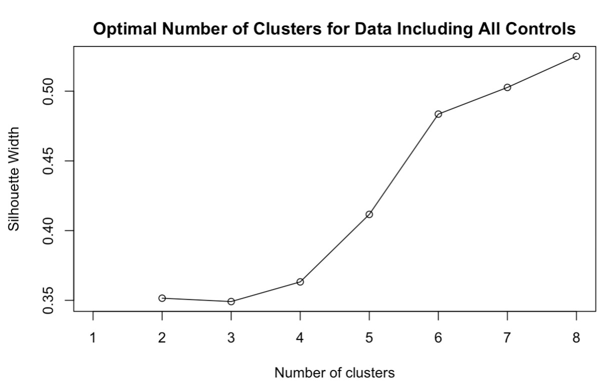


Figure 7: Silhouette Plot for Data Including All Controls

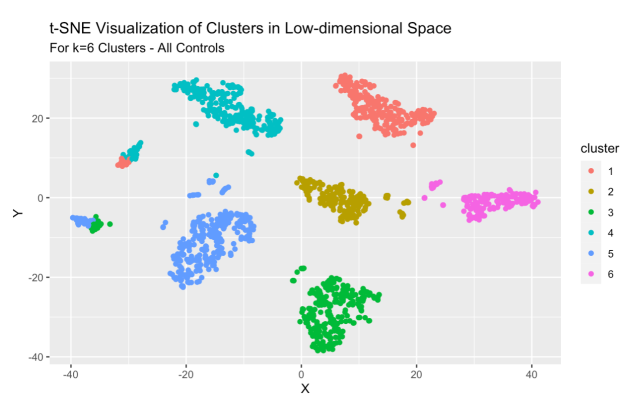


Figure 8: Visualization of Optimal Clustering for Data Including All Controls

The above figure depicts the optimal number of clusters for the dataset including all controls. As shown, it appears from Figure 7 that a value of K = 6 seems optimal. Figure 8's depiction of these clusters shows some distinct regions, though clusters 1 and 4 as well as clusters 3 and 5 do overlap.

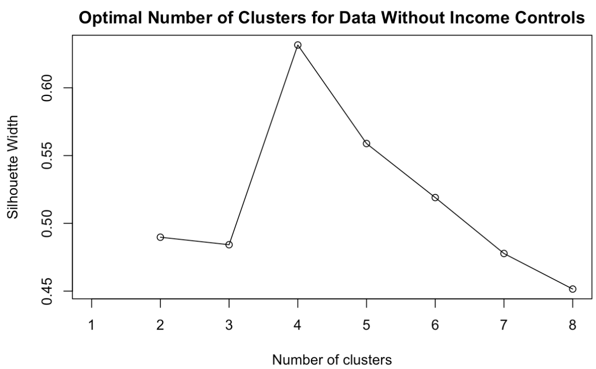


Figure 9: Silhouette Plot for Data Without Controls for Income

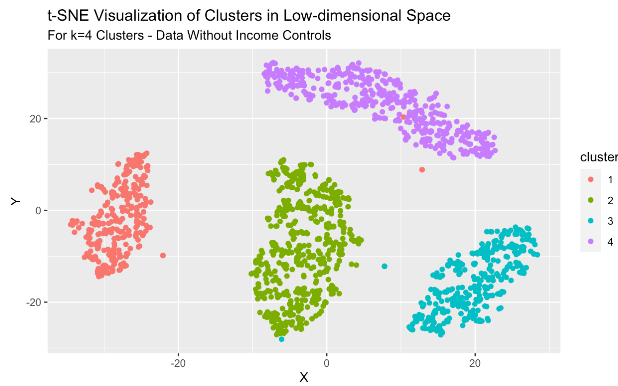


Figure 10: Optimal Clustering for Data Without Controls for Income

The above figure depicts the optimal number of clusters for the dataset without controls for income. As shown, it appears from Figure 9 that a value of K = 4 seems optimal. Figure 10's depiction of these clusters shows clear distinctions between clusters, indicating high levels of convergence in the algorithm and high levels of similarity between points in clusters.

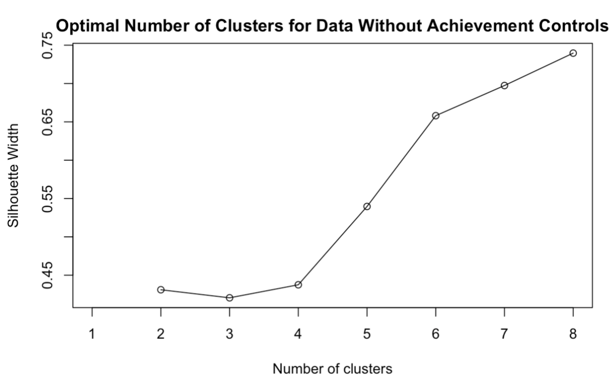


Figure 11: Silhouette Plot for Data Without Controls for Achievement

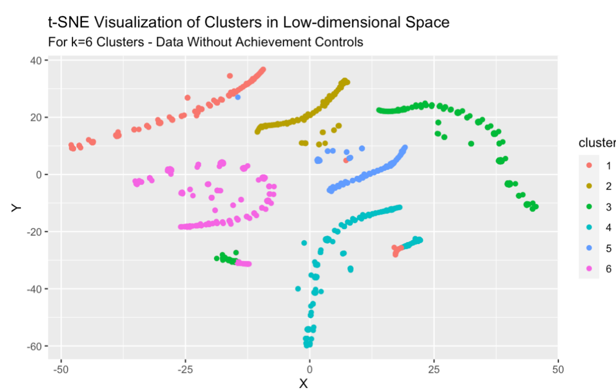


Figure 12: Visualization of Optimal Clustering for Data Without Controls for Achievement

The above figure depicts the optimal number of clusters for the dataset without controls for student achievement. As shown, it appears from Figure 11 that a value of $K = 6$ seems optimal. Figure 12's depiction of these clusters shows some distinct regions, though clusters 1 and 4 as well as clusters 3 and 6 do overlap.

In total, it seems that the clustering is sensitive to the inclusion of controls for income. This is further illustrated in the similarity between Figures 7 and 11. Income may be the driving force behind attrition in STAR as opposed to achievement. It does not seem clear that there are two distinct channels of attrition, instead departures are potentially more random in nature, with little discernable similarities across students.

6 Discussion

In progress.

7 Conclusion

In progress.

8 Bibliography

Boozer, Michael A. and Cacciola, Stephen (2001). Inside the 'Black Box' of Project Star: Estimation of Peer Effects Using Experimental Data.

Dickinson, E. E. (2016). Coleman report set the standard for the study of Public Education. *The Hub.* https://hub.jhu.edu/magazine/2016/winter/coleman-report-public-education/.

Dills, Angela. (2005). Does cream-skimming curdle the milk? A study of peer effects. *Economics of Education Review.* 24. 19-28. 10.1016/j.econedurev.2005.01.002.

Hanushek, Eric A (2008). Education Production Functions. *The New Palgrave Dictionary of Economics.*

Filaire, T. (2018, October 28). Clustering on mixed type data. *Towards Data Science.* Retrieved March 11, 2022, from https://towardsdatascience.com/clustering-on-mixed-type-data-8bbd0a2569c3-data-8bbd0a2569c3

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, 114(2), 497–532. http://www.jstor.org/stable/2587015

Lazear, E. P. (2001). Educational Production. *The Quarterly Journal of Economics*, 116(3), 777–803. http://www.jstor.org/stable/2696418

Preud'homme, G., Duarte, K., Dalleau, K. *et al.* Head-to-head comparison of clustering methods for heterogeneous data: a simulation-driven benchmark. *Sci Rep* 11, 4202 (2021). https://doi.org/10.1038/s41598-021-83340-8

Rohlfs Chris Zilora Melanie (2014). Estimating Parents' Valuations of Class Size Reductions Using Attrition in the Tennessee STAR Experiment. *The B.E. Journal of Economic Analysis Policy, De Gruyter*, vol. 14(3), pages 755-790.

Sacerdote, Bruce (2011). Peer effects in education: how might they work, how big are they and how much do we know thus far? *Handbook of the Economics of Education*, vol. 3, pp. 249–277.

Sojourner, Aaron. (2013). Identification of Peer Effects with Missing Peer Data: Evidence from Project STAR*. *The Economic Journal*.

9 Appendix

9.1 Diagnostic Convergence – AFT Model

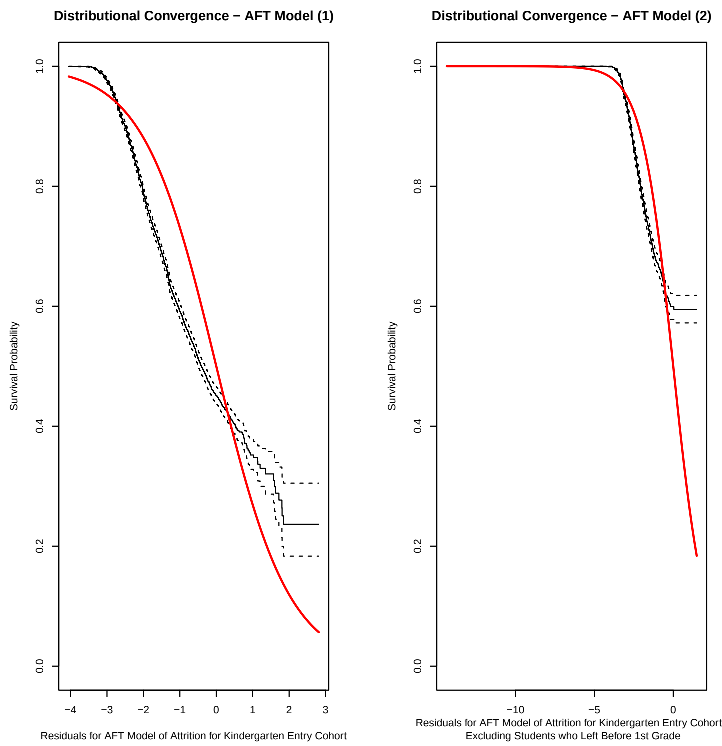


Figure 13: Plot of convergence for AFT Model of Kindergarten Cohort Attrition. Red line represents Weibull distribution, black line represents attrition as modeled by the residuals.

9.2 Diagnostic Convergence – AFT Model of 1-2 Attrition

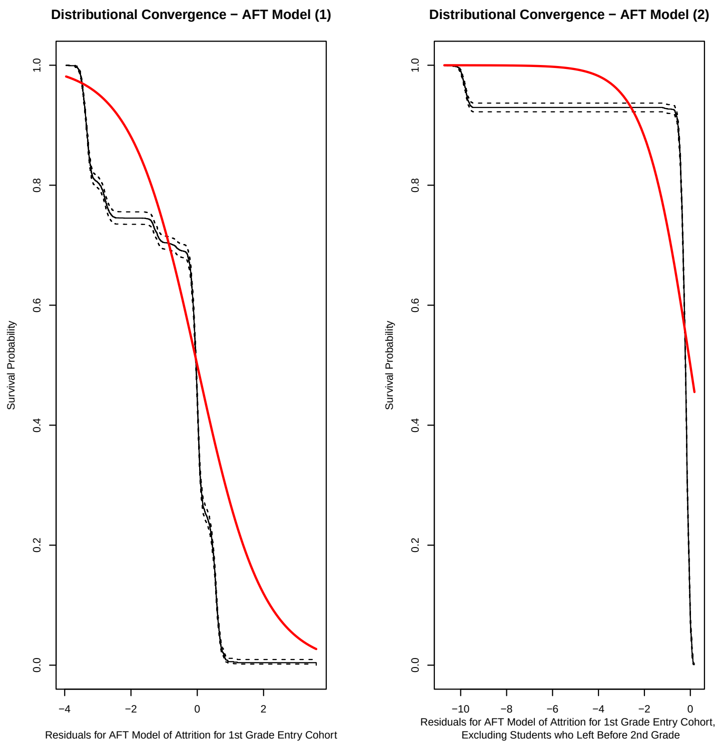


Figure 14: Plot of convergence for AFT Model of 1st Grade and Beyond Attrition. Red line represents Weibull distribution, black line represents attrition as modeled by the residuals.