HIGH SCHOOL DROPOUT RATES IN MASSACHUSETTS: CONTRIBUTING SCHOOL FACTORS AND EFFECTIVENESS OF A CHANGE IN SCHOOL CHOICE MECHANISM

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ORHUN AKBULUT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF ECONOMICS

JUNE 2021

Approval of the thesis:

HIGH SCHOOL DROPOUT RATES IN MASSACHUSETTS: CONTRIBUTING SCHOOL FACTORS AND EFFECTIVENESS OF A CHANGE IN SCHOOL CHOICE MECHANISM

submitted by **ORHUN AKBULUT** in partial fulfillment of the requirements for the degree of **Master of Science in Economics, the Graduate School of Social Sciences of Middle East Technical University** by,

Prof. Dr. Yaşar KONDAKÇI
Dean
Graduate School of Social Sciences ____________________

Prof. Dr. D. Şirin SARAÇOĞLU
Head of Department
Department of Economics ____________________

Prof. Dr. Meltem DAYIOĞLU TAYFUR
Supervisor
Department of Economics ____________________

Examining Committee Members:

Assoc. Prof. Dr. Hakan ERCAN (Head of the Examining Committee)
Middle East Technical University
Department of Economics ____________________

Prof. Dr. Meltem DAYIOĞLU TAYFUR (Supervisor)
Middle East Technical University
Department of Economics ____________________

Prof. Dr. Semih TÜMEN
TED University
Department of Economics ____________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Orhun AKBULUT

Signature:

ABSTRACT

HIGH SCHOOL DROPOUT RATES IN MASSACHUSETTS:
CONTRIBUTING SCHOOL FACTORS AND EFFECTIVENESS OF A CHANGE IN SCHOOL CHOICE MECHANISM

AKBULUT, Orhun
M.Sc., Department of Economics
Supervisor: Prof. Dr. Meltem DAYIOĞLU TAYFUR

June 2021, 102 pages

It is viewed that high dropout rates in high school in the US is still a relevant and persistent problem. Apart from the individual reasons, school-related factors also contribute to the decision of dropping out. School choice mechanisms, the algorithms through which students are placed into high schools, might also be relevant for the decision to drop out due to the importance of placing the right students into right schools. To this end, I study the school related factors influencing the dropout rates of schools in Massachusetts through school level data and then I analyze whether the change of school choice mechanism for Boston high schools in school year 2006-2007, from Boston Mechanism to Student Optimal Stable Mechanism, had any significant effect on dropout rates. For the purpose of understanding the underlying school related factors, I develop two models, one single year OLS model and one multi-year fixed effects panel data model. To understand the effect of the school choice mechanism change, I employ a difference-in-differences methodology. Results indicate that, even after controlling for the other school-related variables and school-time fixed effects, attendance rate is the most significant variable in predicting the dropout rates, with a strong negative association. It is observed that socio-economic composition and racial composition of the students are significant when a single year is considered and insignificant when the school and time fixed effects are controlled for. School choice mechanism change, however, is determined to be statistically insignificant in affecting the dropout rates.

Keywords: High School Dropout Rates, School Choice Mechanisms, Fixed Effects Model, Difference in Differences, Economics of Education

ÖZ

MASSACHUSETTS EYALETİNDE LİSE BIRAKMA ORANLARI: ETKİ EDEN OKULLARLA İLGİLİ FAKTÖRLER VE LİSELERE YERLEŞTİRME SİSTEMİ DEĞİŞİKLİĞİYLE İLİŞKİSİ

AKBULUT, Orhun

Yüksek Lisans, İktisat Bölümü

Tez Yöneticisi: Prof. Dr. Meltem DAYIOĞLU TAYFUR

Haziran 2021, 102 sayfa

Amerika'daki liselerde okulu bırakma oranlarının yüksek olması güncelliğini korumakta olan kalıcı bir problem olarak görülmektedir. Bireysel nedenlerin yanı sıra, okullarla ilgili faktörler de liseyi bırakma kararında etkili olmaktadır. Liselere yerleştirme mekanizmaları, yani öğrencilerin liselere yerleştirilmelerinde kullanılan algoritmalar da doğru öğrenciyi doğru okula yerleştirmenin öneminden dolayı liseyi bırakma kararında etkili olabilmektedir. Bu bağlamda, bu tezde liseyi bırakma oranlarına etki eden okullarla ilgili faktörler, Massachusetts eyaletinin okul seviyesinde verisi üzerinden çalışılmaktadır ve sonrasında Boston şehrindeki liseler için 2006-2007 öğretim yılında gerçekleşen "Boston Mechanism" uygulamasından "Student Optimal Stable Mechanism" uygulamasına geçişin lise bırakma oranları üzerinde anlamlı bir etkisinin olup olmadığı analiz edilmektedir. Etki eden okullarla ilgili faktörleri anlamak amacıyla, iki model geliştirilmiştir, tek yıllı En Küçük Kareler (EKK) Modeli ve çok yıllı Sabit Etkiler Panel Veri Modeli. Liselere yerleştirme sistemi değişikliğinin etkisini anlamak amacıyla ise, Farkların Farkı metodolojisi kullanılmıştır. Sonuçlar, okula devam oranlarının, diğer okullarla ilgili faktörler ve okula özgü sabit etkiler kontrol edildiğinde dahi, okulu bırakma oranlarını tahmin

etmede en etkili değişken olduğunu, güçlü bir negatif ilişkiyle göstermektedir. Diğer okullarla ilgili faktörler ve okula özgü sabit etkiler kontrol edildiğinde, kayıtlı öğrencilerin sosyoekonomik ve ırksal kompozisyonunun tek bir yıla bakıldığında anlamlı olduğu ve okul-zaman etkileri kontrol edildiğinde bırakma oranları üzerinde etkili olmadığı gözlenmiştir. Liselere yerleştirme sistemindeki değişikliğin ise bırakma oranlarını etkilemede istatistiksel olarak anlamlı olmadığı saptanmıştır.

Anahtar Kelimeler: Lise Bırakma Oranları, Liselere Yerleştirme Sistemleri, Sabit Etkiler Modeli, Farkların Farkı, Eğitim Ekonomisi

To my family

ACKNOWLEDGMENTS

Firstly, I would like to express my deepest gratitude to my supervisor, Prof. Dr. Meltem Dayıoğlu Tayfur who guided me with patience and channeled my efforts into the right direction. Without her assistance, this work could not have made it this far. I would also like to thank Prof. Dr. Semih Tümen and Assoc. Prof. Dr. Hakan Ercan for their valuable critiques and suggestions.

In addition, I thank Asu for giving inspirations on how to approach a thesis subject and always being there to support me. I wish her the best on her academic career.

I would like to thank Yakup, Utku, Adil and Fatih for their friendship and all the memories. I owe a great deal to them for their support on both my undergraduate and graduate years.

Last but not the least, I would like to give special thanks to my family for being by my side and motivating me throughout my entire life. I am grateful for their unending love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

High dropout rates in high school in the US remain a serious concern. According to the metric of event dropout rate used by the National Center for Education Statistics (NCES), about 4.7 percent of students dropped out of high schools in the US in the 2016-2017 school year. Event dropout rate is defined as percentage of dropouts in a given school year among the students who are currently studying in any grade between 10 to 12 and in the age group of 15- to 24-year-olds. The act of dropping out is defined as leaving school without a high school diploma or any alternative equivalent credential such as a GED certificate (McFarland et al (2020)). When specific student populations are taken into account, the rates are even more alarming. For example, for the school year 2016-2017, the event dropout rate was 5.5 percent for black students, 6.5 percent for Hispanic students and 6.2 percent for students with disabilities.

Another metric of status dropout rate used by the NCES is defined as the percentage of 16- to 24-year-olds without a high school diploma or equivalent credential who are not registered to any school, among the civilian and noninstitutionalized population. In 2017, the status dropout rate was 5.8 percent. That some student populations have higher dropout rates such as black students hold true for this metric as well.

Figure 1 plots the mentioned rates in the US for each year since 1992. It can be observed that in this period, the status dropout rate decreased more than fifty percent from its top point in 1995 at 12 percent to 5.8 percent in 2017. However, the event dropout rate lingers between 4 and 6 percent, being somewhat stable around 6 percent in 2014-2017. It is important to note that the aforementioned measures may not always follow the same path or move together. The reason is that, by definition, event dropout rate is related to a student’s decision to drop out and measures the number of high school students that dropped out from high school in a given year. Status dropout rate, on the other hand, measures the number of people in the age group of 16- to 24-year-olds that are not enrolled in high schools and do not have a high school diploma/GED. Thus, by definitions, event dropout rate is a flow variable and status dropout rate is a stock variable. A student who has dropped out can still return to high school until age 21 and continue his/her education. After age 21, a student can still earn GED and it serves as a high school diploma. These reasons can explain why status dropout rate may decrease when event dropout rate is constant or increasing. When Figure 1 is taken into account, it can be seen that status dropout rate has fallen from around 11% to around 6%. Nonetheless, the high event dropout rate, being a more direct measure of enrolled students dropping out of school, still appears to be a problem, paralleling the main finding of Heckman and LaFontaine (2010) that graduation rates are, although not very low, not at desirable levels.

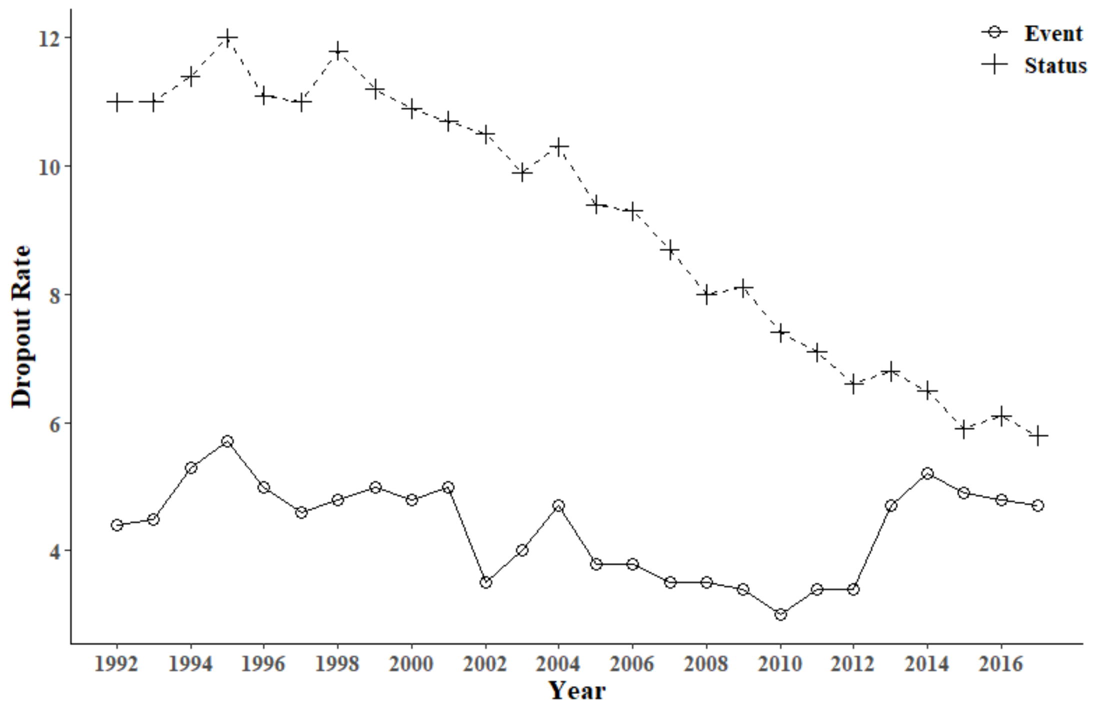


Figure 1: Event and Status Dropout Rates in the United States

Note: Event dropout rate is the percentage of dropouts in a given school year among the students who are currently studying in any grade between 10 to 12 and in the age group of 15- to 24-year-olds, status dropout rate is the percentage of 16- to 24-year-olds without a high school diploma or equivalent credential who are not registered to any school, among the civilian and noninstitutionalized population.
Source: National Center for Education Statistics

Another motivation to study high school dropout rates would be its adverse consequences. U.S. Department of Labor Bureau of Labor Statistics reports that, for the third quarter of 2019, the median weekly earnings of full-time wage and salary workers of age 25 years and older who have dropped out of high school was $606, whereas those who have high school diploma without a college degree was $749. People who are high school dropouts are more prone to unemployment and lower wages compared to high school graduates (Belfield and Levin (2007)). In another study, it is stated that dropouts are more prone to having health problems and being involved in criminal activities compared to high school graduates (Rumberger and Lim (2008)). Moreover, high dropout rates affect the overall economy poorly as well due to social costs. As high school drop outs tend to earn less compared to high school graduates, they pay less taxes and according to Rumberger and Lim (2008), dropouts depend more on government programs for income. Thus, analysis of dropout rates still remains a relevant topic in that regard.

The reasons for dropping out of high school can be numerous. A recent report by the non-profit organization America's Promise Alliance (Hynes 2014) states that academic failure is the biggest reason for an American high school student to drop out of high school. According to the report, 27.6 % of the high school dropouts state that failing too many classes was the biggest reason for dropping out. According to the same report, however, the second biggest reason for dropping out of high school is being bored or disengaged with the school, with a staggering 25.9 % and 17.7 % stated that the reason they quit schooling was that they felt as though no one cared if they attended the school or not. Thus, it is obvious that school-related factors greatly affect whether a student drops out or continues his/her education and studying to what extend school related factors affect the dropout rates would yield important policy implications for high schools.

There is a vast literature on why students drop out of high school and what the predictors and associations are with the dropout rates, as will be detailed in literature review. However, one aspect that is often overlooked in the literature is the connection between high school dropout rates and school choice mechanisms. School choice, defined as the ability to choose the school a child will attend, is one of the most challenging issues parents face when their children reach school age. From the perspective of the market design literature, it is stated in Vulkan et al. (2013) that the school choice problem consists of a finite set of students and schools with finitely many seats available for enrollment, for any student a preference relation over the set of schools and for any school, a priority ordering of the students based on some criteria.

Taking these elements into account, students are placed to schools. This assignment is determined by a matching such that each student is matched with at most one school or remains unmatched while obeying the maximum capacity of each school. The most commonplace matching algorithm used around the world for the school choice problem is Boston Mechanism (BM). In all school districts of Massachusetts except Boston, BM algorithm is in use. In Boston, however, it has been replaced in the 2006-2007 school year by Student Optimal Stable Mechanism (SOSM) proposed by Abdulkadiroğlu and Sönmez (2003). The primary reason for this change is that the BM algorithm may cause some parents to misrepresent their true preferences in terms of schools to take advantage of the inherent drawbacks of the system, which may harm the truth-telling parents. The details of these matching algorithms and how education system in Massachusetts works are given in Chapter 3.

Therefore, a fairer school choice mechanism has been implemented in Boston starting with 2006-2007 school year. However, the existing literature studying the connection between school choice mechanisms and dropout rates, which is detailed in Chapter 2, almost entirely focuses on the effects of introducing a randomized lottery to the school choice system to give the students a chance at attending a high school other than their assigned schools. The effects of this specific policy change on student achievements and dropout rates, to my knowledge, has not been studied. The results of such a research may result in important policy recommendations. If it is the case that the fairer mechanism translated into improved schooling outcomes such as reduced dropout rates, this would pave the way for policy-makers to encourage other school districts in Massachusetts to implement the Student Optimal Stable Mechanism instead of the Boston Mechanism.

The purpose of this thesis is thus twofold. The first aim is to understand how school-related characteristics contribute to predicting the high school dropout rates using school-level data from Massachusetts. I present two statistical models which, hopefully, shed light on the determinants of high school dropout rates. The second aim is to analyze the connection between dropout rates and school choice mechanisms, mainly seeking an answer to the question of whether and to what extent dropout rates have been affected from the aforementioned algorithm change in Boston high schools, using a difference-in-differences estimation.

The remainder of this thesis is structured as follows. Chapter 2 outlines the theoretical and empirical literature on high school dropout rates and factors affecting it, together with the literature on the effect of school choice on dropouts and student achievement. Chapter 3 introduces the data together with some descriptive and exploratory analysis and details the models used. Chapter 4 presents the results of these models. Finally, Chapter 5 concludes.

CHAPTER 2

LITERATURE REVIEW

2.1. Theoretical Literature

There is considerable research on why people go to school. Sen (1980) defines the concept of capability as one's liberty to achieve valuable things one desires to do or be. For instance, reading and processing information can be thought of as capabilities crucial for a normal life. Checchi (2006) argues that a simple explanation of the demand for schooling is that receiving education creates minimal capabilities one needs to have to maintain a normal life and makes the following ordinary acts that requires some level of education; using public transport, finding a street address, checking a bill in a restaurant, signing a cheque, enrolling your child at school, reading the instructions on an electrical appliance, and so on. Thus, nearly all the nations implement compulsory schooling until a certain degree in order to equip their citizens with basic capabilities for a normal life.

Despite being a good starting point, education being the medium of acquiring minimal capabilities do not offer much insight into the question of why so many people participate in schooling beyond the levels where they acquire these capabilities. In this respect, economists have made important contributions to the way we think about schooling, considering it as investment rather than a pure consumption good. One of the most widely accepted theories of educational demand is the human capital theory formalized by Becker (1962). Human capital can be defined as the knowledge and skills the labor force possesses which help in production of goods and services (Goode 1959). Firstly, Mincer (1958) observed that education and training significantly explain the differences in personal income across individuals and found that wages increase as the level of education rises. He concluded that education is the primary source of any person to increase his/her human capital. Then, building upon these insights, Becker (1962) formalized what is known as the human capital theory.

According to this theory, the decision to receive education can be viewed as an investment decision by an individual in his/her human capital. Through this investment decision, income that can be received by being employed today is waived in order to increase potential future income. This theory posits that the demand for schooling increases with future expected gains and decreases with cost of schooling. Also, it predicts that a more talented person will be more willing to receive education because his/her marginal return is higher (Checchi (2006)). It assumes that each person will invest in education to a certain level if, at time of enrollment, net present value of that education level in terms of both monetary and non-monetary benefits and costs is positive.

Human capital theory is built on the assumptions that individuals make this choice based on both monetary and non-monetary benefits/costs of receiving education and individuals are perfectly able to calculate these benefits and costs (Aina et al. (2018)). However, as the education decision process is much more complex, it is argued that there are many reasons why this process may lead to drop-outs although returns to education are high. Two of the mostly reviewed reasons are basic credit constraints and myopic behavior of individuals. Firstly, an individual may decide to not receive education purely based on monetary reasons. Empirical studies suggest that, in recent years, credit constraints became more and more noteworthy as the demand for credit boomed in comparison to available credit opportunities. The reason for that is that cost of receiving education rose considerably, along with the both monetary and non-monetary benefits of receiving education (Lochner and Monge-Naranjo (2012)). Since the educational costs of receiving education is high and the returns are realized many years after the cost is incurred, especially poor families may not afford to send their children to school.

Secondly, although human capital theory assumes that individuals are perfectly able to calculate the benefits and costs of receiving one more year of education, it is well known that some individuals may have myopic behavior. Since the returns to the education are uncertain and occurs many years after the investment, it is possible for the individuals to incorrectly assess the returns. In most of the cases, the investment decision is made by parents and the ones who receive the returns are children. Thus, some poor parents might not have the sufficient knowledge to correctly assess the size of benefits their children will have with another year of schooling. Thus, myopic behavior of the decision-makers might also lead to the dropping out of school (Attanasio (2015)).

Another widely referenced model related to the demand for schooling from economic literature is its role in being a signal of ability. Personal abilities are regarded as private information in that firms cannot perfectly observe applicants' abilities. However, firms are primarily concerned with its employees' abilities as productivity and hence profits increase with abilities. This phenomenon of parties having different information sets is called imperfect information and can lead to market failures in the worst case. Spence (1973) argues that to overcome the imperfect information problem, people send a signal about their ability to the firms by receiving higher levels of education. From the firms' perspective, since receiving education is more costly (more difficult) for people with lower abilities, having a high education level is a credible signal that the potential employee has a high ability level. This way, education helps the firms distinguish low and high ability workers, and thus people receive education to increase the likelihood of being employed by the firms.

When the educational literature on dropping out is examined, it is observed that models that explain the dropout behavior fall into two broad categories: models that highlight individual characteristics and models that highlight institutional characteristics as determinants of school dropout. The former is related to individual student characteristics that affect their decision to drop out and the latter is related to students' families, schools and communities.

Nearly all the models that fall in the first category of models mentioned above suggest that dropping out of school is affected by various factors related to: educational and general attitudes, early and recent school performance, and academic and social behaviors. A commonly referred model developed by Wehlage et al. (1989) suggests that two general factors conjointly influence dropping out process, school membership and educational engagement. School membership is about the social aspect of student mentality that covers characteristics such as relationship with other students, participation in school activities, and having a general positive attitude about the school. Educational engagement is, on the other hand, about the academic aspect of student mentality that covers the external rewards of academic achievement and internal satisfaction with the subjects thought in school and the way in which these subjects are programmed around a certain path based on students' capabilities. It is suggested that a student who is not engaged in schooling both socially and academically is a potential dropout.

In another paper, Finn (1989) describes two different models to conceptualize dropping out as a progressional process that potentially begins at earliest grades. In his first model called frustration-self-esteem model, he argues that school failure is the starting point of the process that eventually leads to a student's dropout. He suggests that early school failure results in low self-esteem and this in turn results in problem behaviors in students. These problem behaviors further damage how a student performs in the school and thus further exacerbate his/her self-esteem, which eventually leads to dropout. The second model, called participation-identification model concentrates on involvement of a student in schooling activities both behaviorally and emotionally. This model suggests that risk of dropout is minimized when a student participates in school relevant activities. These activities include conforming to teacher directions and class requirements, doing homework, taking part in other learning activities, participation in extracurricular activities and taking part in the governance of school. According to the participation-identification model, lack of participation in schooling results in poor school performance and this in turn leads to lack of identification with the school, thus resulting in a dropout.

The core factors associated with dropouts in these models are student engagement and student motivation. The models of student engagement and student motivation are often related and both incorporate concepts from each other (Rumberger and Lim (2008)). Newman et al. (1992) defines academic engagement in their model as a student's effort and emotional investment on learning and understanding the material. As being an inner qualification based on effort and investment, engagement is indirectly observed through such variables as attendance and time spent on academic work. The model argues that need for competence, actuality of the work they are required to complete and how much they identify themselves with the school are the major influencers of academic engagement. Connell (1990) also came up with a model for student motivation, arguing that students are more encouraged to participate in education if their psychological needs in the form of self-determination, proficiency and relatedness are met. This implies that if they are not met, students may be more prone to dropping out.

As mentioned, one can look at the theoretical literature on dropout behavior from two broad aspects, individual characteristic models and institutional characteristic models. Tinto (1987) develops a widely accepted connection between these two categories. Labeled as the theory of institutional departure at the postsecondary level, Tinto's model argues that the process of dropping out is firstly affected by students' characteristics such as commitment, expectations and family background. The reasoning is that these attributes collectively precondition students to behave in certain ways to different conditions and thus affecting the decision to drop out. Tinto argues that given these attributes, there are two aspects of an institution affecting whether a student remains in school or becomes a dropout. First aspect is social aspects of a school, which is about how well students are integrated socially to school and to the schooling concept it conveys to students through its social practices. The second one is about academic aspects of a school, how well students are integrated academically to the school and its ability to engage students academically through its academic practices. Both aspects cover the formal as well as informal practices carried out by the school. The model posits that some degree of engagement of a student with the institution in terms of either academic aspects or social aspects is a necessity for remaining in school. If a student is engaged through academic aspect and not through the other, as long as social interaction is done elsewhere, student is likely to remain in school. Also, if a high social interaction is coupled with a minimum academic performance, drop out is less likely.

As argued above, Tinto's model establishes a bridge between individual characteristics models and institutional characteristics models. While individual attributes are the main determinants in high school dropout models, it is generally accepted that institutional features related to schools such as composition, structure, resources, and practices shape students' academic performance and attitudes in a way that would lead to dropping out. Among these features, composition indicates the characteristics of the students within a school, structure covers the size and location of a given school, resources indicate all physical, fiscal and human resources and practices mean the general process and environment through which schooling functions are implemented (Rumberger & Palardy (2004)).

These features can be modeled in the framework of what Rumberger & Palardy (2004) calls an economic model of schooling. According to this model, Hanushek (1986) argues that there are three major components of schooling; inputs, educational process and outputs. Schooling is modeled as taking the inputs of students, teachers and other resources and through the educational process, which describes how these inputs are used, outputs the states of either graduation or dropping out. A visual representation of this model is given in Figure 2.

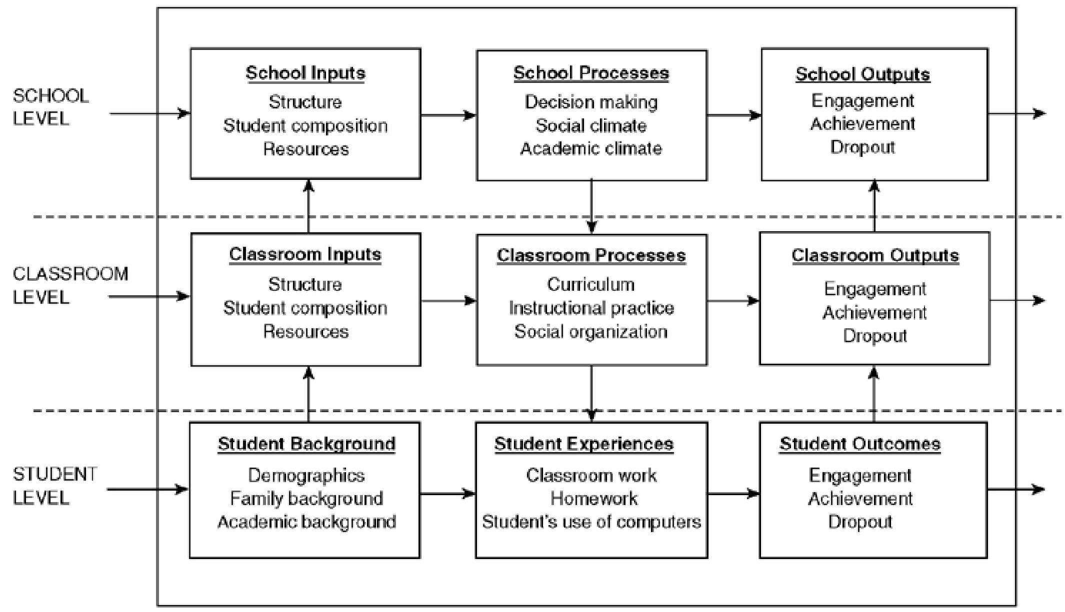


Figure 2: Framework of Economic Model of Schooling

Source: Rumberger, R. W. & Palardy, G. J. (2004). Multilevel models for school effectiveness research.

According to this framework, schooling is modeled to be operating at three hierarchical levels, which are students, classrooms and schools. At each level, inputs are converted to schooling outputs through relevant processes. Inputs are generally considered as given to the schools. Thus, it is assumed that schools do not have control over these inputs. However, schools do have control over the school processes such as academic and social climate and therefore can have an impact on school outputs. School inputs are largely "given" to a school and therefore are not alterable by the school itself.

2.2. Empirical Literature

Since the focus of this paper is understanding the association between school related factors and dropout rates, only the empirical literature on institutional perspective is outlined. When the empirical research on dropouts is examined, it can be seen that there is a broad empirical literature studying the contributors and predictors of high school dropouts via above mentioned statistical techniques with mixed results.

Rumberger (1983) finds in his study that the background of a student's family plays a vital role in dropping out. As the data for his study, he uses the results of a survey conducted in the first half of 1979, namely National Longitudinal Survey (NLS) of Youth Market Experience (Center for Human Resource Research (1980)). The survey consists of interviews with the respondents who were 18 to 21 years old and not enrolled in high school and he identifies the respondents as either a high school graduate or a high school dropout. The data includes, in addition to respondents' individual traits such as race and gender, a variety of information on background characteristics, attitudes, aspirations, educational and labor market experiences and personal characteristics. With dependent variable being the probability of dropping out of high school, he develops a multivariate probit model to analyze the predictors of the high school dropouts. He observes that students coming from lower social class families tend to have higher dropout rates compared to students from high social class families. He also points out in his simulations that differences in dropout rates across racial groups can in most part be explained by family backgrounds of these students.

McNeal Jr (1999) studies a longitudinal data (National Educational Longitudinal Study) about students in 8th grades beginning in 1988 and continues to be updated every two years. He models parental involvement as social capital and analyzes parent involvement effects on dropping out. He uses the data for the students that meet the following conditions; they must be from public schools, they must have the basic achievement tests and must have attended in at least three interviews (in their 8th grade, 10th grade and 12th grade). He uses principal-components factor analysis to analyze parental involvement and comes up with four different factors; supervising of children, conversations with children, parent-child discussion, parent-teacher organization (PTO) involvement and educational assistance. Then, these factors are included in OLS and logistic regression models to ascertain parental involvement's effect on dropping out. He demonstrates that if parents are involved with childrens' schools in the form of PTO and continues supervising the children, students are less likely to drop out.

Rumberger (1995) also uses the National Education Longitudinal Survey of 1988 as base year data and 1990 as the follow-up year data. The data consists of both individual level variables and school level variables for each student in the dataset. To analyze dropout rates, he develops two different sets of models. He first develops a set of logistic regression models using only the individual level variables and secondly a set of hierarchical linear models (HLM) with nonlinear setting, using results of the first set of models as a benchmark for individual-level variables, in order to differentiate between schools using school-level variables. Through these models' results, he observes that students coming from single parent families, step-families, and non-English- speaking families all have higher possibility of dropping out. Pong and Ju (2000) with the same dataset of NELS 1988, on the other hand, observes by means of a logistic regression that separation of two parent families, controlling for the income loss, does not seem to be increasing the likelihood of a drop out.

There are quite a large number of school-related variables that is used in the literature to predict dropout rates. One of the factors considered as a proxy of a school's resources is the number of pupils per teacher enrolled in that school. McNeal Jr (1997), using a hierarchical logistic regression model, observes that an increase in the student/teacher ratio significantly increases the risk of dropouts in high schools. With the binary variable of whether a student dropped out or not being the dependent variable, he considers the High School and Beyond (HSB) dataset of the NCES by restricting his attention to sophomores from regular public high schools. He uses a set of independent variables concerning individual traits such as race, gender, socioeconomic status of families, school structure variables such as number of students, pupil/teacher ratio, and school context variables such as academic emphasis and percent minorities. The findings also coincide with Rumberger (1995) that even controlling for social class differences and student composition, high student/teacher ratio increases the likelihood of dropping out.

Apart from observing the same effect, Rumberger and Thomas (2000) models the binary variable of whether the student dropped out or not by means of a hierarchical generalized linear model (HGLM) for the same dataset of NELS 1988 and finds that for the schools whose students perceive the teachers as high quality, the dropout rates are lower. However, in the same paper, it is found that schools whose principals think their teachers are more qualified had higher dropout rates, which indicates how students and management of schools think differently in terms of the teachers.

Li (2007) also reinforces the effect of student/teacher ratio. He uses the HSB dataset of NCES for 1980 and the follow-up survey of 1982. He models the high school dropout by means of a hazard analysis, which indicates that high school behavior is characterized by strong state dependence so that the probability of a student dropping out depends on how long he/she has been in that school. With a set of individual, family and school-related independent variables, he observes that dropout hazard increases 1.3 percent as result of a unit increase in students per teacher, despite finding insignificant effects for percentage of teachers with MA/PhD or having more than 10 years of experience.

Another school related factor analyzed in literature is student composition of schools. As Gamoran (1992) puts it, it is also possible to see the effects of social mixture of students influencing student success, in addition to individual social characteristics of students. Sander (2001) employs a regression model in a similar fashion employed for this thesis. With dependent variable being the annual dropout rates of 1995-1996 school year in Chicago public schools, he runs an OLS regression using a set of school-level characteristics. As in this thesis, unit of analysis is schools. He reports that an increase in the proportion of limited English proficiency students and low-income students increases the dropout rates. Also, he observes a decrease in dropout rates with the proportion of Asian students, although no significant difference is found for Hispanic or Black students' proportion to the whole student population.

McNeal Jr (1997) also finds that the percentage of minorities enrolled in a school directly affects the dropout rate. Rumberger and Thomas (2000), however, finds that although the composition of students based on their social classes had strong effects on dropout rates, the racial composition did not. It is a well-known and researched idea that the social and academic environment of a school also plays a vital role in predicting the dropout rates. As discussed in Rumberger and Lim (2008), an index created from the questions which are asked to students about the school climate successfully predicted that positive environment in schools reduces the likelihood of dropping out, according to the study of Worrell and Hale (2001).

An indicator used in the literature to quantify the overall environment in school is the overall attendance rate, as in Rumberger and Thomas (2000) and Christle et al. (2007). Rumberger and Thomas (2000) argues that even after specific student level features are controlled for, school-level measure of attendance remains a strong predictor of school dropout rates. It is observed in the study that schools with higher attendance rates had lower dropout rates than did schools with lower attendance rates. Christle et al. (2007) also observes that school level attendance rate is negatively correlated with dropout rates. He points out that next to academic achievement, the rate of school attendance showed the strongest relationship to dropout rates.

As per the second aim of this paper, which is to analyze the effect of a school choice algorithm change in Boston high schools on dropout rates, it is observed that the existing literature almost entirely focuses on the effects of introducing a randomized lottery to the school choice system to give the students a chance at attending a high school other than their assigned schools. Deming et al. (2014) study the impact of a public-school choice lottery in Charlotte-Mecklenburg Schools (CMS) on degree completion. Their main purpose is to understand the effects of a change in school assignments from predetermined assignments to Boston Mechanism that took place in 2002. They use student level administrative data from CMS linked to the National Student Clearinghouse (NCS), which is a national database of postsecondary enrollment that records college enrollment and degree completion for almost all colleges in the US. They match the lottery applicant and non-applicant files taken from CMS, which contain individual choices, lottery numbers, priority groupings and admission outcomes, to NCS data and data taken from CMS database that contain demographics, enrollment histories, test scores etc. for its students. To estimate the average impact of winning the lottery across multiple schools and grades, they employ a 2 stage least squares estimation (2SLS) methodology. Firstly, they use a variable indicating whether the lottery number assigned to a student is a winner number or not as an instrumental variable (IV). Together with pre-lottery student variables and lottery fixed effects, they estimate enrollment variable in the first equation. Then, they use the estimated enrollment as independent variable in estimating the academic outcomes of interest in the second equation of 2SLS methodology. As a result, they find a significant overall increase in the rate of finishing school among lottery winners who attend their first-choice school.

Lavy (2010) examines the effects of a program change in Tel-Aviv, Israel. The old mechanism takes the zones students are currently residing into account and the new one makes school choice free from these restrictions to a certain extend. He uses the student-level administrative records of Israeli public schools during the 1992-1994 school years. The dataset contains individual level data on students' schools, class levels, family-background variables and achievement data including dropout rates and test scores. Due to the gradual implementation of the new choice program, school district 9 is chosen as a pilot district and the new program is first applied in this district. Thus, the school districts that joined the program later compared to the school district 9 forms the control group and the school district 9 is the treatment group. In a much similar fashion to this thesis, he uses difference-in-differences (DID) estimation and finds that the program led to reduced dropout rates, together with lower levels of classroom disruption and violence.

Cullen et al. (2006) used student-level administrative data on applications submitted to Chicago Public Schools (CPS) to participate in a randomized lottery that makes it possible for the students to choose a high school other than the one they are automatically assigned to for the years of spring 2000 and spring 2001. Including covariates such as student demographics or prior achievements, they run OLS regressions for a set of outcomes including dropout rates. By including the variable of whether a given student became the winner of a given lottery, contrary to the findings of Lavy (2010), they found that the difference in dropout rates for lottery winners and losers is not statistically significant.

In another empirical paper, Hastings et al. (2012) use daily student-level data from a low-income school district. The federal No Child Left Behind Act (NCLB) of 2001 introduced a new system allowing parents whose children are at unsuccessful schools an option to participate in a lottery which may result in a better school for their children. By using a difference-in-difference estimation, they observe that winners of the lottery after they are informed that they won the lottery has better outcomes in terms of absenteeism and discipline problems. Thus, they conclude that even the possibility of attending a better school motivates students to be better engaged to schooling.

CHAPTER 3

DATA AND METHODOLOGY

3.1. School Choice System in the US

In Massachusetts, there are 351 municipalities. 57 of these are officially regarded as city and the others are towns. Population, number of high schools and number of high school students of these cities can be found in Table 14 given in Appendix A. From a geographical perspective, Massachusetts is divided into 14 different counties, whose population and high school statistics can be found in Table 15 given in Appendix B. From an educational point of view, in the United States, each state is divided into several school districts. A school district is the governing administration unit serving a particular geographic area, established to implement the public education system in that area. The district boundaries are determined by the state education agencies and usually follow county governments and boundaries. All public schools are operated under a school district. Apart from public schools, there are also charter schools in the US, which receive government funding and accountable for their educational outcomes but operate independently from the educational rules in their state. Since charter schools are independent, each one is counted as a district.

As of 2019-2020 school year, there are 401 active school districts in Massachusetts and 78 of these 401 active districts are charter schools, usually following municipality boundaries. However, it is important to note that the numbers may vary for each year as some districts may be closed, opened, merged with others and charter schools may be opened/closed. In the school districts of Massachusetts, students seeking a seat in Kindergarten, Grades 1, 6, and 9 are required to submit a preference ranking of schools in the spring semester of each school year to their respective school district offices. Unless they request and receive a transfer to another school, students who are in the remaining transition grades (which are the grades other than Kindergarten, Grades 1, 6, and 9) are able to continue their education in their current schools.

Policies on how students are assigned to schools within a school district are set by the school committee of that district. The differences in these policies across the school district of Massachusetts originate from two sources. The first one is that for each school, there are some priority rules that determine how the applicants to that school are ordered and these rules are set by the school committee of the district in which that school operates. Each school district applies a different set of the priority rules, although there are common priority rules across the districts. For instance, most of the districts do not accept applications from students living in another student districts and the ones that accept these applications give the priority to the students living in their district boundaries and only consider the applicants of different districts in case there are empty seats after the in-district students are registered. However, there are different priority rules across the districts as well. For instance, for some of the school districts, a student who already has a sibling already enrolled in a school takes priority over remaining students. Living in the walk zone of a school, which is applied for schools in Boston district, is another example of priority rule of schools over applicants that generate differences in student-school assignment process across the districts. A random lottery number is used to break the ties for any priority rule. Also, in some districts, there are several special admission high schools that process applicants separately from the centralized assignment mechanism. These include schools that require an interview, presentation of a portfolio or scores on an entrance examination.

Second source of differences in student-school assignments across school districts is the algorithm by which the assignment of students to schools is done, where the students' preferences, schools' priority ordering on the students and capacities are taken as given. This assignment is determined by a matching such that each student is matched with at most one school or remains unmatched while obeying the maximum capacity of each school. The most commonplace matching algorithm used around the world for the school choice problem is Boston Mechanism (BM). In all school districts of Massachusetts except Boston, the BM algorithm is used. According to the BM, firstly, each school considers only the students who chose that school in first place and assigns its seats to these students based on their preferences on students. In the second step, only the ones who are not assigned in the first step and who placed that school second in their list are considered and the same procedures in the previous step are applied. This is repeated in each step until there are no more students unassigned.

Although the BM algorithm has been widely used, it has some drawbacks. As Abdulkadiroğlu and Sönmez (2003) argues, BM is not strategy-proof in the sense that it may be profitable for some parents to misrepresent their preferences over schools. For instance, by ranking the overdemanded schools lower and relatively safe schools higher, they can guarantee seats at these schools. However, the truth-telling parents are possibly harmed by these misrepresenting parents, as there is the possibility that parents who rank an overdemanded school first and a safer school second do not get assigned to either of them as they lose their seat listed as their second choice to the misrepresenting parents' child. It is reported in the empirical studies of Abdulkadiroğlu et al. (2006) that 19% of parents do not behave strategically and may be harmed by the system. Thus, there is an inherent incentive in BM for some parents to lie about their preferences in order to "game" the system.

Due mostly to the abovementioned drawback of the system, the Boston School Committee changed its school choice mechanism in 2006, which has been used since 1999, from BM to the algorithm proposed by Abdulkadiroğlu and Sönmez (2003), the strategy-proof Student Optimal Stable Mechanism (SOSM). According to the SOSM, firstly, each school considers only the students who chose that school in first place and assigns its seats to these students tentatively based on their preferences on students and rejects remaining ones. In the second step, however, the student set from which schools choose tentative seat-holders are different from the BM algorithm. Each school considers a student set consisting of; the students who are rejected in the first step, students who placed that school second in their list and its tentatively accepted students. According to their preferences, seats are assigned tentatively again and the remaining ones are rejected. This is repeated in each step until there are no more students unassigned.

3.2. Data and Exploratory Analysis

In this study, school-level data from Massachusetts is used. The data come from the statewide reports of Massachusetts Department of Elementary and Secondary Education (hereafter referred to as MDoE). Most of the variables reported by MDoE are at the school-level and remaining variables are reported at district-level. Dataset consists of data on schools, beginning with the 2003-2004 school year until 2017-2018 school year. Throughout this paper, I refer to a school year by its second component. For example, 2006-2007 school year is meant by the year 2007.

As the theory on schooling argues, the inputs of composition of students, structure, resources and general and academic practices/processes are believed to influence the outcome of whether a student graduates or drops out of high school. In this respect and in line with the research aims, statewide reports of MDoE are examined for the necessary variables. Firstly, MDoE defines dropouts as students who leave school prior to graduation for reasons other than transfer to another school. The dropout rate for the schools in Massachusetts for a given year is defined as percentage of students in grades 9-12 who dropped out of school between July 1 and June 30 prior to the listed year and who did not return to school by the following October 1. This means that students who were registered to the school on June 30 of the last school year and are not present once July 1 of a school year is reached are recorded as dropped-out if they do not register by October 1, except for the cases of transferring to another school or graduating. This can be understood more easily via the following example. Suppose that for a school, there are 100 students registered on the 30th of June 2019 who are not graduating in the 2019-2020 school year. Suppose also that, on the 1st of July 2020, 10 of these students are not registered yet and it is known that they did not transfer to another school. Also assume that they do not complete their registration until 1st of October 2020. Then, these 10 students are recorded as dropouts and the dropout rate for the school year 2019-2020 is calculated as 10%. Thus, the definition of dropout rate used by MDoE coincides with the event dropout rate definition of NCES given in the introduction.

MDoE reports several measures for the student composition characteristics of a given school. White enrollment percentage is the percentage of students having origins in any of the original peoples of Europe, the Middle East, or North Africa in the total number of students and gives the racial composition of students. Economically disadvantaged enrollment percentage is percentage of economically disadvantaged students in the total number of students. A student is recorded as an economically disadvantaged student if he/she meets any one of the following criteria;

- The student is eligible for free or reduced-price lunch,
- The student receives Transitional Aid to Families benefits,
- The student is eligible for food stamps.

This measure gives an idea about the student composition in terms of their economic conditions. Another variable reported in a yearly basis is the percent of enrollment of Limited English Proficiency students, which is defined by MDoE as the percent of enrollment who are coded as a student whose first language is a language other than English and who is unable to perform ordinary classroom work in English. I believe these three percentage measures capture most of the student composition part of the inputs of schooling process that influence dropout rates.

As for the school resources, fiscal data such as expenditure per pupil or average teacher salaries are reported at the district level rather than school level, thus could not be used in this thesis. However, student teacher ratio is a good proxy variable in determining how much a school has resources to allocate to schooling. MDoE defines the student teacher ratio as the student enrollment as of October 1 of a school year in relation to the total number of teachers.

As discussed in the theoretical literature, it is the general environment and practices through which above-mentioned inputs are converted to outputs of graduation or dropout. As the general/academic environment in a school is not a directly observable characteristic, I use the average attendance rate at a given school as a proxy to represent overall environment in schools, which would indicate how much the implemented practices motivate students to engage in schooling. MDoE defines the attendance rate as the average percentage of days in attendance for students enrolled with at least 20 days in membership (20 days of being a registered student). Any student who is identified as registered to school for less than 20 days is not taken into account in any calculation MDoE reports.

As mentioned above, statewide reports database of Massachusetts Department of Elementary and Secondary Education is employed. The dataset consists of school-level data on the variables defined above between school years 2003-2004 and 2017-2018. The initial database has 5887 observations. From the initial database, schools with missing values for any one of the above variables for a given year are removed, which amounts to 266 (4.5%) observations. Also, from the remaining 5621 observations, schools that show one any one of the following characteristics for a given year are removed from the dataset for being outliers and not conforming with the great majority of values:

- A school has a dropout rate of above 0.70 (5 observations),
- A school has student teacher ratio of above 100 (5 observations).

Table 1 summarizes the number of schools in the remaining data for each year together with the average dropout rates.

Table 1: Number of Schools and Average Dropout Rate

Year	Number of Schools in Massachusetts	Average Dropout Rate in Massachusetts	Number of Schools in Boston	Average Dropout Rate in Boston
2004	333	4.36	29	10.27
2005	342	4.35	29	8.78
2006	357	4.66	37	11.54
2007	359	5.38	38	11.28
2008	373	4.95	39	8.80
2009	371	4.49	39	8.26
2010	375	4.82	40	9.39
2011	375	4.31	40	8.45
2012	379	4.51	34	8.62
2013	382	4.18	36	8.81
2014	391	3.99	37	6.66
2015	395	4.02	36	6.68
2016	391	4.26	35	7.05
2017	389	3.71	37	6.37
2018	397	3.89	37	7.17

Table 2, on the other hand, gives the descriptive statistics on all the variables in the dataset.

Table 2: Descriptive Statistics (Pooled data of 2004-2018)

	Mean	**Std. Dev.**	**Min**	**Max**
Dropout Rate (%)	4.39	8.45	0.00	66.70
Attendance Rate (%)	91.98	5.99	43.30	100.00
Student Teacher Ratio	12.72	3.87	1.30	76.30
White Enrollment Rate (%)	67.33	31.63	0.00	100.00
Economically Disadvantaged Enrollment Rate (%)	31.86	25.44	0.00	100.00
Limited English Proficiency Enrollment Rate (%)	4.69	9.85	0.00	100.00

Figure 3 plots the average dropout rates in Massachusetts given in Table 1 to see how it behaved over time. It can be inferred from the figure that overall, in Massachusetts, average dropout rate increases sharply from 2004 to 2007 and then decreases until 2017 around a trend with fluctuations.

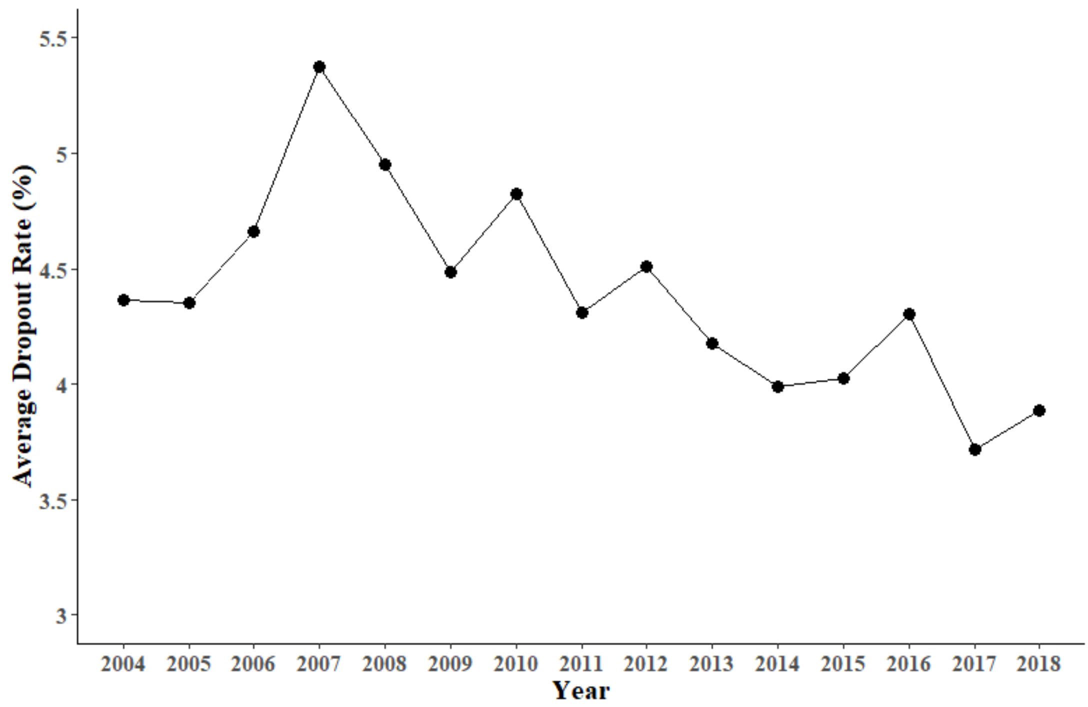


Figure 3: Average Dropout Rates in Massachusetts

Note: Dropout rates are calculated annually according to the definition given above. Source: Own Calculations based on Massachusetts Department of Elementary and Secondary Education, Statewide Reports

Average dropout rates in Boston between 2004-2018 is given in Figure 4. To understand what drives this behavior in dropout rates, there is a need to go into detail. Since the dataset contains all districts and all cities in Massachusetts, it would be troublesome to group the schools based on all the districts or all the cities. Instead, I naively consider the ten biggest cities in Massachusetts and consider the other 47 cities as ‘Massachusetts-Other’. This grouping comes with two benefits. First, here in this exploratory analysis, it helps us understand the data more clearly as grouping the data can point out how dropout rates differ based on each city and how they behave in terms of cities. Second, it helps us in the later part of this paper where I will look for the cities with similar trends to apply difference-in-difference analysis, which is to be detailed below.

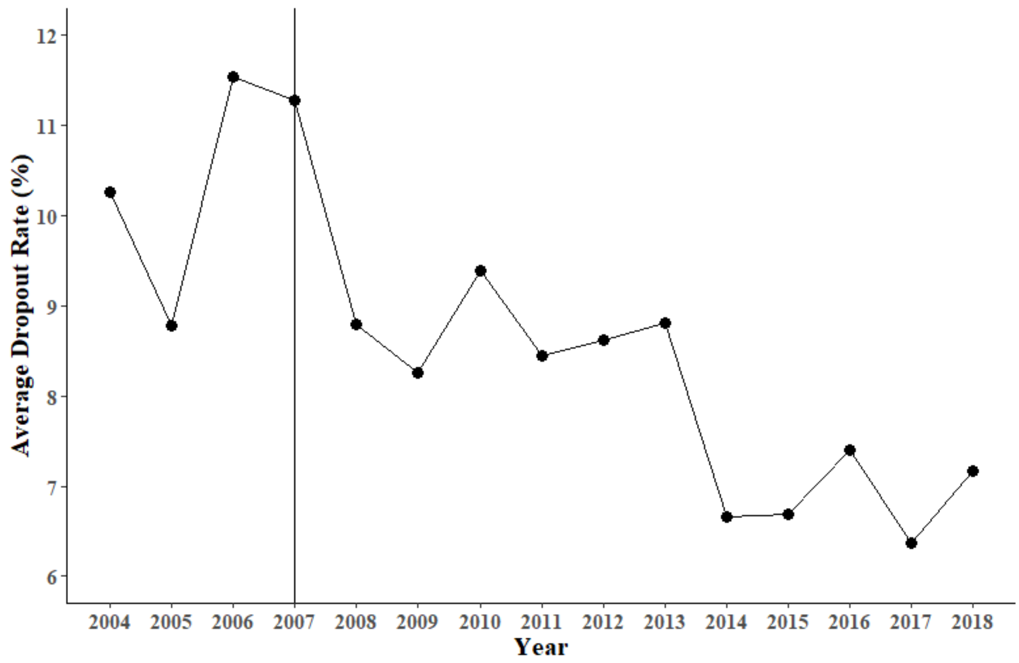


Figure 4: Average Dropout Rates in Boston

Note: Vertical line indicates the year with which new school choice mechanism has been used. Before the school year 2006-2007, Boston Mechanism has been used and starting with the school year 2006-2007, Student Optimal Stable Mechanism has been implemented in Boston high schools. Dropout rates are calculated annually according to the definition given above.
Source: Own Calculations based on Massachusetts Department of Elementary and Secondary Education, Statewide Reports

In this respect, the dataset is grouped into ten biggest cities in Massachusetts, which are Boston, New Bedford, Quincy, Worcester, Brockton, Lowell, Fall River, Springfield, Cambridge and Lynn. The districts that are located through other cities are grouped into a city called Massachusetts-Other. I give the average dropout rates for the cities other than Boston in Figures 5-14. The average dropout rates used in these figures come from own calculations based on Massachusetts Department of Elementary and Secondary Education, Statewide Reports, which are calculated annually according to the definition given above.

In the cities of Boston, Massachusetts-Other, New Bedford, Quincy and Worcester, the average dropout rates seem to be following a similar pattern. For these cities for the time period between years 2004-2018, average dropout rates seem to be first decreasing, then increasing to a peak around years 2006 and 2007 and decreasing to their lowest point around years 2017 and 2018.

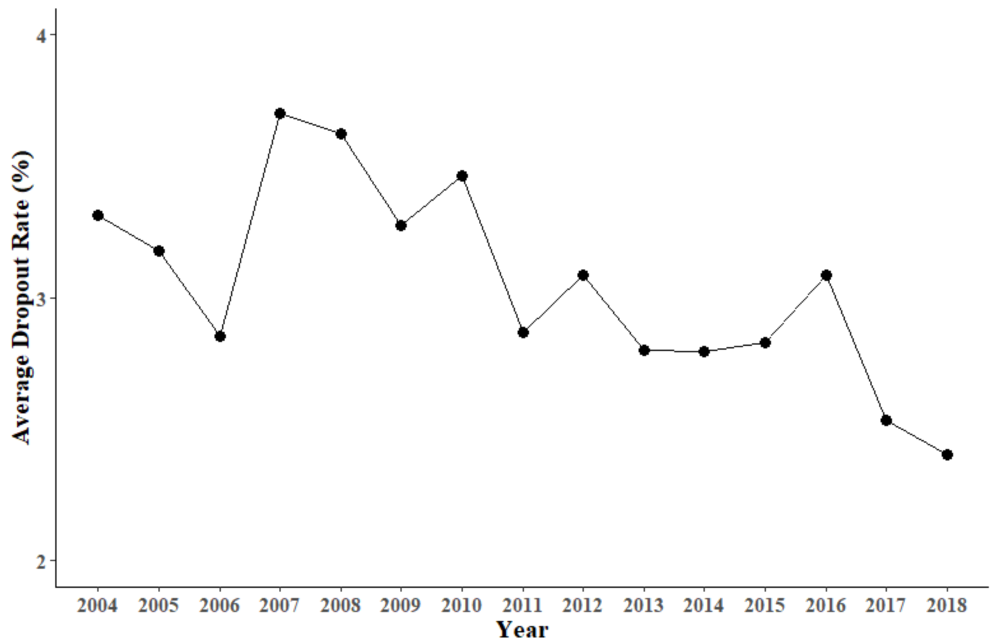


Figure 5: Average Dropout Rates in Massachusetts-Other

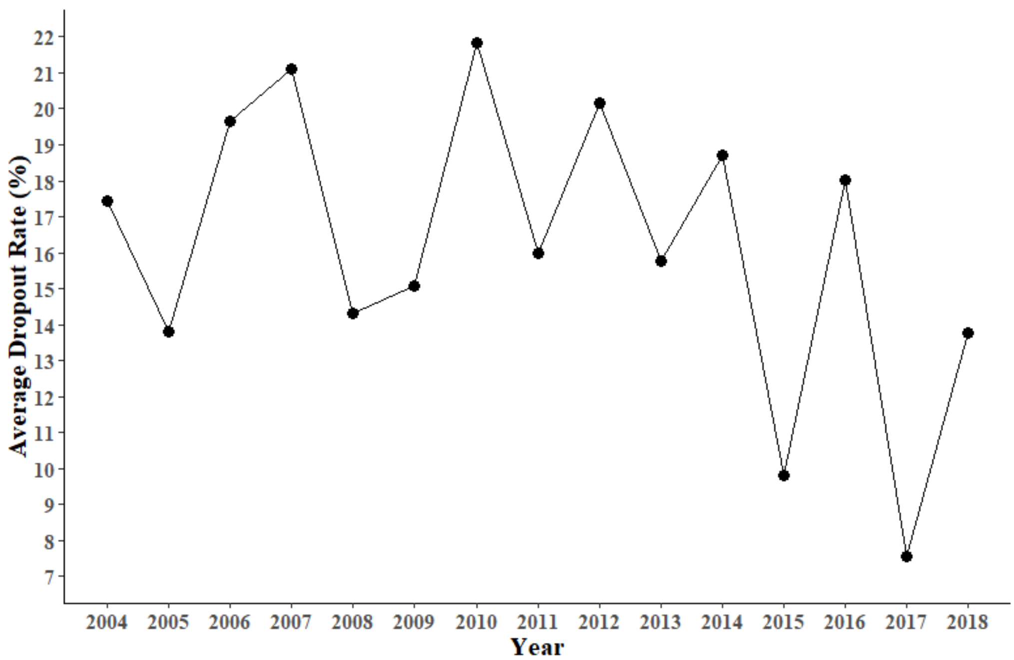


Figure 6: Average Dropout Rates in New Bedford

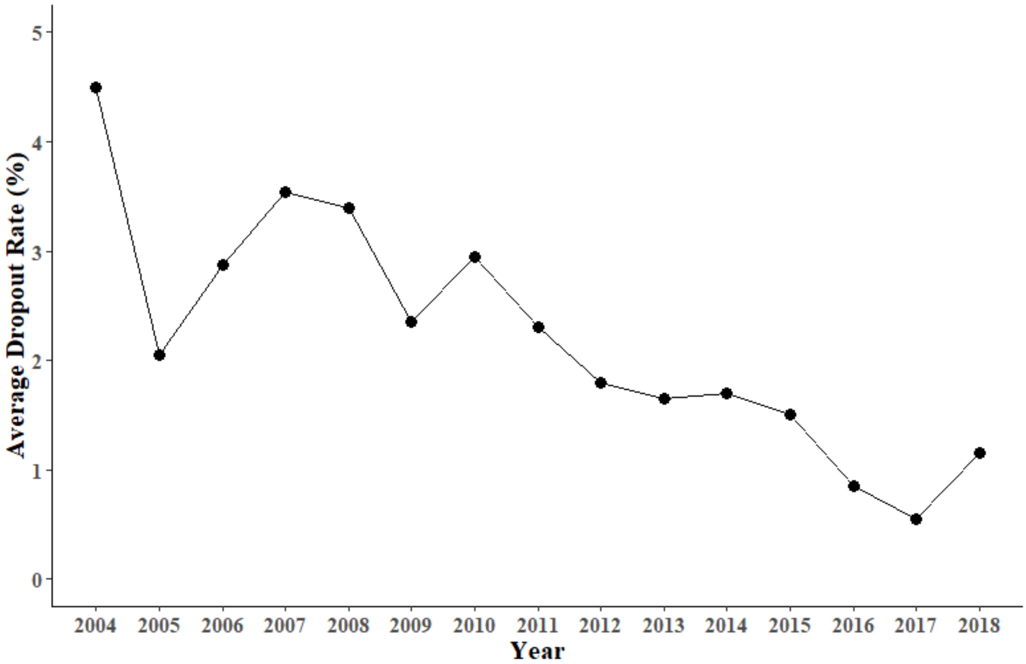


Figure 7: Average Dropout Rates in Quincy

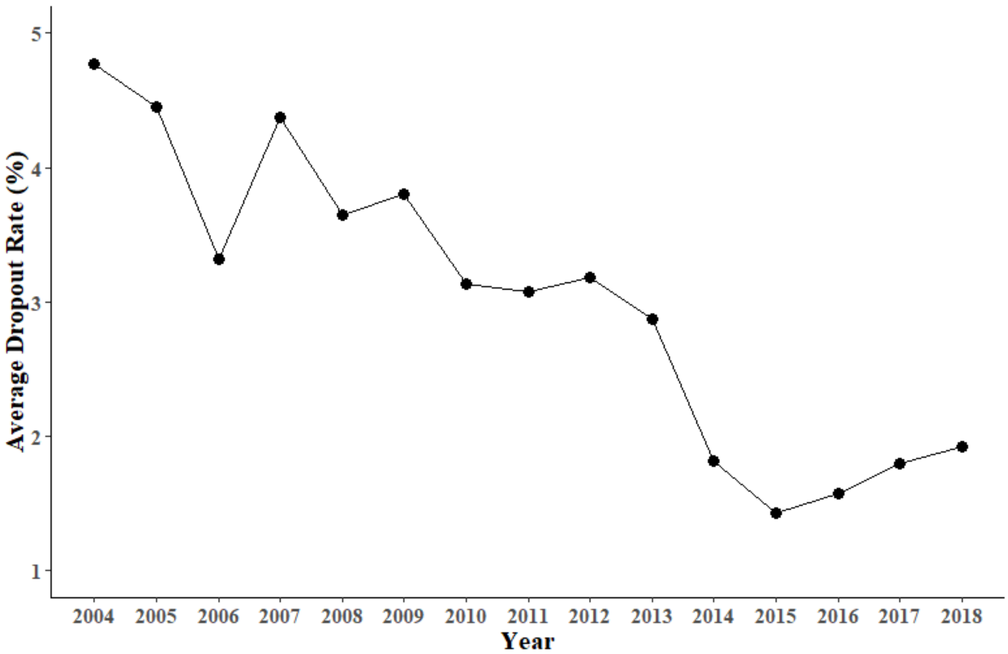


Figure 8: Average Dropout Rates in Worcester

Average dropout rates in the cities of Brockton, Lowell and Fall River, on the other hand, seem to be following a trend that is different from those observed for the cities Boston, Massachusetts-Other, New Bedford, Quincy and Worcester during these years. For Brockton and Lowell, rates first increase until the years 2005 for the second and 2006 for the first one. Then, for both of these cities, the rates decrease and then increase to make a peak in 2011. After 2011, rates seem to be declining for both cities, although there is a sharp increase in 2018 for Lowell. For Fall River, however, the dropout rates seem, on the average, to be much higher compared to the other cities in Massachusetts, well above the state averages.

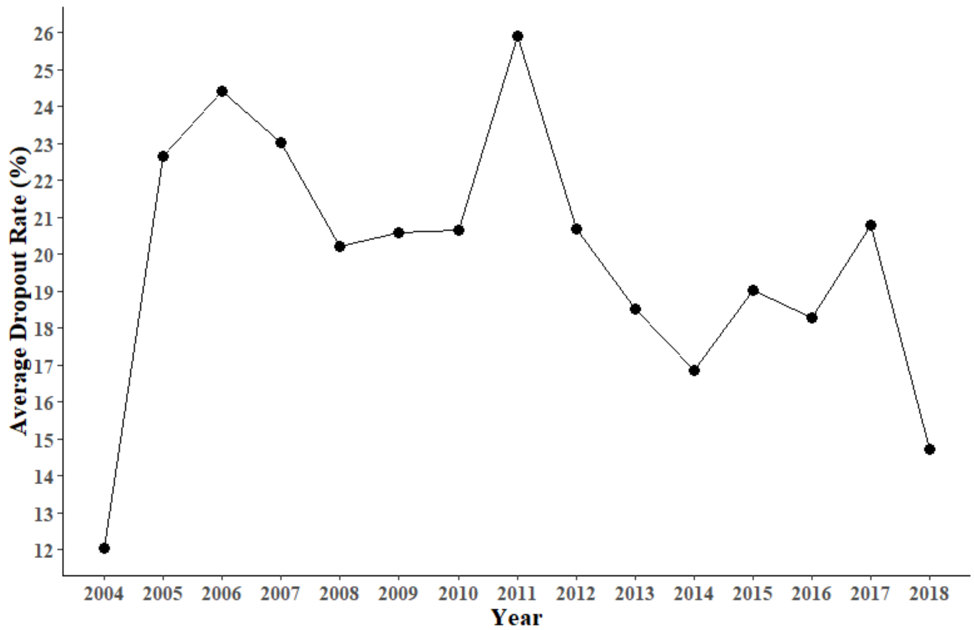


Figure 9: Average Dropout Rates in Brockton

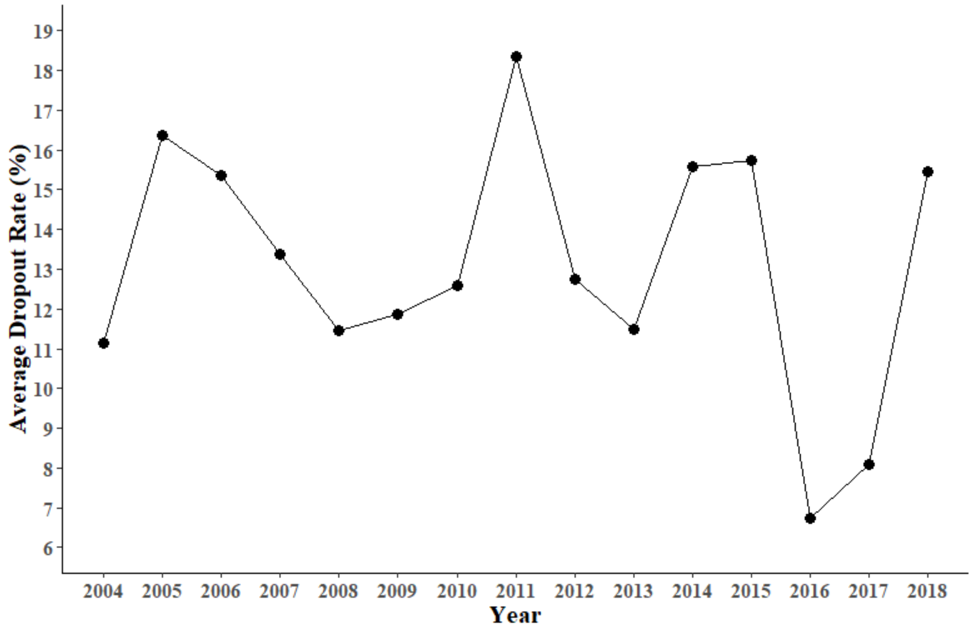


Figure 10: Average Dropout Rates in Lowell

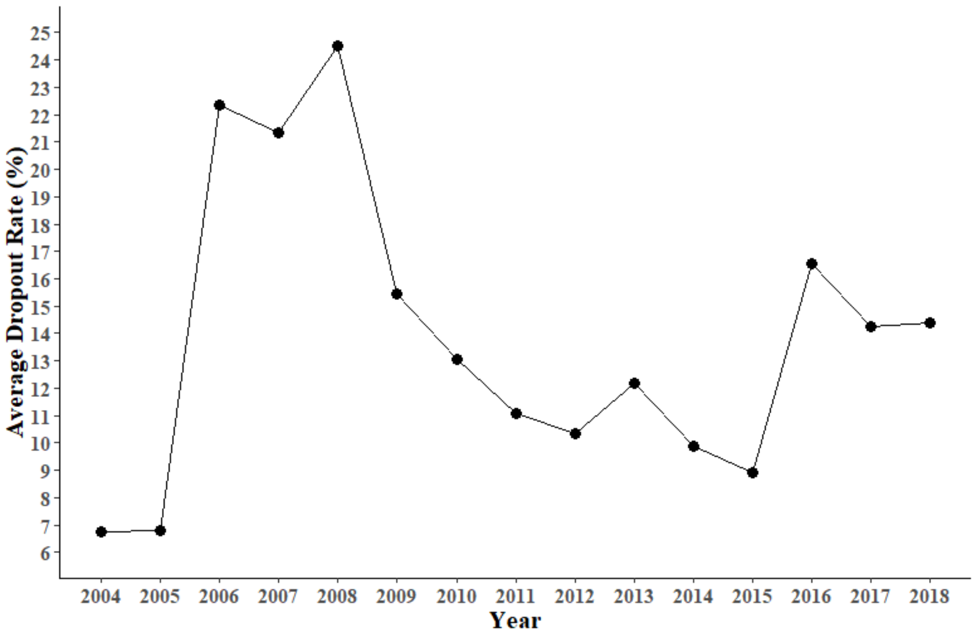


Figure 11: Average Dropout Rates in Fall River

Springfield seems to have a general upward trend starting from 2004. Although the rates show a steady decline between 2007 and 2009, a generally higher average dropout rate is observed in the following years.

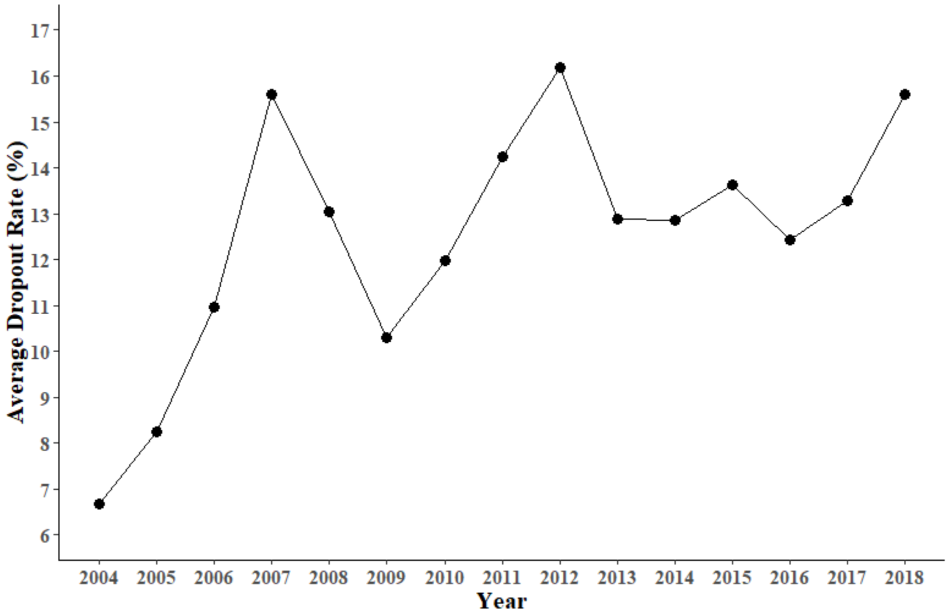


Figure 12: Average Dropout Rates in Springfield

On the contrary, Cambridge seems to enjoy an overall downward trend in average dropout rates starting from 2004. For this city, the average seems to have increased only for the year 2010 and apart from that, a generally lower average dropout rate is observed in recent years.

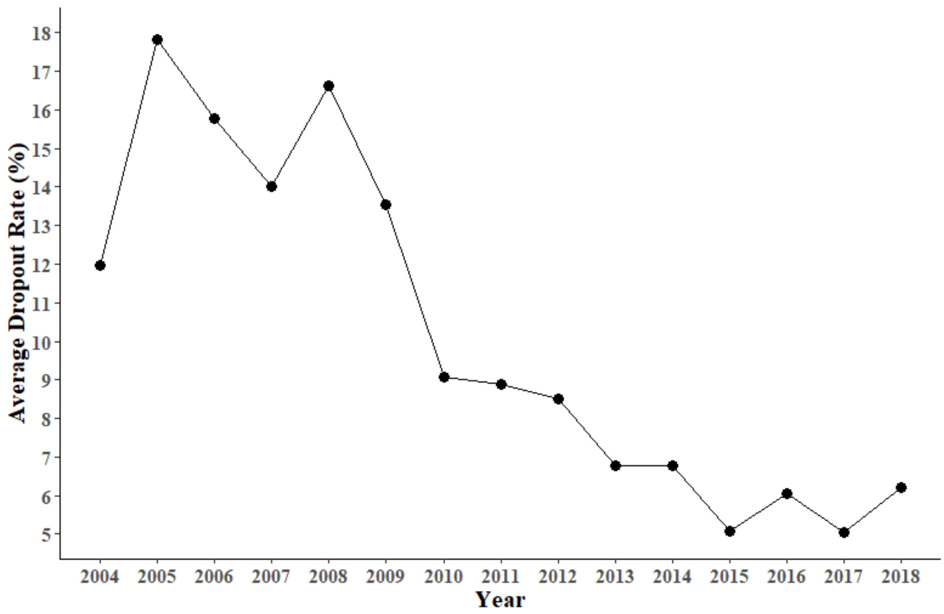


Figure 13: Average Dropout Rates in Cambridge

Average dropout rates in Lynn seem to be following a similar pattern to Cambridge. Although starting at much higher averages, the average dropout rate seems to be following a decreasing trend, approaching the state average.

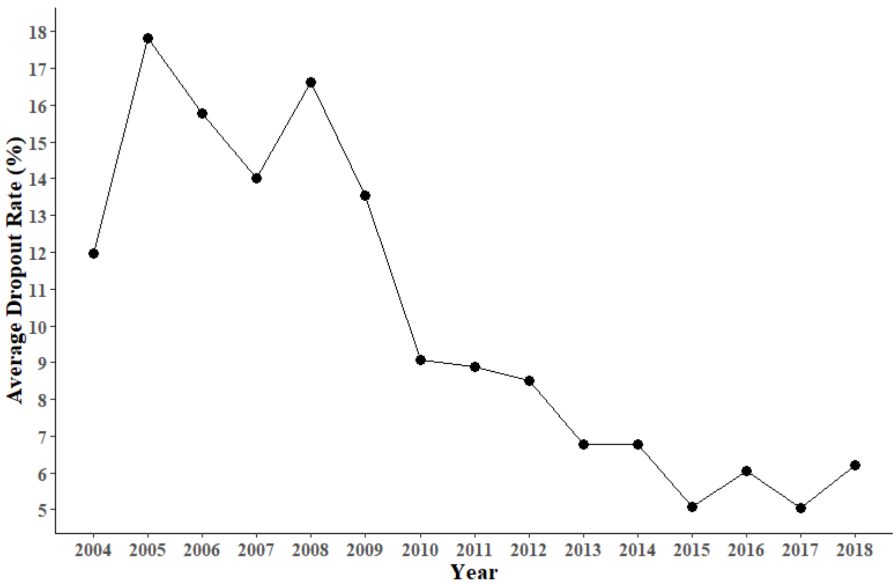


Figure 14: Average Dropout Rates in Lynn

Figure 15 plots the relationship between attendance rates and dropout rates with a linearly fitted line. It can be observed that there is a strong negative correlation between attendance rate and dropout rate of a school. The fitted line suggests that schools with higher attendance rates tend to have significantly lower dropout rates. This seems to be in line with the literature in that, being an overall proxy for the general academic/social environment of a school, a higher attendance rate would signal lower dropout rate.

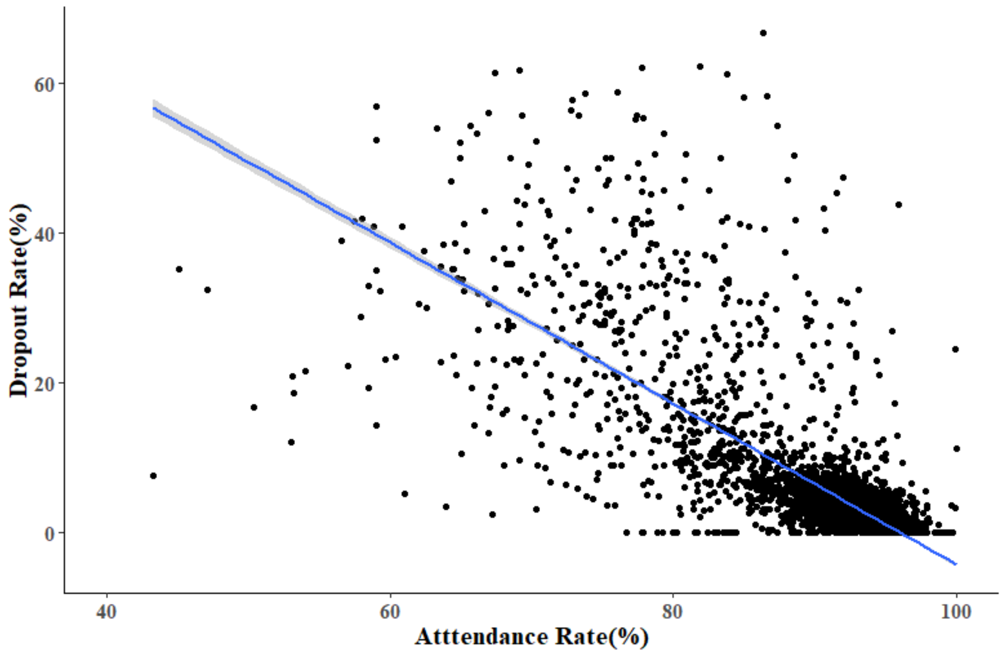


Figure 15: Relationship between Dropout Rate and Attendance Rate

Note: Dropout rates and attendance rates are calculated annually according to the definitions given above. Each point represents the values of relevant data pair for one school in the dataset for a year between 2004-2018 and the blue line represents the fitted line with 95 percent confidence interval around it.
Source: Own Calculations based on Massachusetts Department of Elementary and Secondary Education, Statewide Reports

Figures 16 to 18 plot the relationship between the three variables about the student composition of schools and dropout rates. Figure 16 suggests that there is a positive relationship between percentage of economically disadvantaged students of a school with the dropout rate. This also supports the literature that as economic background of the students in a school worsens, one would expect higher dropout rates.

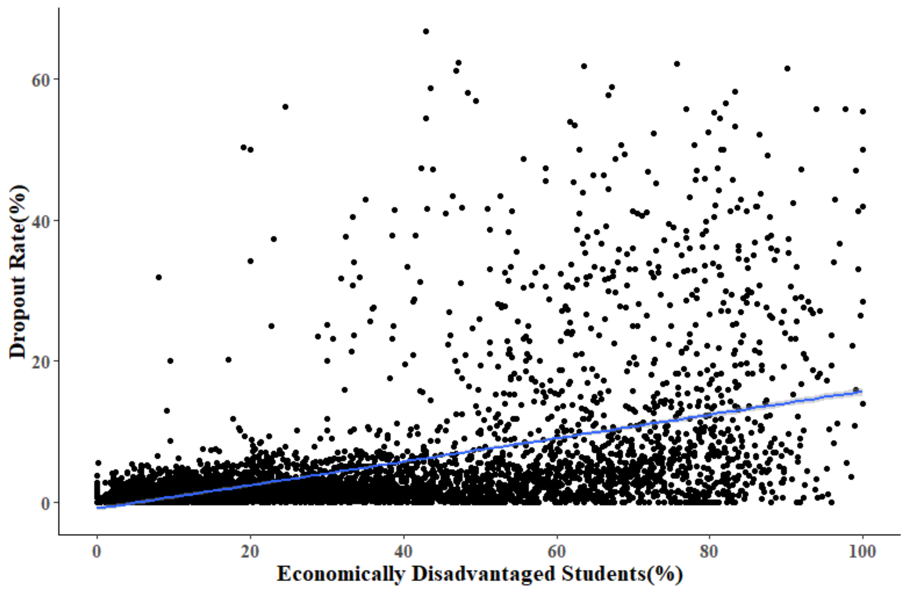


Figure 16: Relationship between Dropout Rate and Percentage of Economically Disadvantaged Students in Student Population

Note: Dropout rates and economically disadvantaged student rates are calculated annually according to the definitions given above. Each point represents the values of relevant data pair for one school in the dataset for a year between 2004-2018 and the blue line represents the fitted line with 95 percent confidence interval around it.
Source: Own Calculations based on Massachusetts Department of Elementary and Secondary Education, Statewide Reports

The same argument is valid for the percentage of limited English proficiency students. Figure 17 suggests that as the number of students who have limited English capabilities gets higher compared to the school population, which would make the teaching process more challenging, the dropout rates are expected to be higher.

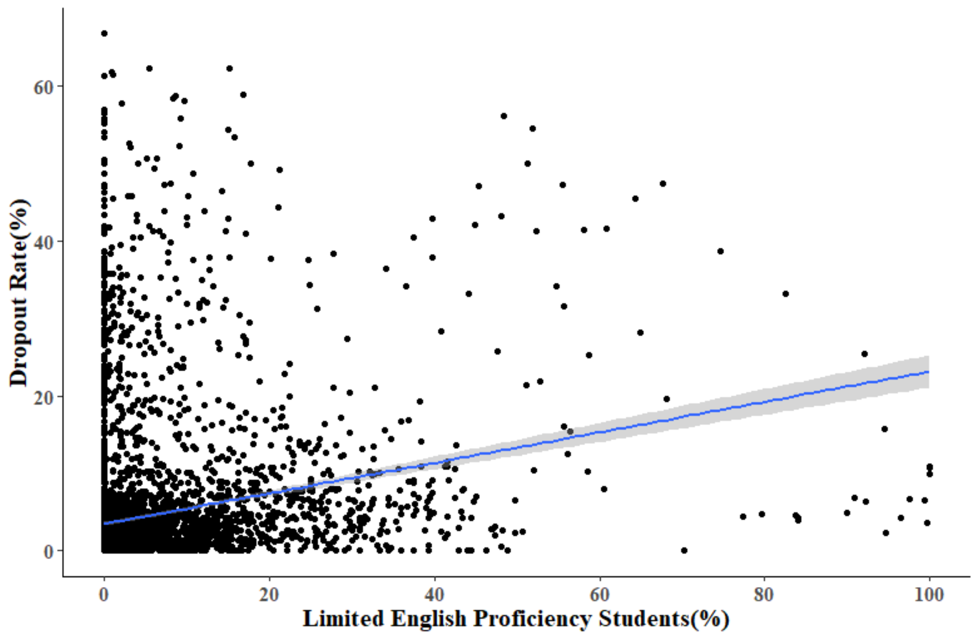


Figure 17: Relationship between Dropout Rate and Percentage of Limited English Proficiency Students in Student Population

Note: Dropout rates and limited English proficiency student rates are calculated annually according to the definitions given above. Each point represents the values of relevant data pair for one school in the dataset for a year between 2004-2018 and the blue line represents the fitted line with 95 percent confidence interval around it.
Source: Own Calculations based on Massachusetts Department of Elementary and Secondary Education, Statewide Reports

Figure 18, on the other hand, gives an idea about how dropout rates and percentage of white students in a school are related. According to the figure, when the racial composition of students in a school is higher in favor of white students, it can be seen that the dropout rates are lower.

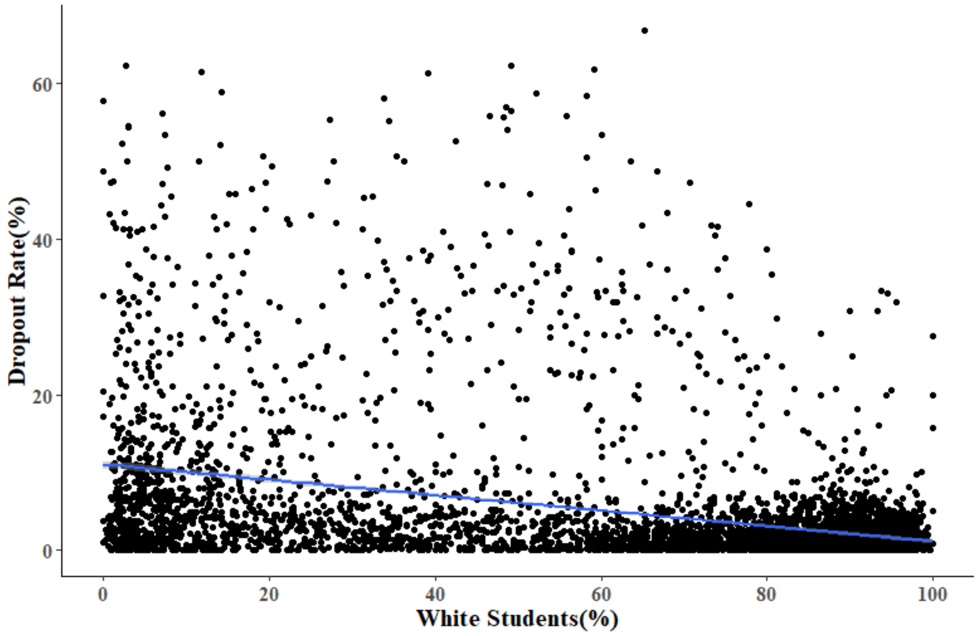


Figure 18: Relationship between Dropout Rate and Percentage of White Students in Student Population

Note: Dropout rates and white student rates are calculated annually according to the definitions given above. Each point represents the values of relevant data pair for one school in the dataset for a year between 2004-2018 and the blue line represents the fitted line with 95 percent confidence interval around it.
Source: Own Calculations based on Massachusetts Department of Elementary and Secondary Education, Statewide Reports

Figure 19 plots the relationship between student teacher ratio and dropout rates. Although the literature suggests that there is a positive relationship between dropout rates and student teacher ratio, which is thought as a proxy for school resources, the figure suggests a negative relationship. This may have to do with other organizational and structural characteristics of high schools affecting the dropout rates, as pointed out in Lee and Burkham (2003).

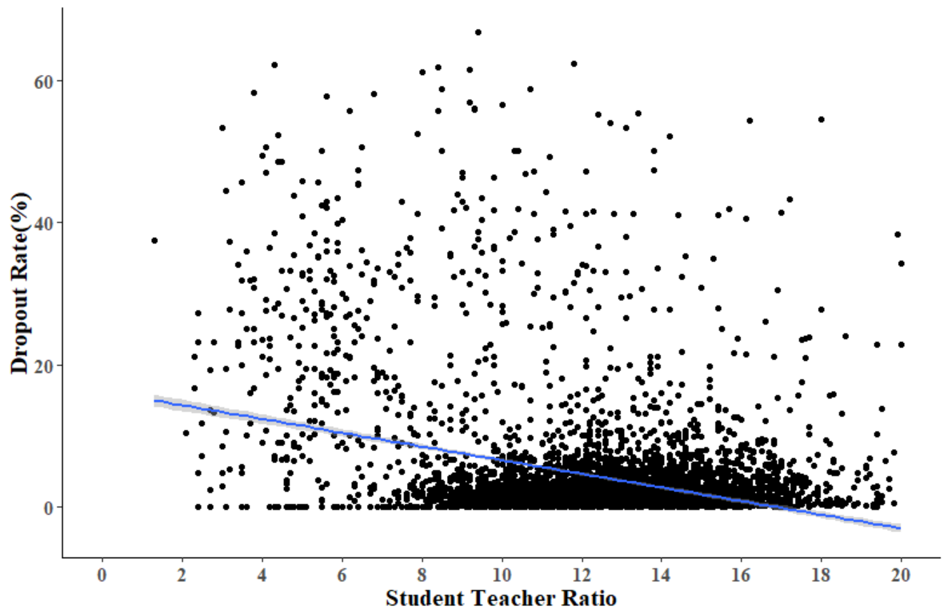


Figure 19: Relationship between Dropout Rate and Student Teacher Ratio

Note: Dropout rates and white student rates are calculated annually according to the definitions given above. Each point represents the values of relevant data pair for one school in the dataset for a year between 2004-2018 and the blue line represents the fitted line with 95 percent confidence interval around it.
Source: Own Calculations based on Massachusetts Department of Elementary and Secondary Education, Statewide Reports

It is important to point out that these graphs merely serve the purpose of exploratory data analysis, as they do not take into account the effect of between schools and across year differences on dropout rates. Following the figures explaining the relationship between several explanatory variables and dropout rates, Table 3 presents the correlation matrix for the variables defined above. When the table is examined, there is nothing whose absolute value is above 0.8. Thus, strict relation between any two variables is not present according to the data.

Table 3: Correlation Matrix for the Variables (Pooled data of 2004-2018)

	Dropout Rate	Attend. Rate	Student Teacher Ratio	White Enr.	Econ. Disadv. Enr.	Lim. Eng. Proficiency Enr.
Dropout Rate	1	-0.754	-0.163	-0.370	0.501	0.230
Attend. Rate		1	0.221	0.464	-0.589	-0.289
Student Teacher Ratio			1	0.071	-0.195	-0.041
White Enr.				1	-0.715	-0.632
Econ. Disadv. Enr.					1	0.493
Lim. Eng. Proficiency Enr.						1

3.3. Single Year Regression Models

As mentioned, there are two aims in this study. First is to understand how a set of school factors are correlated with and helpful in predicting the dropout rate of a high school. The second one is to understand whether and to what extent the change in the mechanism Boston Public Schools made from Boston Mechanism to Student Optimal Stable Mechanism affected the dropout rates in Boston High Schools. In this part, I give the methodology and models employed in this thesis towards these aims.

For the aim of understanding the relationship between the variables and dropout rates, I first ignore the differences across time and try to understand how these variables affect dropout rate by focusing on 2018 data. I apply OLS regression to 2018 data and analyze the regression output trying to see how well given variables explain the variation in dropout rates, how the signs of coefficients of these variables coincide with expectations. The same regression is also done for 2004 data to serve as a robustness check. For the OLS regression, the model that is estimated takes the form;

$$DRP_j = \beta_0 + \beta_1 ATT_j + \beta_2 STR_j + \beta_3 WHI_j + \beta_4 ECO_j + \beta_5 LEP_j + \beta_6 COUNTY_j + u_j \tag{3.1}$$

and the variables in the model are;

- DRP_j: Dropout rate for school j (%),
- ATT_j: Attendance rate for school j (%),
- STR_j: Student teacher ratio for school j,
- WHI_j: White enrollment rate for school j (%),
- ECO_j: Economically disadvantaged enrollment rate for school j (%),
- LEP_j: Limited English proficiency enrollment rate for school j (%),
- $COUNTY_j$: County for school j.

In this setting, COUNTY is a categorical variable indicating the county of the school. This dummy variable is included in the model to control for the local labor market effects.

Figure 20 plots the residuals against fitted values. When the figure is examined, one suspects that the model is suffering from heteroscedasticity problem. I thus report the White's heteroscedasticity robust standard errors in the next section.

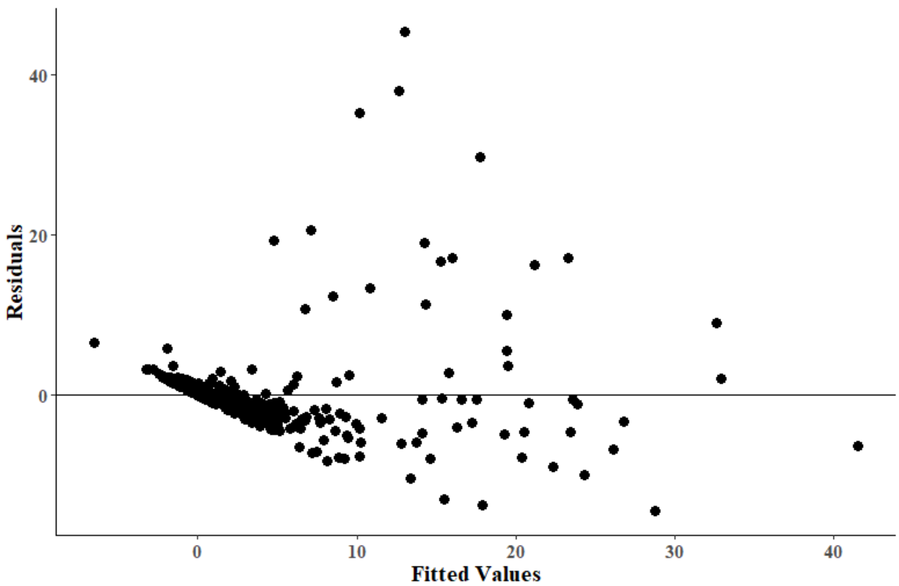


Figure 20: Residuals vs Fitted Values, 2018

3.4. Multi-Year Panel Data Model

Although single year OLS model may give a pretty good intuition about the dropout rates and its predictors, it is prone to have some limitations. Single year setting does not account for the heterogeneity across different schools as it does not fully capture the school fixed effects. Moreover, single year models may suffer from omitted variable bias, as there are a lot of factors that would affect the annual dropout rate of a school. Due to these limitations, I extend the single year regression model to a panel data regression model between years 2004 and 2018.

The dataset can be described as an unbalanced, short panel data. It is natural for the panel data to be unbalanced in the context of this thesis as new schools may be opened while some schools may be closed or merged with other schools in different years. Dataset consists of an unbalanced panel with n = 453 total number of schools and T = [2004, 2018] number of time periods as years for a total of N = 5,611 observations. As developed in Ahrens and Pincus (1981), there are two measures of unbalancedness, namely γ and ν, and they are equal to one for the balanced data. The smaller they are, the bigger the impact of unbalancedness on estimators. Table 4 reports the values for the model, which will be detailed in the next paragraph. As seen in Table 4, it is arguably safe to continue the analysis despite unbalancedness.

Table 4: Ahrens and Pincus (1981) Measures of Unbiasedness

γ	ν
0.696	0.894

The following equation describes the basic linear panel data model that one can use;

$$DRP_{jt} = \alpha_{jt} + \beta_{1jt}ATT_{jt} + \beta_{2jt}STR_{jt} + \beta_{3jt}WHI_{jt} + \beta_{4jt}ECO_{jt} + \beta_{5jt}LEP_{jt} + u_{jt} \tag{3.2}$$

where $j = 1, \dots, n$ is the school index, $t = 1, \dots, T$ is the time index and u_{jt} is the random disturbance term, which is assumed to have mean 0. Some assumptions are made about errors and exogeneity of regressors and parameters, which turns the model into different panel data models. Most common assumption is that the parameters are constant across all units and time;

$$\alpha_{jt} = \alpha,\ \beta_{kjt} = \beta_k\ \forall j \in [1, n], t \in [1, T], k \in [1,5] \tag{3.3}$$

which turns Equation 3.2 into;

$$\begin{aligned} DRP_{jt} &= \alpha + \beta_1 ATT_{jt} + \beta_2 STR_{jt} + \beta_3 WHI_{jt} + \beta_4 ECO_{jt} \\ &+ \beta_5 LEP_{jt} + u_{jt} \end{aligned} \tag{3.4}$$

It is obvious that, Equation 3.4 is no different from the single year model given in Equation 3.1. Thus, this is known as Pooled model, since all the data are pooled across schools and time periods. As indicated earlier, this model does not take the heterogeneities among the schools and across time into account. To model these heterogeneities, one often makes the assumption that the error term has three separate parts, one specific to individual and does not change over time, one specific to time period and does not change across individuals and an idiosyncratic error component as in;

$$u_{jt} = \mu_j + \lambda_t + \varepsilon_{jt} \tag{3.5}$$

which turns Equation 3.4 into;

$$\begin{aligned} DRP_{jt} &= \beta_j + \lambda_t + \beta_1 ATT_{jt} + \beta_2 STR_{jt} + \beta_3 WHI_{jt} + \beta_4 ECO_{jt} \\ &+ \beta_5 LEP_{jt} + \varepsilon_{jt} \end{aligned} \tag{3.6}$$

where,

$$\beta_j = \alpha + \mu_j \tag{3.7}$$

Thus, it can be seen that in the final model, one has the notion that each of the schools and time periods are different with their own fixed effects, reflected through the respective intercept terms.

If one assumes that the individual specific effects are random variables which are uncorrelated with the regressors, the model is called the Random Effects Model (RE), otherwise the model is called the Fixed Effects Model (FE). In this study, I model the dropout rates of schools as a fixed effects model, since I do not expect the individual effects of schools to be uncorrelated with the regressors. However, one needs to convince himself of this choice more technically through formal testing procedures. I first test the poolability, i.e., the validity of assumptions given in Equation 3.3. Formally, it is written as;

$$H_0 : \mu_j = 0 \; \forall \, j \in [1, n] \tag{3.8}$$

Note that this is a standard F test, with F-statistic given by;

$$F = \frac{(\mathrm{ESS}_R - \mathrm{ESS}_U)/(n-1)}{\mathrm{ESS}_U/(nT - n - K)} \tag{3.9}$$

where ESS_R is the residual sum of squares of the pooled model and ESS_U is the residual sum of squares of the FE model. Table 5 presents the result of the test, implying that there are strong individual fixed effects to be taken into account, rejecting the null of Pooled OLS model.

Table 5: Poolability Test Results

F	**df 1**	**df 2**	**p-value**
22.582	466	5139	<0.001

There is also a need to test for fixed effects vs. random effects model, i.e., the assumption for the random effects model that the individual effects are exogenous to the regressors, by the means of Hausman-type specification tests. I estimate both the fixed and random effects functions and test the null hypothesis that the individual specific effects are exogenous. From the test results that are presented in Table 6, the null hypothesis is strongly rejected. Thus, I find that the random effects model is inconsistent, justifying the model choice of fixed effects.

Table 6: Hausman Test Results

χ^2	df	p-value
22.582	466	<0.001

After specifying the model choice, I test the model with three objectives in mind. First is, I expect there to be cross-sectional dependence, as schools in the same district can have common factors affecting their respective dropout rates. Secondly, since I now have multi-year observations for schools, one should expect serial correlation to be present. Lastly, one needs to be alert to heteroscedasticity. Thus, one needs to show caution in the inference and standard errors about the individual coefficients.

One of the most common tests in the literature on cross-sectional dependence is Pesaran's cross-sectional dependence (Pesaran's CD) test, originating from Pesaran (2004), with null hypothesis of residuals across schools are not correlated. The test has good properties for any n-T (number of individuals, number of time periods) consideration and robust to a variety of settings, providing that the model does not contain time-specific dummies (Croissant et al. (2019)). The results of Pesaran's CD test for model is given in Table 7, from which one concludes that the data contains significant correlations among schools.

Table 7: Pesaran's CD Test Results

z	p-value
57.891	<0.001

For the serial correlation test, I use Wooldridge's test for serial correlation in FE panels, which can be applied under general not-restrictive assumptions for short panels (Croissant et al. (2019), Wooldridge (2002)). The results presented in Table 8 clearly shows that dataset suffers from serial autocorrelation.

Table 8: Wooldridge Serial Correlation Test Results

χ^2	p-value
86.108	<0.001

Through these tests, it can be understood that the inferences that will be made about the coefficients will be erroneous, due to the observed heteroscedasticity, serial correlation and cross-sectional dependence observed in the data. I thus use standard errors that are robust to these effects. In the literature, it is pointed out that spatial correlation consistent (SCC) covariance estimators, originated from Driscoll and Kraay (1998), are robust against heteroscedasticity, autocorrelation and cross-sectionally dependent data (Hoechle (2007)). Thus, the standard errors from SCC estimators are reported in the next section.

In addition to the fixed effects panel data estimation methodology, a dynamic panel data model can also be considered. For a dynamic panel data regression model, the dynamics are introduced via including a lag of the dependent variable into the regressors. By this inclusion, one makes the entire history of the dependent variable a regressor and the interpretation of the other independent variables changes. In dynamic panel data setting, the coefficients of the independent variables indicate the correlations of each of these variables conditional on the history of the dependent variable. This approach is particularly suited if it is the case that past values of the dependent variable would affect its present value.

Within the context of this thesis, I believe that past values of the dropout rates for a given school may affect the dropout rates of a given year. However, this effect is already accounted for in the fixed effects model. Nonetheless, it may be of value to also consider a dynamic panel data model of the form;

$$DRP_{jt} = \beta_j + \lambda_t + \gamma DRP_{jt-1} + \beta_1 ATT_{jt} + \beta_2 STR_{jt} + \beta_3 WHI_{jt} + \beta_4 ECO_{jt} + \beta_5 LEP_{jt} + \varepsilon_{jt} \quad (3.10)$$

where γ is the coefficient of the lagged dependent variable. The inclusion of the lagged dependent variable on the right-hand side of regression equation, however, violates exogeneity assumption in that now there is a regressor that is most likely to be correlated with the error term. Due to this, estimation of such a model is done by a Generalized Method of Moment (GMM) approach, proposed by Arellano and Bond (1991). I thus present the GMM estimation results of the panel data model given in 3.10 in the next section as well.

3.5. Difference in Differences

As it can be seen from Figure 4, average dropout rates in Boston is around 11% for the year 2004. After decreasing to about 10% in 2005, it sharply increases to around 12.5% for 2006. Then it slightly reduces to 12% in 2007 and a massive decrease to around 9% is observed for the year 2008. Finally, another slight decrease to around %8.5 is seen for 2009. State average dropout rates can be seen in Figure 2. In comparing Boston to the state averages, there are two striking differences. First is that in Boston, average dropout rates are way above the state averages between 2004 and 2009. Secondly, immediately after Boston changes its school choice mechanism, a slight decrease occurs while state average sharply increases for the year 2007. Apart from these differences, averages in Boston seems to be moving together with the state averages.

As mentioned, the mechanism change in Boston occurred in 2006 (school year 2005-2006) and introduced to the system starting with 2007 (school year 2006-2007), while the other districts in Massachusetts remained with the Boston Mechanism. Thus, for the aim of understanding the effects of the school choice mechanism change in Boston, I make use of the difference in difference methodology for the years 2006 and 2007. The main motivation for the difference-in-differences model given below is to understand whether the decrease can be attributed to the mechanism change.

Since the change in school choice mechanism took place in 2006, school year 2005-2006 will be denoted as before-treatment year and school year 2006-2007 will be denoted as after-treatment year. Also, the schools which are in Boston and adopted the new mechanism are denoted as treatment group whereas the other schools are control group. Suppose the four averages of the dropout rates are denoted as;

- $\overline{D}_{T,B}$: Average dropout rate of the treatment group for the before-treatment year,
- $\overline{D}_{C,B}$: Average dropout rate of the control group for the before-treatment year,
- $\overline{D}_{T,A}$: Average dropout rate of the treatment group for the after-treatment year,
- $\overline{D}_{C,A}$: Average dropout rate of the control group for the after-treatment year.

Then, the difference-in-difference estimator $\hat{\delta}$ is defined as;

$$\hat{\delta} = \left(\overline{D}_{T,A} - \overline{D}_{C,A}\right) - \left(\overline{D}_{T,B} - \overline{D}_{C,B}\right) \tag{3.11}$$

$\hat{\delta}$ is the difference between (i) the difference in the means of the treatment (Boston Schools) and control (Other Schools) groups in the response variable (dropout rate) after the treatment (change of the school choice mechanism), and (ii) the difference in the means of the treatment (Boston Schools) and control (Other Schools) groups in the response variable (dropout rate) before the treatment (change of the school choice mechanism), thus named as difference-in-differences (DID) estimator. Basically, it measures the effects of a treatment comparing the before and after treatment differences in the outcome of a treatment and a control group. Suppose one has the following equation;

$$\begin{aligned} DRP_{it} = \beta X_{it} + \beta_1 BOSTON_i + \beta_2 2007_t + \\ \beta_3 (BOSTON_i \times 2007_t) + \varepsilon_{it} \end{aligned} \tag{3.12}$$

where DRP_{it} is the dropout rate of school i in time t, X_{it} is the matrix of Attendance Rate, Student Teacher Ratio, Percentage White Enrollment, Percentage Economically Disadvantaged Enrollment and Percentage Limited English Proficiency Enrollment variables for school i in time t, $BOSTON_i$ is a binary variable which equals 1 if school i is in the treatment group (Boston schools) and 2007_t is a binary variable which equals 1 if time t is in the after-treatment year (2007). As outlined in Table 9, DID estimator is the $\widehat{\beta_3}$.

Table 9: Difference in Difference Coefficients

	Before Treatment	After Treatment	Difference
Treatment Group	$\beta+\beta_1$	$\beta+\beta_1+\beta_2+\beta_3$	$\beta_2+\beta_3$
Control Group	β	$\beta+\beta_2$	β_2
Difference	β_1	$\beta_1+\beta_3$	β_3

Along with the usual OLS assumptions, the most important assumption that needs to be satisfied is the so-called parallel trend assumption. Parallel trend assumption basically states that the difference between the control group and the treatment group would be the same if there had not been a treatment, which means the two groups would have had parallel trends if one group had not been exposed to a treatment. Thus, I look for districts that would satisfy the parallel trend assumption to be included in the control group. The districts that satisfy the parallel trend assumption is presented alongside the analysis results in the next section.

CHAPTER 4

RESULTS

In line with the purpose of this study, this chapter is divided into three parts. In the first part, I examine the dropout rates in a single year context, for the year 2018 and also for the sake of robustness, for the year 2004. In the second part, I use panel data estimation to control for time and school fixed effects, for the years between 2004 and 2018. Finally, in the third part, I perform DID analysis.

4.1. Single Year Regression Model for Dropout Rates

Descriptive statistics for 2018 data are given in Table 10.

Table 10: Descriptive Statistics (2018 Data)

	Mean	Std.Dev.	Min	Max
Dropout Rate (%)	3.889	8.347	0.000	58.300
Attendance Rate (%)	91.622	6.965	45.100	99.400
Student Teacher Ratio	12.096	3.832	1.300	43.900
White Enrollment Rate (%)	60.706	31.576	0.300	97.500
Economically Disadvantaged Enrollment Rate (%)	34.069	22.780	3.500	89.700
Limited English Proficiency Enrollment Rate (%)	7.340	12.129	0.000	92.300

Results of the single year regression model for 2018, equation 3.1, are presented in Table 11. Note that to save space, the intercepts of each of the 14 counties are not given. Firstly, as the overall F statistic suggests, the overall model is significant. The adjusted R^2 is about 0.58, meaning the model explains about 58 percent variation in the dropout rates, which implies that overall, these variables perform reasonably well in estimating dropout rates.

It can be observed that controlling for the other variables, attendance rate has the most significant association with dropout rates, a finding that is consistent with the plots in exploratory data analysis part. It is observed that this association is statistically significant at 5% level and 1 percentage point increase in a school's attendance rate is associated with a decrease in dropout rates with a magnitude of about 0.76 percentage points.

According to the results, percentage of enrolled students who are economically disadvantaged has the second most significant association with dropout rates. This association is statistically significant at 1% level and 1 percentage point increase in a school's percentage enrollment of economically disadvantaged students causes an increase in dropout rates with a magnitude of about 0.1 percentage points.

Racial decomposition of students, on the other hand, seems affecting the dropout rate, however not in line with the general literature and sign expectations. The results suggest a small positive effect of percentage enrollment of white students. One percentage point increase in percent white enrollment seems to be increasing the dropout rate by 0.06 percentage points, which is statistically significant at 5% level. However, it should be noted that its magnitude is not very large, considering the mean dropout rate of 3.889.

According to the regression results, percentage of students who have limited English proficiency seems to have statistically insignificant effects (at 5% level) on the dropout rates, after controlling for other variables. Another statistically insignificant effect is for the variable student teacher ratio. According to the regression results, any change in the number of students per teacher does not seem to be affecting significantly the dropout rate of that school at 5% level.

Table 11: Single Year Regression Model Results, 2018 and 2004

	Dependent Variable: Dropout Rate (%)	
	2018 Mean: 3.889	**2004** Mean: 4.364
Attendance Rate (%)	-0.764** (0.074)	-0.874** (0.144)
Student Teacher Ratio	0.112 (0.162)	0.391* (0.193)
White Enrollment Rate (%)	0.055* (0.028)	0.092* (0.038)
Economically Disadvantaged Enrollment Rate (%)	0.100* (0.047)	0.101* (0.043)
Limited English Proficiency Rate (%)	0.074 (0.074)	0.018 (0.140)
Observations	397	333
R^2	0.582	0.590
Adjusted R^2	0.562	0.566
Residual Std. Error (df = 378/314)	5.526	4.398
F Statistic (df = 18; 378/314)	29.194**	25.075**

Note: *p<0.05; **p<0.01

Since I have the results that are for the single year which is 2018, the same regression of equation 3.1, only this time using 2004 data, is estimated to see whether the signs and magnitudes of the coefficients change. Firstly, the descriptive statistics for 2004 data are given in Table 12.

Table 12: Descriptive Statistics (2004 Data)

	Mean	Std.Dev.	Min	Max
Dropout Rate (%)	4.364	6.677	0.000	55.200
Attendance Rate (%)	91.970	4.729	58.500	98.000
Student Teacher Ratio	13.456	4.211	3.100	63.300
White Enrollment Rate (%)	77.160	28.293	1.200	100.000
Economically Disadvantaged Enrollment Rate (%)	23.462	23.868	0.000	100.000
Limited English Proficiency Enrollment Rate (%)	3.019	6.543	0.000	67.600

The results of 2004 data are also given in Table 11. It is observed that attendance rate still has the strongest statistically significant association with dropout rates, although its coefficient is larger this time, around -0.87. It is also observed that enrollment percentage of economically disadvantaged students is statistically significant at 5% level again and still, the coefficient is nearly the same as that of 2018 results. White enrollment rate is also still statistically significant at 5% level; however, its magnitude differs from that of 2018 regression. From the 2004 regression results, the coefficient of white enrollment rate is 0.092, nearly double that of the 2018 regression result.

Results of regression on 2004 data suggest the percent enrollment of limited English proficiency students still does not seem to be statistically significant in affecting the dropout rates, controlling for other variables. However, the biggest difference between 2004 and 2018 regression results is that although student teacher ratio seemed statistically insignificant in predicting the dropout rates, 2004 regression results suggest that it has a statistically significant positive association with dropout rates with a magnitude of about 0.39 percentage points.

There are four takeaways from the single year regression models formed for the years 2004 and 2018. First is that, even controlling for the other school-related variables, attendance rate is a strong predictor of dropout rates. Both models suggest that 1 percentage point increase in attendance rates in a school corresponds to around 0.75-0.85 percentage point decrease in its dropout rates, which indicates a near one-to-one correspondence. Also, we observe that both models suggest around 0.1 percentage point increase in dropout rates in response to 1 percentage point increase in proportion of economically disadvantaged students, which is not very small when we take into account the mean dropout rates of 3.889 and 4.364 for years 2018 and 2004 respectively. This finding coincides with the literature in terms of sign expectations. Second is, racial composition of schools is also significant for both models, however, both models find results that are different from sign expectations. Although literature suggests a negative relationship between percentage enrollment of white students, both models suggest a statistically significant positive, albeit low in magnitude, relationship between percentage of white students and dropout rates. 2018 data model predicts a 0.05 percentage point increase and 2004 data predicts around 0.09 percentage point increase in dropout rates for a 1 percentage point increase in percentage enrollment of white students. This may simply point out that in Massachusetts high schools, dropout rates tend to increase with percentage of white students overall, contrary to the general US population. Another possibility is that white enrollment rate is correlated with another variable which is not accounted for in the model.

The third takeaway is that although 2018 model does not find student per teacher variable statistically significant, 2004 regression results suggest a positive significant association between number of students per teacher and dropout rates. Lastly, both models suggest that percentage of limited English Proficiency students does not significantly affect dropout rates for a school at any significant level below 5%.

A word of caution is necessary about these findings. These regressions, by their nature, treat every school the same and assumes that at any given year, there is no other variable apart from the stated ones that may differ among these schools. Thus, although these regressions give pretty good intuition on dropout rates and the school level factors associated with them, the coefficients may be biased to some extent. As the school-level variables I could find are limited and data on student-level are not available, there is always a risk of omitted variable bias. However, a considerable part of the omitted-variable bias problem can be solved through multi-year fixed effects regression, the results for which are given in the next part.

4.2. Multi-Year Panel Data Regression Model for Dropout Rates

The results of the fixed effects panel regression model, Equation 3.6, are presented in Table 13. There are three features of the fixed effects regression model. Firstly, results suggest that attendance rate has a statistically significant negative association with dropout rate of a given school, even after controlling for the other variables and time/school fixed effects. However, its coefficient is much smaller compared to the single year regression results given in the previous section. Results of the fixed effects regression suggest that one percentage point increase in attendance rates is associated with a 0.16 percentage point decrease in dropout rates. This is not surprising in that, as mentioned in the previous part, single year regressions do not take individual effects into account and may suffer from omitted variable bias. To give an example, suppose that a school has a high crime rate and thus has a low attendance rate, high dropout rate. In this possible scenario, since we do not have a crime rate variable, its effects are observed through attendance rate variable and cause its coefficient to be larger in magnitude. However, this problem is solved when we introduce school level fixed effects in the model.

Secondly, I notice that after controlling for both school and time fixed effects, the model suggests that both the white enrollment rate and economically disadvantaged enrollment rate variables become insignificant at any level below 5%. This means that the significant effect found in the single year models of these variables are not observed when we take the school and time fixed effects into account. Lastly, the variables limited English proficiency rate and student per teacher continue to be statistically insignificant for any significance level below 5%, although 2004 regression results suggest a positive statistically significant relation between student teacher ratio and dropout rates.

Table 13: School and Time Level Fixed Effects Panel Model, 2004-2018

	Dependent Variable: Dropout Rate (%) Mean: 4.385
Attendance Rate	-0.167** (0.028)
Student Teacher Ratio	-0.034 (0.041)
White Enrollment Rate (%)	-0.007 (0.028)
Economically Disadvantaged Enrollment Rate (%)	-0.003 (0.010)
Limited English Proficiency Rate (%)	-0.021 (0.023)
Observations	5,611
R^2	0.012
Adjusted R^2	-0.078
F Statistic (df = 5; 5139)	12.797**

Note: *p<0.05; **p<0.01

In addition to the above findings, in Table 14, I present the results of the dynamic panel data regression estimation given in 3.10. As mentioned in the previous chapter, the factors that may affect the past performance of schools in terms of dropout rates are already accounted for in the fixed effects model and I observe that all of the explanatory power of the model is observed through the lagged dependent variable.

Table 14: Dynamic Panel Data Model, 2004-2018

	Dependent Variable: **Dropout Rate (%)** Mean: 4.385
Dropout Rate (1 Year Lagged)	-0.250** (0.042)
Attendance Rate	-0.020 (0.117)
Student Teacher Ratio	-0.050 (0.056)
White Enrollment Rate (%)	0.011 (0.062)
Economically Disadvantaged Enrollment Rate (%)	0.006 (0.010)
Limited English Proficiency Rate (%)	-0.024 (0.063)
Observations	5,611
Sargan Test $\chi^2_{(103)}$	165.8**

Note: *p<0.05; **p<0.01

4.3. Difference in Differences Estimation

To find a sample satisfying parallel trend assumption, after examining the average dropout rates for each school district between 2004-2006, I find that 27 of these districts have the similar pattern to the Boston averages in the sense that the average dropout rate first decreases from 2004 to 2005 and then increases from 2005 to 2006.

Figure 21 plots the average dropout rates in Boston between 2004-2007 compared to the other 27 districts. Comparison of each of these districts to Boston averages are given in Appendix C. As it can be observed from Figure 21, these districts' average follows nearly the same trend as in Boston school district.

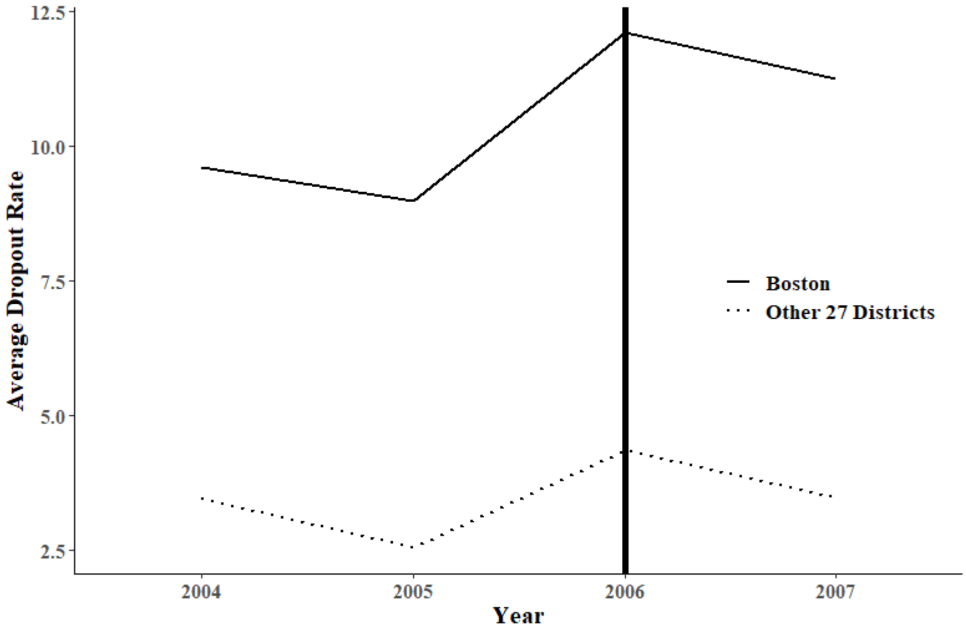


Figure 21: Average Dropout Rates in Boston and Other 27 Districts, 2004-2007

Note: Vertical line indicates the year after which new school choice mechanism has been used. Before the school year 2006-2007, Boston Mechanism has been used and starting with the school year 2006-2007, Student Optimal Stable Mechanism has been implemented in Boston high schools. Dropout rates are calculated annually according to the definition given above.

Note: The 27 districts are: Belchertown, Belmont, Blackstone Valley Regional Vocational Technical, Bourne, Bridgewater, Bristol County Agricultural, Cambridge, Danvers, East Bridgewater, Essex Agricultural Technical, Hampshire, Harwich, Holyoke, Hudson, Marblehead, Milford, Millbury, Minuteman Regional Vocational Technical, North Central Charter Essential (District), Pioneer Valley, Quaboag Regional, Saugus, South Hadley, Southern Berkshire, Sturgis Charter Public (District), Swampscott, Winchendon

Source: Own Calculations based on Massachusetts Department of Elementary and Secondary Education, Statewide Reports

Another way to test whether parallel trend assumption holds for these districts is through the following regression equation;

$$DRP_{it} = \beta X_{it} + \beta_1 BOSTON_i + \beta_2 (BOSTON_i \times 2007_t) + \beta_3 (BOSTON_i \times 2006_t) + \beta_4 (BOSTON_i \times 2005_t) + \beta_5 2007_t + \beta_6 2006_t + \beta_7 2005_t + \varepsilon_{it} \quad (4.1)$$

where DRP_{it} is the dropout rate of school i in time t, X_{it} is the matrix of Attendance Rate, Student Teacher Ratio, Percentage White Enrollment, Percentage Economically Disadvantaged Enrollment and Percentage Limited English Proficiency Enrollment variables for school i in time t, $BOSTON_i$ is a binary variable which equals 1 if school i is in the treatment group (Boston schools), 2007_t, 2006_t, and 2005_t are binary variables denoting years. In this setting, 2004 is the base year and the variables 2006_t and 2005_t must be insignificant for the parallel trend assumption to hold. With the 27 districts forming the control group, results of the regression of Equation 4.1 are given in Table 15. Since these variables are found to be insignificant, we conclude that parallel trend assumption holds with these 27 districts given in Figure 21 forming the control group as opposed to the treatment group of Boston school district.

Table 15: Testing the Parallel Trend Assumption

	Dependent Variable: **Dropout Rate (%)**
Attendance Rate (%)	-0.852** (0.113)
Student Teacher Ratio	0.010 (0.122)
White Enrollment Rate (%)	-0.131** (0.054)
Economically Disadvantaged Enrollment Rate (%)	-0.185** (0.065)
Limited English Proficiency Rate (%)	0.140** (0.056)
Boston	-0.538 (2.066)
2005	-1.201 (0.828)
2006	1.158 (0.792)
2007	0.118 (0.759)
Boston x 2005	-0.464 (1.818)
Boston x 2006	1.025 (2.149)
Boston x 2007	0.251 (1.814)
Constant	96.362** (10.906)
Observations	241
R^2	0.556
Adjusted R^2	0.532
Residual Std. Error (df = 228)	5.995
F Statistic (df = 12; 228)	23.770**

Note: *p<0.05; **p<0.01

Before presenting the regression results, I give the mean values for the used sample. With the 27 districts given in Figure 21 forming the control group in the DID analysis, Table 16 summarizes the means of each variable in both treatment and control groups in 2006 (before the algorithm change) and 2007 (after the algorithm change);

Table 16: Mean Values (2006, 2007 Boston and Control Group)

	BOSTON HIGH SCHOOLS			CONTROL GROUP HIGH SCHOOLS		
	2006	2007	POOLED	2006	2007	POOLED
Dropout Rate (%)	12.1	11.3	11.7	4.4	3.5	3.9
Attendance Rate (%)	87.1	86.6	86.8	92.8	92.7	92.8
Student Teacher Ratio	14.2	13.7	13.9	12.7	12.2	12.4
White Enrollment Rate (%)	12.0	11.1	11.5	81.6	81.2	81.4
Economically Disadvantaged Enrollment Rate (%)	65.3	63.8	64.6	21.2	21.1	21.2
Limited English Proficiency Enrollment Rate (%)	12.8	14.1	13.4	3.1	2.8	2.9

We observe that in Boston high schools, average dropout rate dropped from 12.1% to 11.3% and we also see a decrease of average dropout rates in the control group, formed by the 27 districts given in Figure 21. With this in mind, I run the regression of Equation 3.12. The result of the regression is given in Table 17 with the heteroscedasticity and autocorrelation robust HAC standard errors.

Table 17: Difference in Difference Model, 2006-2007

	Dependent Variable: **Dropout Rate (%)**
Attendance Rate (%)	-1.054** (0.163)
Student Teacher Ratio	-0.268** (0.095)
White Enrollment Rate (%)	-0.181* (0.083)
Economically Disadvantaged Enrollment Rate (%)	-0.279** (0.104)
Limited English Proficiency Rate (%)	0.132* (0.065)
Boston	0.430 (2.422)
2007	-1.242 (0.875)
Boston x 2007	-0.983 (2.076)
Constant	125.953** (17.154)
Observations	128
R^2	0.609
Adjusted R^2	0.583
Residual Std. Error (df = 119)	6.277
F Statistic (df = 8; 119)	23.376**

Note: *p<0.05; **p<0.01

The parameter of interest, $\widehat{\beta_3}$, is about -0.98, which implies that the stated change in school choice matching algorithm resulted in 0.98 percentage points decrease in the dropout rates in Boston high schools. Although its sign seems to be in line with the expectations, it does not appear to be statistically significant. Thus, it is concluded that the algorithm change from Boston Mechanism to Student Optimal Stable Mechanism does not seem to result in a decrease in dropout rates of Boston high schools in the statistical sense.

It is important to note that the insignificance result is possibly about the imbalance in the observations used for the model. Among the 128 observations, 70 of them are for Boston high schools and the remaining 58 are for the other schools in the control group. Since there are only 27 districts satisfying the parallel trend assumption, the control group's sample size is considerably low compared to that of treatment group. Thus, as a cautionary note, it may be too strict to conclude directly from the model that the algorithm change was ineffective in reducing the dropout rates. Nonetheless, according to the model, its effect appears to be not statistically significant.

CHAPTER 5

CONCLUSION

In this thesis, I study the high school dropout rates in Massachusetts, trying to identify how certain school-related characteristics influence the dropout rates of schools and whether a change in the school choice mechanism possibly affected the rates. The data covers the schools in Massachusetts, for the school years between 2003-2004 and 2017-2018. For each school in a given year, the following variables are taken into account; (i) attendance rate, an indicator of overall academic environment, (ii) student-teacher ratio, indicating the school resources, and the variables about the racial and socio-economic composition of students enrolled; (iii) percentage enrollment of white students, (iv) percentage enrollment of economically disadvantaged students, and (v) percentage enrollment of limited English proficiency students.

To understand the predictors of the high school dropout rates, I develop two models. The first is the single-year model, which is based on the 2018 data. I also do a robustness check of this model using 2004 data. The second model is a fixed-effects panel data model for the years between 2004 and 2018. In this model, I further analyze the dropout rates controlling for the school and time fixed effects. Then, for the aim of trying to identify how the change in school choice mechanism in Boston, from Boston Mechanism to Student Optimal Stable Mechanism, affected the high school dropout rates, I conduct a difference-in-difference analysis. I first compare average dropout rates between 2004 and 2006 for each school district to the averages of Boston to find out which of these districts satisfy the parallel trend assumption. I show, both graphically and by means of an auxiliary regression, that 27 districts' dropout rate averages share the same trend as Boston averages. Then, I apply DID analysis to these dropout rates, with Boston being the treatment group and the 27 districts who share a common trend with Boston school district forming the control group.

In terms of the aim of understanding the predictors of dropout rates, results indicate that, when only a single year is considered, attendance rate, percentage of economically disadvantaged students and white enrollment percentage are statistically significant at 5% level in predicting the dropout rates. 1 percentage point increase in attendance rate is associated with about 0.75 to 0.85 percentage points decrease in dropout rates, 1 percentage point increase in percentage of economically disadvantaged students seems to be increasing the dropout rate by about 0.1 percentage points and 1 percentage point increase in white enrollment percentage is associated with about 0.05 to 0.09 percentage point increase in dropout rates. It is observed that controlling for these variables, the percentage of limited English proficiency students have insignificant effects in dropout rates. Single year models of 2004 data and 2018 data suggest mixed results about the number of students per teacher, regression of 2004 data finds positive significant association and regression of 2018 data finds this association insignificant.

When the nature of analysis is switched to a multi-year setting using data between years 2004-2018 and controlling for the school and time fixed effects, I observe that the results change drastically. Firstly, results suggest that even after controlling for the school and time fixed effects, attendance rate has statistically significant negative association with dropout rates. However, I observe that the magnitude of this negative association is much smaller, 0.16 percentage point, compared to the single year regression results where school fixed effects are not taken into account. Secondly, when the school and time fixed effects are controlled for, the other variables become statistically insignificant. Thus, the overall conclusion is that, even after school level fixed effects are accounted for, increasing the overall attendance rate in a given school would be successful in decreasing the dropout rates.

In terms of understanding the effects of a mechanism change in Boston on dropout rates, it is observed that DID analysis, with the districts for which average dropout rates behave similar to that of Boston averages between school years 2003-2004 and 2005-2006 forming the control group, does not find a statistically significant change in dropout rates. Although the coefficient of the DID estimator, -0.98, coincides with the expectations, I conclude that the stated mechanism change did not result in a statistically significant decrease in dropout rates.

This thesis contributes to the literature in two ways. First is that it furthers the literature on the school-related factors contributing to the dropout rates. It supports the negative association of attendance rates, both in the single-year setting and multi-year school and time fixed effects setting and casts doubts on the effects of the other variables when school and time fixed effects are controlled for. Secondly, it is concluded that the stated mechanism change in Boston high schools did not result in a statistically significant decrease in dropout rates.

Although the models given in this thesis furthers the intuition on school dropout rates, they are not without their limitations and potentials for improvement. Firstly, due to data limitations, the variables such as number of students per teacher and attendance rates are taken to be proxy variables for the resources and general/academic environment of a school respectively. By their nature, they may contain some information that are also correlated with dropout rates and thus may result in biased estimation. Although I believe school level fixed effects solve a great deal of this problem, as the coefficient of attendance rate drops significantly despite still being significant, there may still be the possibility of estimation errors. Also, as student-level data is not available and school-level data is only limited to the variables given in this thesis between 2004-2018, it was not possible to include additional information potentially related with the dropout rates. Thus, future work on the subject can be greatly strengthened if more and more variables are available in both student and school level. Secondly, as pointed out, the data used on DID analysis is very imbalanced in that as opposed to 70 treatment group observations, there are 58 control group observations that satisfy the parallel trend assumption, which makes it extremely difficult to obtain a significant estimate. Data on schools in other states that share similar characteristics and parallel average dropout rate trends with Boston high schools would increase the odds of a stronger analysis. Lastly, it is important to note that, the analysis looks at the difference only between 2006 and 2007. Time period of the analysis is not extended further to make sure that there is less possibility of another policy change occurring during the same time. However, since the dropout rates are defined for all grades between 9-12 and grade-by-grade dropout rates of Massachusetts high schools are not available for the years before 2008, only the 9th graders in school year 2006-2007 are affected by the mechanism change in 2007. Thus, the analysis is done without observing the full effect of the change, which may affect both the estimate itself and whether it is significant or not. For potential future work, data on grade-by-grade dropout rates for the years 2006 and 2007 would make the analysis much more reliable in terms of its conclusion.

REFERENCES

Abdulkadiroğlu, Atila, Parag Pathak, Alvin E Roth, and Tayfun Sönmez (2006) "Changing the Boston school choice mechanism," Technical report, National Bureau of Economic Research.

Abdulkadiroğlu, Atila and Tayfun Sönmez (2003) "School choice: A mechanism design approach," American economic review, 93 (3), 729–747.

Ahrens, Henry and Robert Pincus (1981) "On two measures of unbalancedness in a one-way model and their relation to efficiency," Biometrical Journal, 23 (3), 227–235.

Aina, C., Baici, E., Casalone, G., & Pastore, F. (2018). The economics of university dropouts and delayed graduation: a survey.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. The review of economic studies, 58(2), 277-297.

Attanasio, O. P. (2015). The determinants of human capital formation during the early years of life: Theory, measurement, and policies. Journal of the European Economic Association, 13(6), 949-997.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. Journal of political economy, 70(5, Part 2), 9-49.

Belfield, Clive R and Henry M Levin (2007) The price we pay: Economic and social consequences of inadequate education: Brookings Institution Press.

Checchi, D. (2006). The economics of education: Human capital, family background and inequality. Cambridge University Press.

Christle, Christine A, Kristine Jolivette, and C Michael Nelson (2007) "School characteristics related to high school dropout rates," Remedial and Special education, 28 (6), 325–339.

Connell, J. P. (1990). Context, self, and action: A motivational analysis of self-system processes across the life span. In D. Cicchetti (Ed.), The self in transition: From infancy to childhood (pp. 61-97). Chicago: University of Chicago Press.

Croissant, Yves, Giovanni Millo et al. (2019) Panel data econometrics with R: Wiley Online Library.

Cullen, Julie Berry, Brian A Jacob, and Steven Levitt (2006) “The effect of school choice on participants: Evidence from randomized lotteries,” Econometrica, 74 (5), 1191–1230.

Deming, David J, Justine S Hastings, Thomas J Kane, and Douglas O Staiger (2014) “School choice, school quality, and postsecondary attainment,” American Economic Review, 104 (3), 991–1013.

Driscoll, John C and Aart C Kraay (1998) “Consistent covariance matrix estimation with spatially dependent panel data,” Review of economics and statistics, 80 (4), 549–560.

Finn, J. D. (1989). Withdrawing from school. Review of educational research, 59(2), 117-142.

Gamoran, Adam (1992) “Social factors in education,” Encyclopedia of educational research, 6, 1222–1229.

Goode, R. B. (1959). Adding to the stock of physical and human capital. The American Economic Review, 49(2), 147-155.

Hanushek, E.A. (1986). The economics of schooling: Production and efficiency in public schools. Journal of Economic Literature, 24, 1141-1177.

Hastings, Justine S, Christopher A Neilson, and Seth D Zimmerman (2012) “The effect of school choice on intrinsic motivation and academic outcomes,” Technical report, National Bureau of Economic Research.

Heckman, James J and Paul A LaFontaine (2010) “The American high school graduation rate: Trends and levels,” The review of economics and statistics, 92 (2), 244–262.

Hoechle, Daniel (2007) “Robust standard errors for panel regressions with cross-sectional dependence,” The stata journal, 7 (3), 281–312.

Hynes, M. (2014). Don't Call Them Dropouts: Understanding the Experiences of Young People Who Leave High School before Graduation. A Report from America's Promise Alliance and its Center for Promise at Tufts University. America's Promise Alliance.

Lavy, Victor (2010) “Effects of free choice among public schools,” The Review of Economic Studies, 77 (3), 1164–1191.

Lee, V. E., & Burkam, D. T. (2003). Dropping out of high school: The role of school organization and structure. American educational research journal, 40(2), 353-393.

Li, Mingliang (2007) “Bayesian proportional hazard analysis of the timing of high school dropout decisions,” Econometric Reviews, 26 (5), 529–556.

Lochner, L., & Monge-Naranjo, A. (2012). Credit constraints in education. Annu. Rev. Econ., 4(1), 225-256.

McFarland, Joel, Jiashan Cui, Juliet Holmes, and Xiaolei Wang (2020) “Trends in High School Dropout and Completion Rates in the United States: 2019. Compendium Report. NCES 2020-117.,” National Center for Education Statistics.

McNeal Jr, Ralph B (1997) “High school dropouts: A closer examination of school effects,” Social Science Quarterly, 209–222.

McNeal Jr, R. B. (1999). Parental involvement as social capital: Differential effectiveness on science achievement, truancy, and dropping out. Social forces, 78(1), 117-144.

Mincer, J. (1958). Investment in human capital and personal income distribution. Journal of political economy, 66(4), 281-302.

Newman, F.M., Wehlage, G.G. and Lamborn, S.D. (1992) The Significance and Sources of Student Engagement. In: Newman, F.M., Ed., Student Engagement and Achievement in American Secondary Schools, Teachers College Press, New York, 11-39.

Pesaran, M Hashem (2004) “General diagnostic tests for cross section dependence in panels.”

Pong, Suet-Ling and Dong-Beom Ju (2000) “The effects of change in family structure and income on dropping out of middle and high school,” Journal of family issues, 21 (2), 147–169.

Rumberger, Russell W (1983) “Dropping out of high school: The influence of race, sex, and family background,” American educational research journal, 20 (2), 199–220.

Rumberger, Russell W (1995) “Dropping out of middle school: A multilevel analysis of students and schools,” American educational Research journal, 32 (3), 583–625.

Rumberger, R. W. & Palardy, G. J. (2004). Multilevel models for school effectiveness research. In D. Kaplan (Ed.), Handbook of Quantitative Methodology for the Social Sciences (pp.235-258). Thousand Oaks, CA: Sage Publications.

Rumberger, Russell W and Sun Ah Lim (2008) “Why students drop out of school: A review of 25 years of research.”

Rumberger, Russell W and Scott L Thomas (2000) “The distribution of dropout and turnover rates among urban and suburban high schools,” Sociology of education, 39–67.

Sander, William (2001) “Chicago public schools and student achievement,” Urban Education, 36 (1), 27–38.

Sen, A. (1980). Equality of what?. The Tanner lecture on human values, 1, 197-220.

Spence, M. 1973. Job market signaling. Quarterly Journal of Economics 87, 355–79.

Tinto, V. (1987). Leaving college: Rethinking the causes and cures of student attrition. University of Chicago Press, 5801 S. Ellis Avenue, Chicago, IL 60637.

Vulkan, Nir, Alvin E Roth, and Zvika Neeman (2013) The handbook of market design: OUP Oxford.

Wehlage, G. G., Rutter, R. A., Smith, G. A., Lesko, N. L., & Fernandez, R. R. (1989). Reducing the risk: Schools as communities of support. Philadelphia: Farmer Press.

Wooldridge, Jeffrey M (2002) "Econometric analysis of cross section and panel data MIT press," Cambridge, MA, 108.

Worrell, Frank C and Robert L Hale (2001) "The relationship of hope in the future and perceived school climate to school completion.," School Psychology Quarterly, 16 (4), 370.

APPENDICES

A. CITIES IN MASSACHUSETTS

Table 18 gives the total population, number of high schools and number of high school students in Massachusetts as of July 1, 2019. The total population data comes from the U.S. Census Bureau's most recent sub-county population estimates release on May 21, 2020, which includes estimates for Massachusetts cities and towns as of July 1, 2019. Number of schools and students data comes from the Enrollment Report of Massachusetts Department of Elementary and Secondary Education.

Table 18: Population, Number of High Schools and Students of Massachusetts Cities

City	Total Population	Number of High Schools	Number of High School Students
Boston	692,600	33	15,035
Worcester	185,428	7	7,158
Springfield	153,606	13	6,955
Cambridge	118,927	1	1,981
Lowell	110,997	4	3,299
Brockton	95,708	5	4,419
New Bedford	95,363	3	2,311
Quincy	94,470	2	2,768
Lynn	94,299	4	4,559
Fall River	89,541	3	2,265
Newton	88,414	2	4,016
Somerville	81,360	2	1,289
Lawrence	80,028	3	3,426
Framingham	74,416	1	2,271
Haverhill	64,014	3	1,854
Waltham	62,495	1	1,608
Malden	60,470	1	1,832
Weymouth	57,746	1	1,861
Taunton	57,464	2	2,083

Table 18 (Continued)

Medford	57,341	2	1,292
Chicopee	55,126	3	2,300
Revere	53,073	2	2,061
Peabody	53,070	1	1,425
Methuen	50,706	1	1,979
Everett	46,451	2	2,002
Attleboro	45,237	2	1,790
Barnstable	44,477	1	1,426
Salem	43,226	3	982
Beverly	42,174	1	1,250
Pittsfield	42,142	2	1,603
Bridgewater-Raynham	42,089	2	1,473
Leominster	41,716	3	1,811
Westfield	41,204	2	1,731
Fitchburg	40,638	2	1,320
Woburn	40,228	1	1,301
Holyoke	40,117	1	1,508
Amherst	39,924	1	917
Chelsea	39,690	2	1,410
Marlborough	39,597	1	1,067
Braintree	37,190	1	1,654
Watertown	35,939	1	652
Randolph	34,362	1	646
Franklin	34,087	1	1,744
Gloucester	30,430	1	804
Agawam	28,613	1	1,104
West Springfield	28,517	1	1,199
Northampton	28,451	1	848
Melrose	28,016	1	1,008
Gardner	20,683	2	563
Winthrop	18,544	1	595
Newburyport	18,289	1	754
Amesbury	17,532	2	611
Greenfield	17,258	1	361
Southbridge	16,878	2	511
Easthampton	15,829	1	469
North Adams	12,730	1	304
Palmer	12,232	1	329

B. COUNTIES IN MASSACHUSETTS

Table 19 gives the total population, number of high schools and number of high school students in Massachusetts counties as of July 1, 2019. The total population data comes from the U.S. Census Bureau's most recent county population estimates release on May 21, 2020, which includes estimates for Massachusetts counties as of July 1, 2019. Number of schools and students data comes from the Enrollment Report of Massachusetts Department of Elementary and Secondary Education.

Table 19: Population, Number of High Schools and Students of Massachusetts Counties

County	Total Population	Number of High Schools	Number of High School Students
Middlesex	1,600,842	68	67,281
Worcester	824,772	60	38,282
Suffolk	796,605	52	23,539
Essex	783,676	42	35,014
Norfolk	700,437	32	31,365
Bristol	561,037	31	26,061
Plymouth	515,303	35	25,026
Hampden	467,871	35	21,741
Barnstable	213,496	11	8,143
Hampshire	161,032	14	5,748
Berkshire	126,425	12	4,937
Franklin	70,577	11	2,754
Dukes	17,312	2	673
Nantucket	11,168	1	532

C. CONTROL GROUP DISTRICTS COMPARISON WITH BOSTON

In this section, graphical comparison of average dropout rates in Boston high schools with each of the 27 districts given in Figure 21 which are found to have similar trend are provided. Dropout rates are calculated annually according to the definition given in Chapter 3.2. The values are calculated based on Massachusetts Department of Elementary and Secondary Education, Statewide Reports.

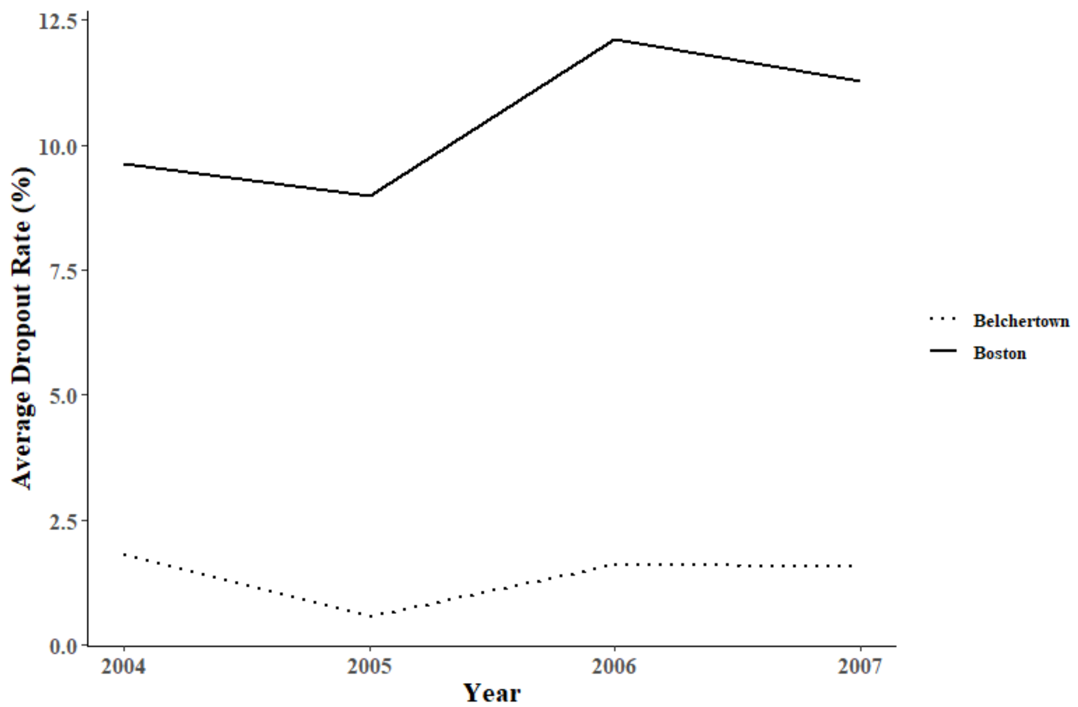
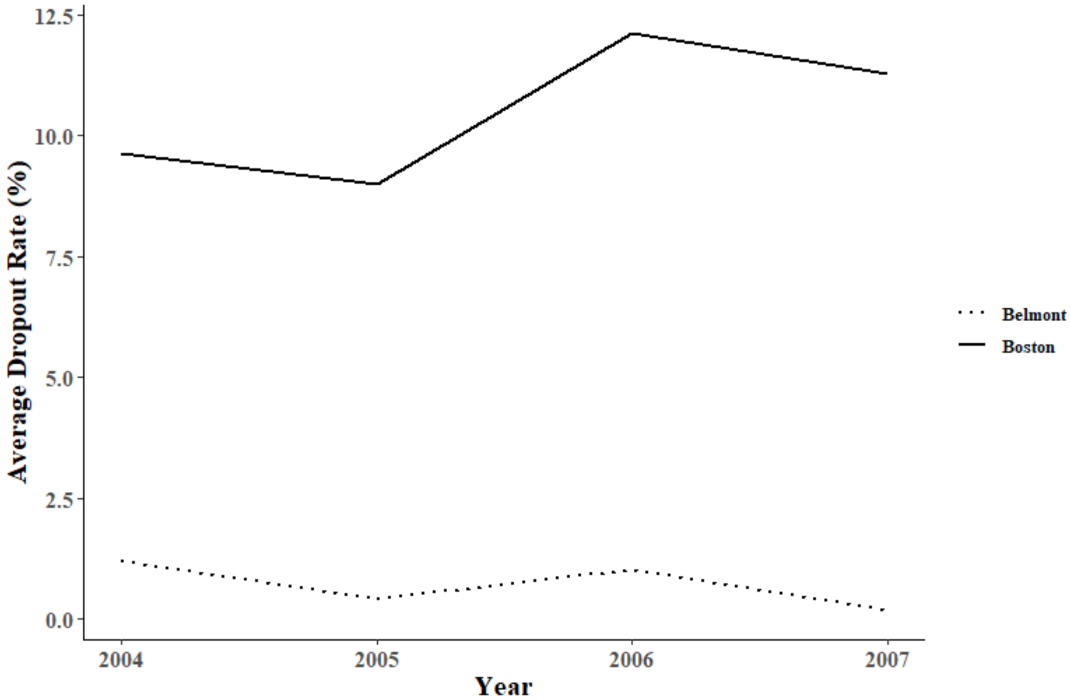


Figure 22: Average Dropout Rates in Boston and Belchertown



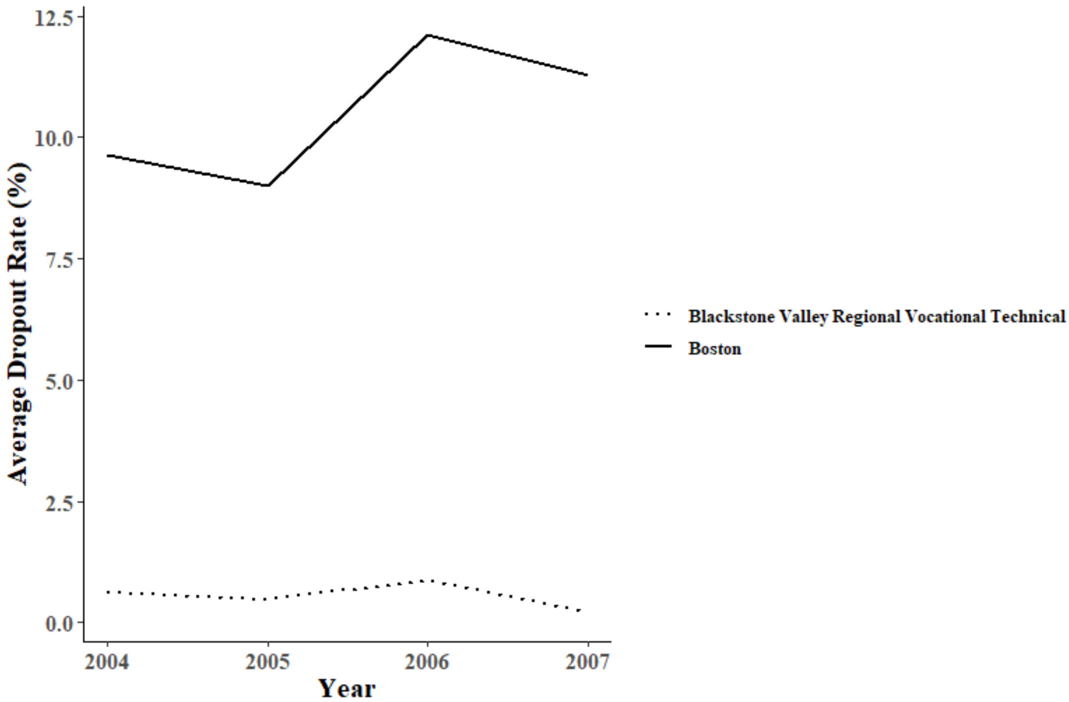
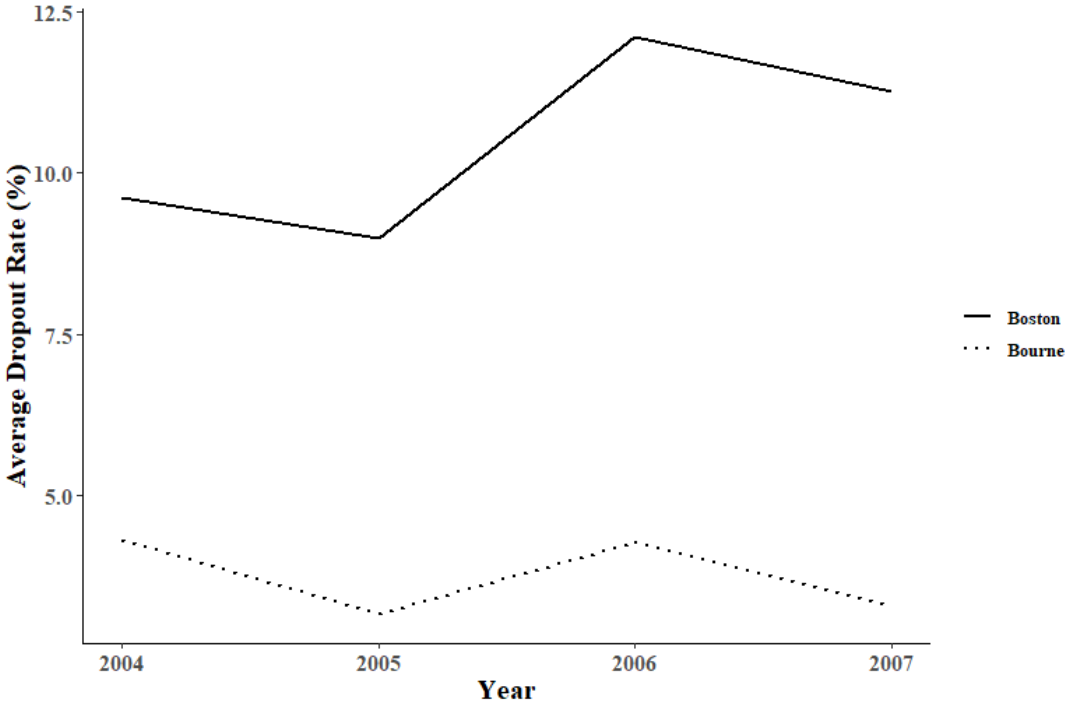


Figure 23: Average Dropout Rates in Boston, Belmont and Blackstone Valley Regional Vocational Technical



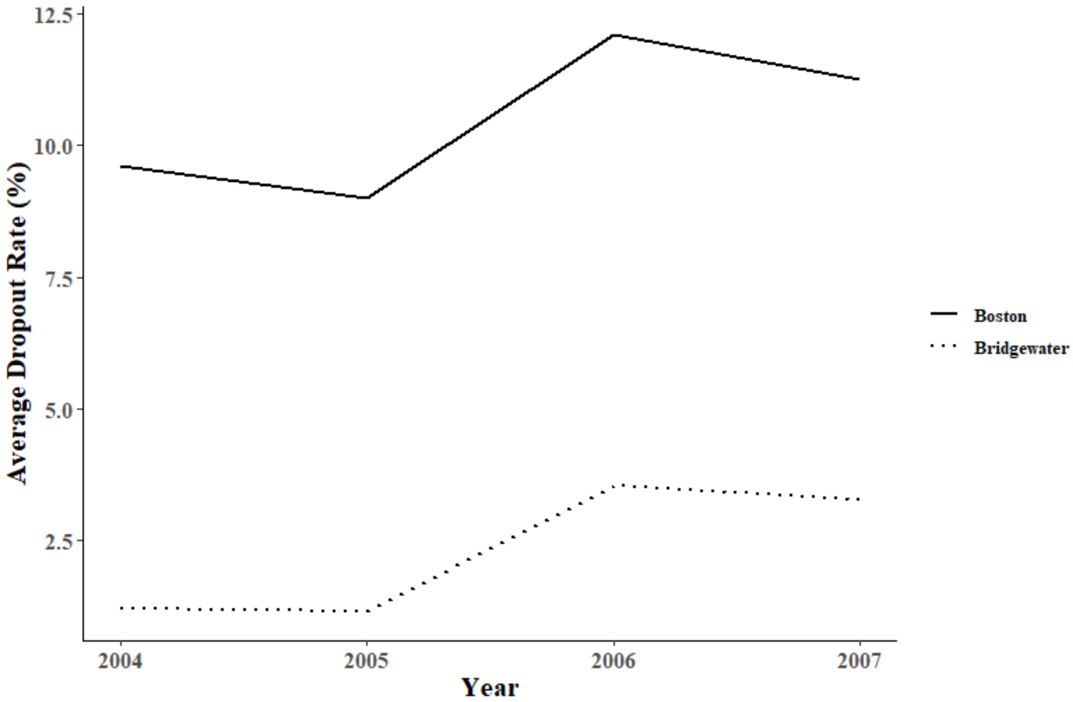


Figure 24: Average Dropout Rates in Boston, Bourne and Bridgewater

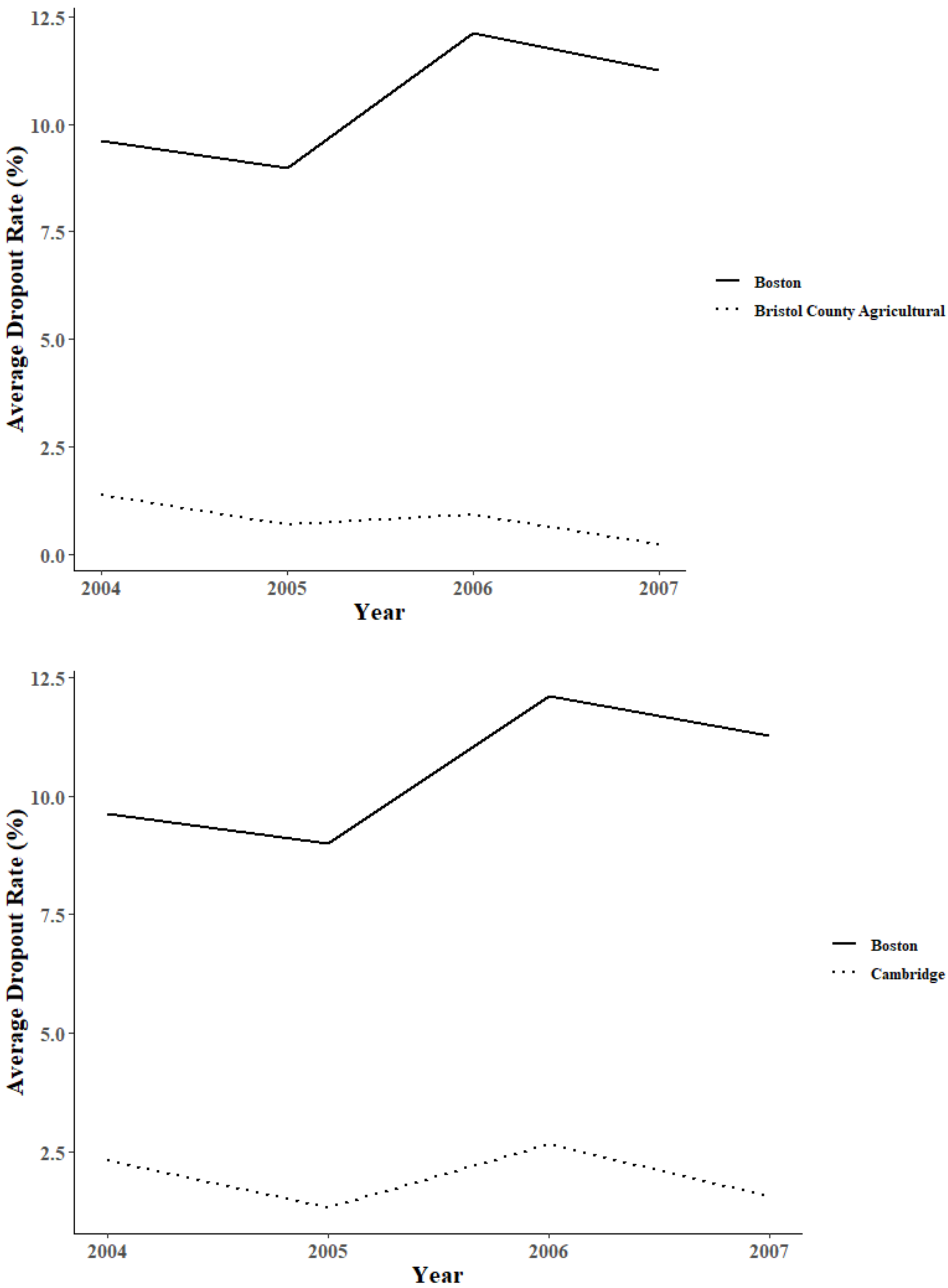
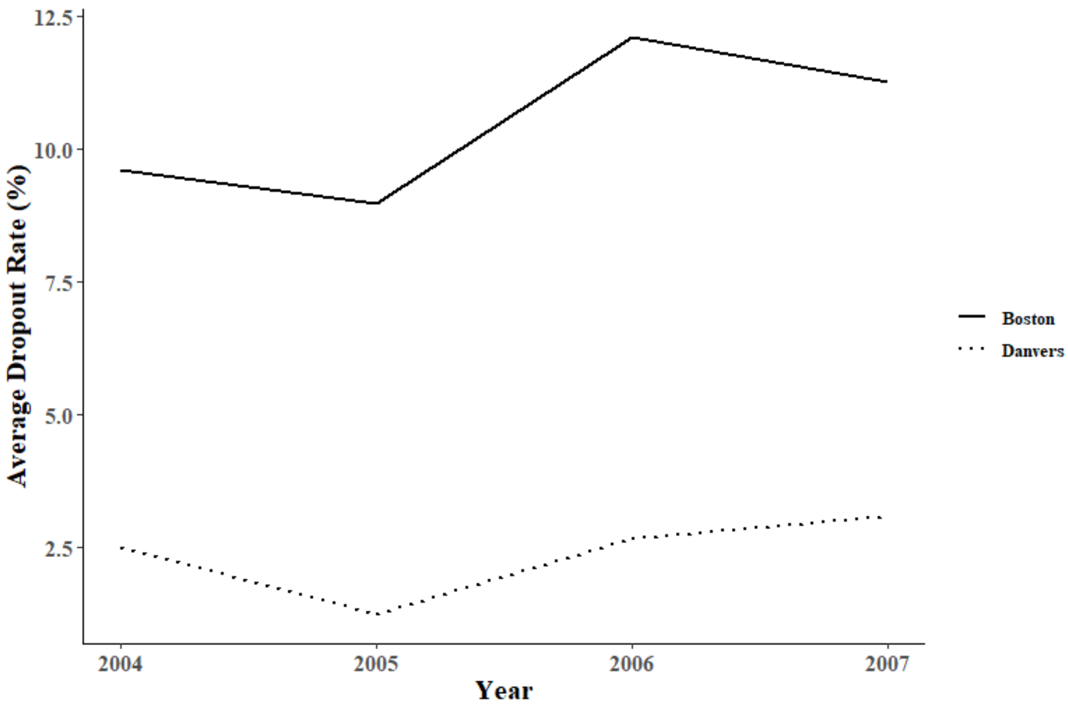


Figure 25: Average Dropout Rates in Boston, Bristol County Agricultural and Cambridge



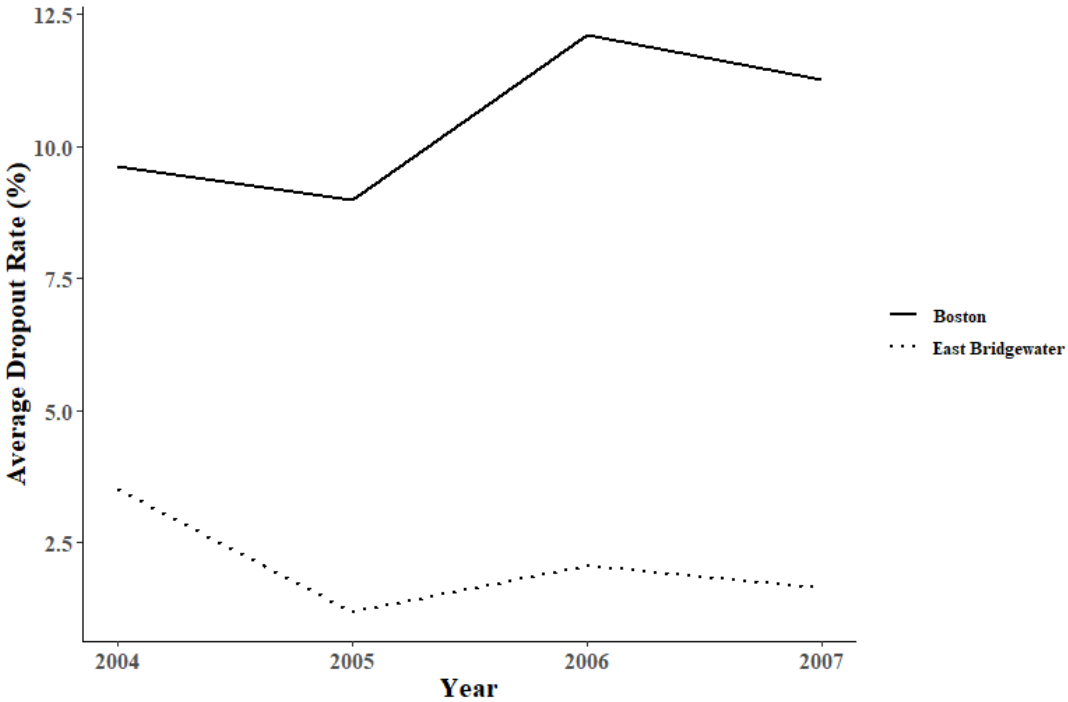


Figure 26: Average Dropout Rates in Boston, Danvers and East Bridgewater

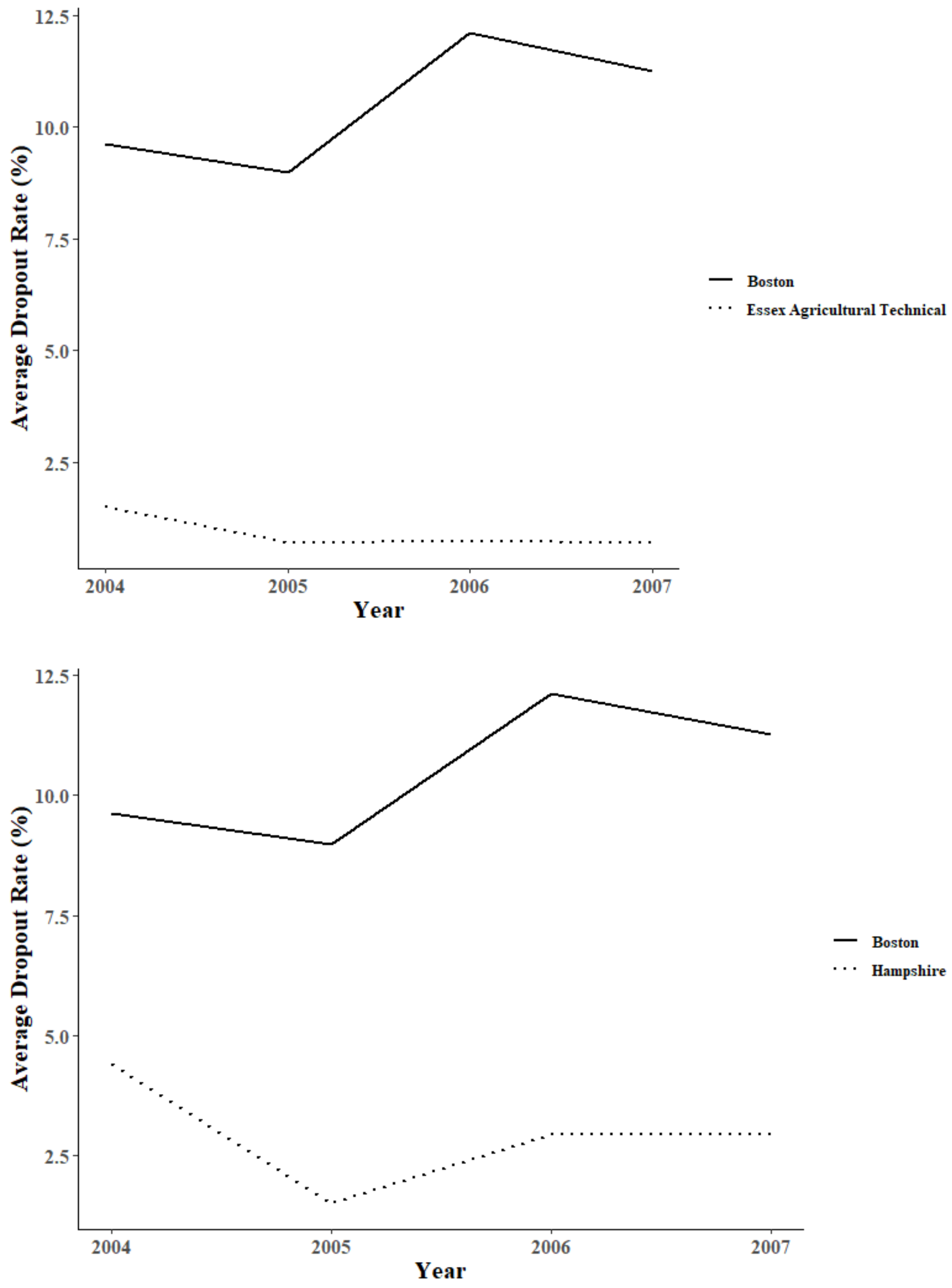
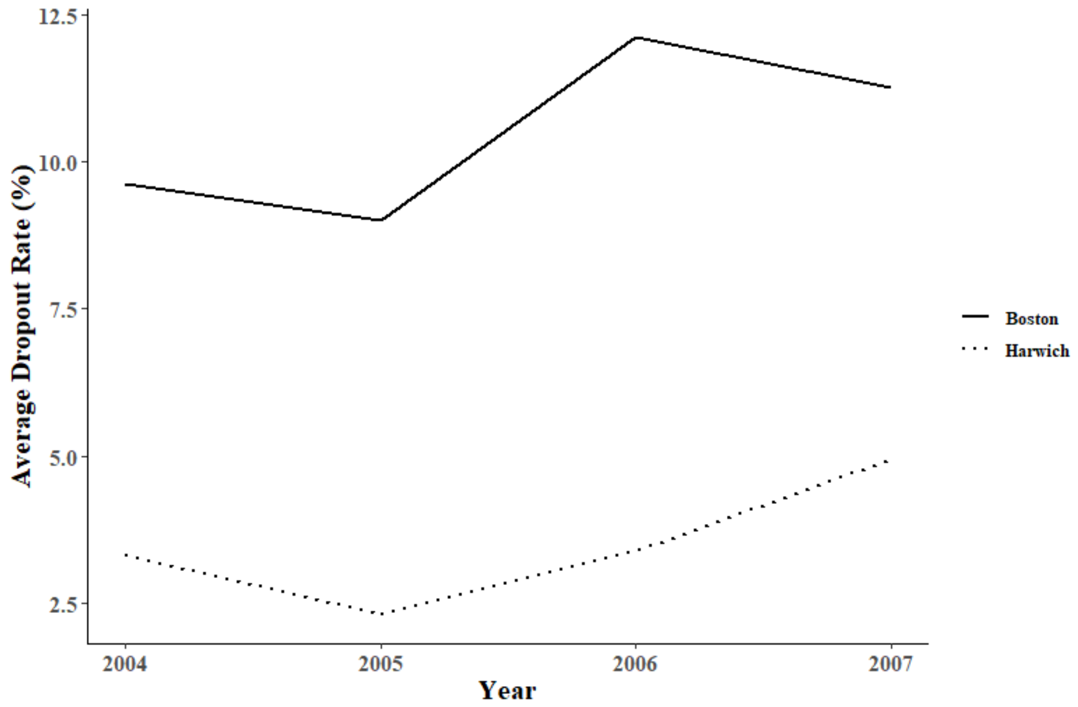


Figure 27: Average Dropout Rates in Boston, Essex Agricultural Technical and Hampshire



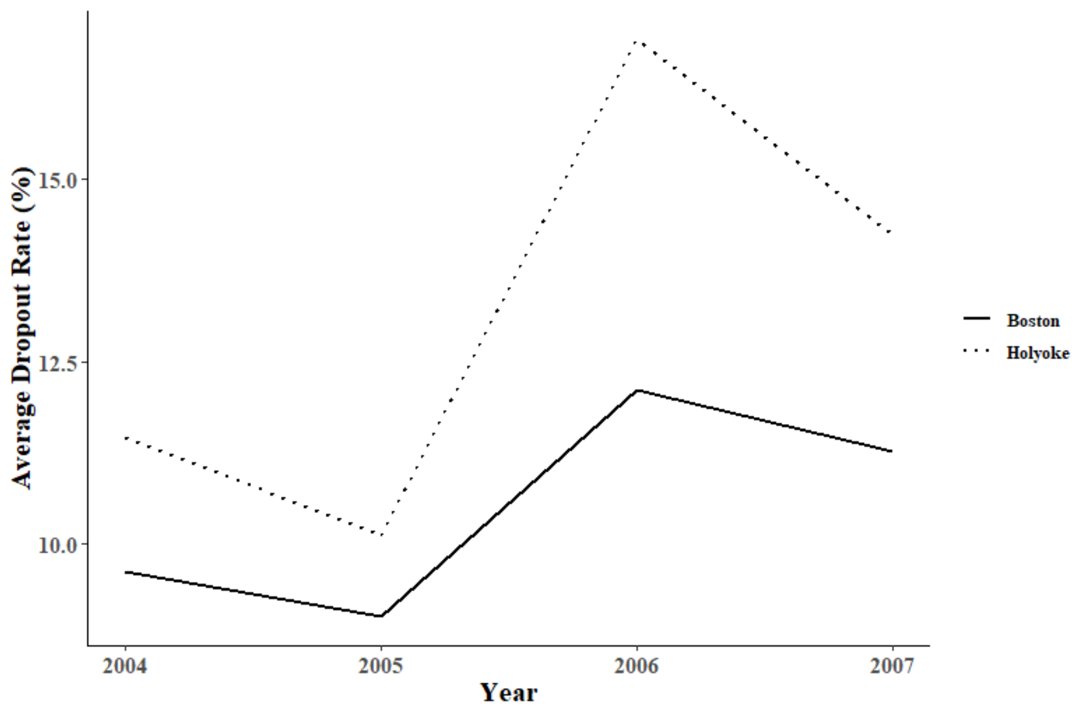


Figure 28: Average Dropout Rates in Boston, Harwich and Holyoke

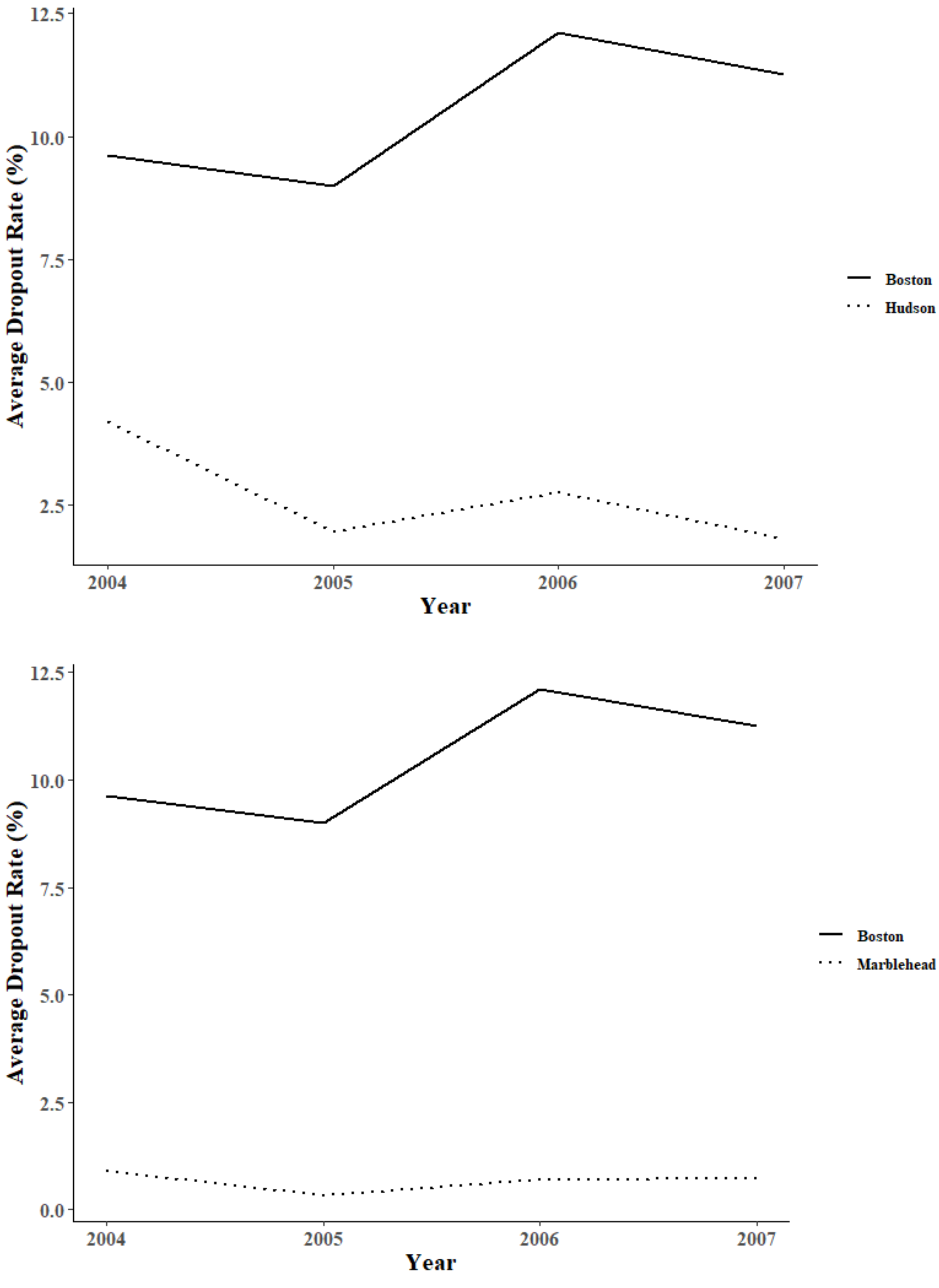
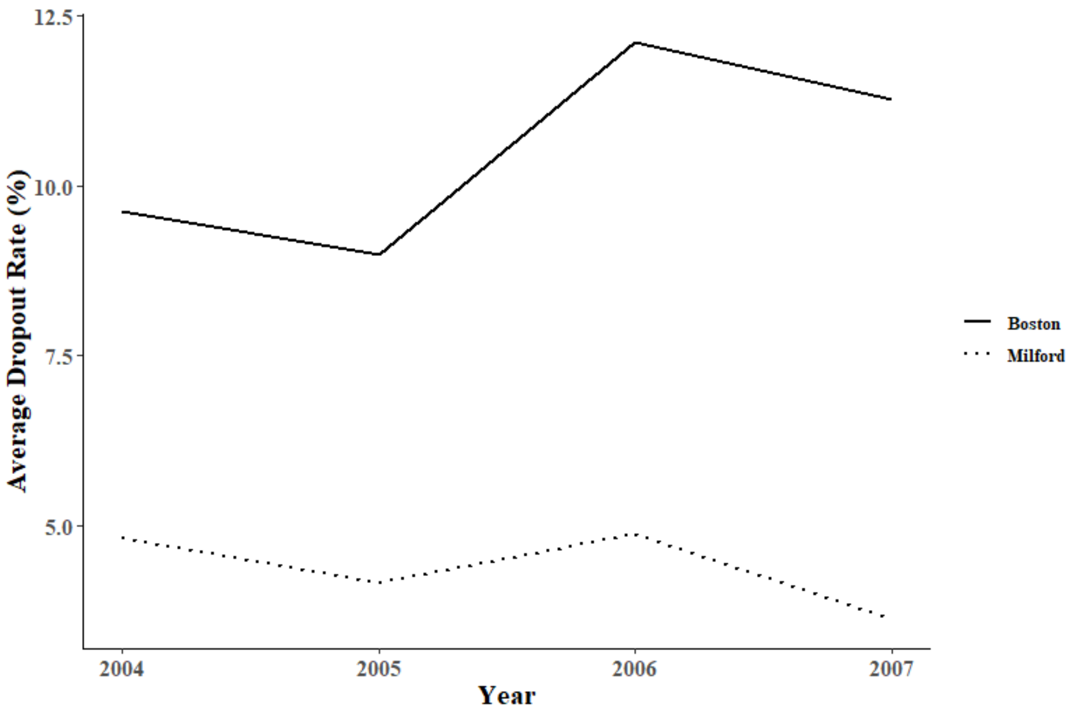


Figure 29: Average Dropout Rates in Boston, Hudson and Marblehead



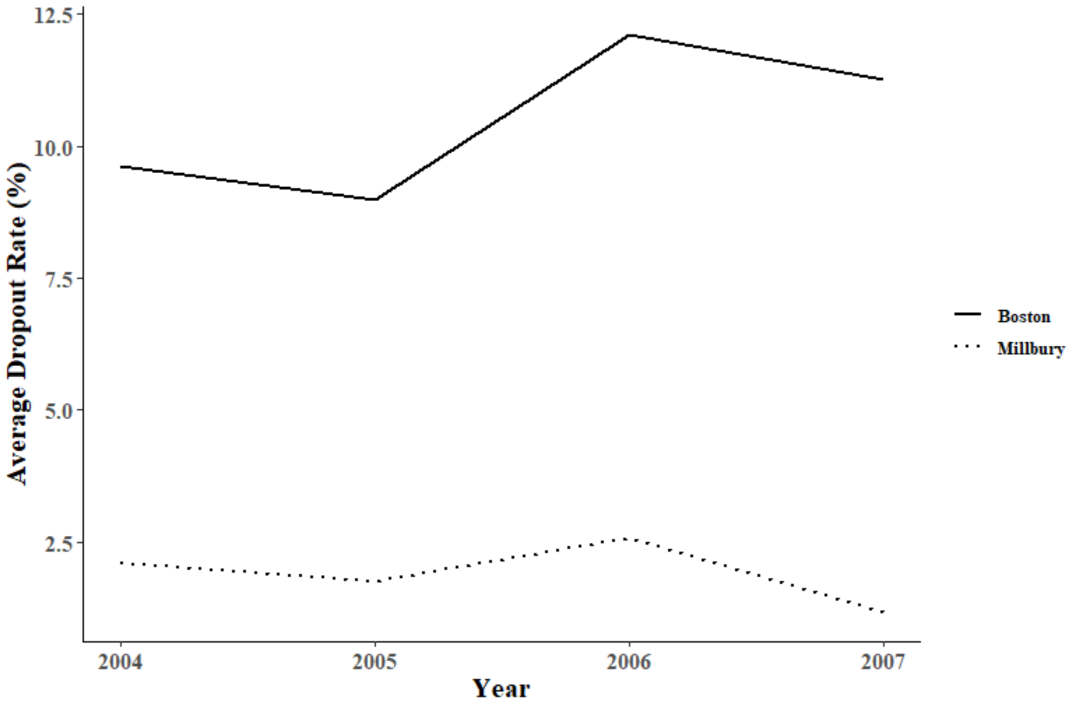
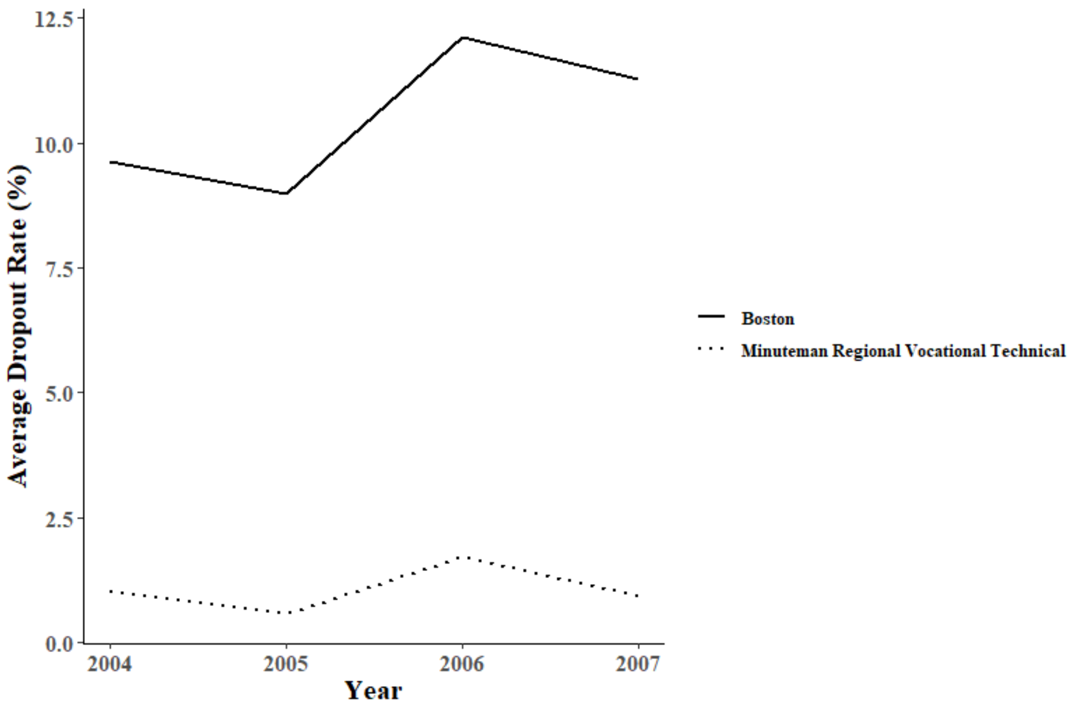


Figure 30: Average Dropout Rates in Boston, Milford and Millbury



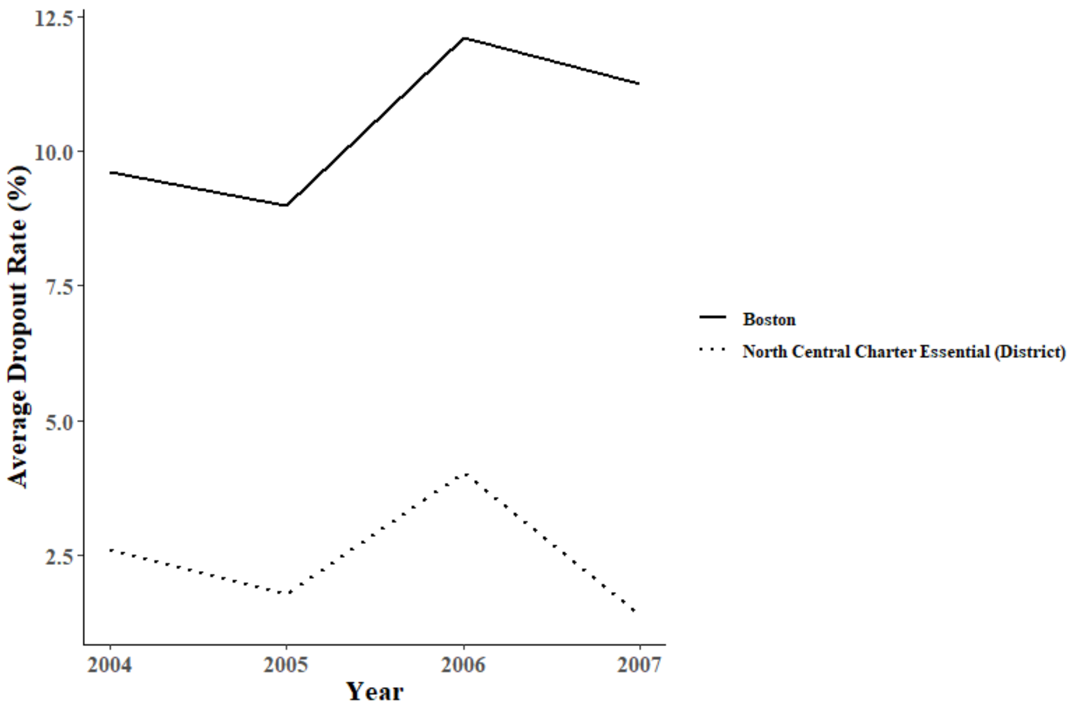


Figure 31: Average Dropout Rates in Boston, Minuteman Regional Vocational Technical and North Central Charter Essential (District)

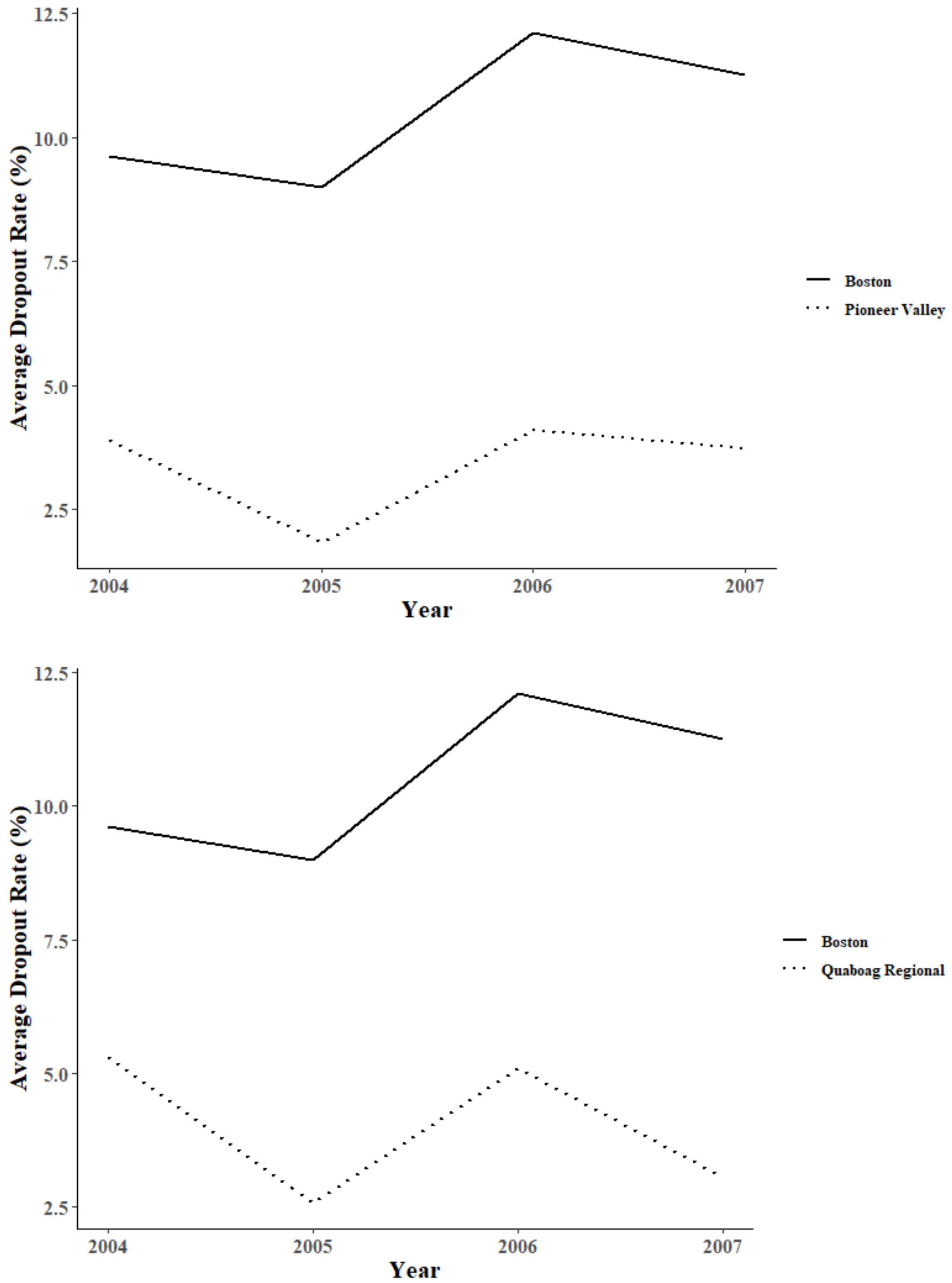
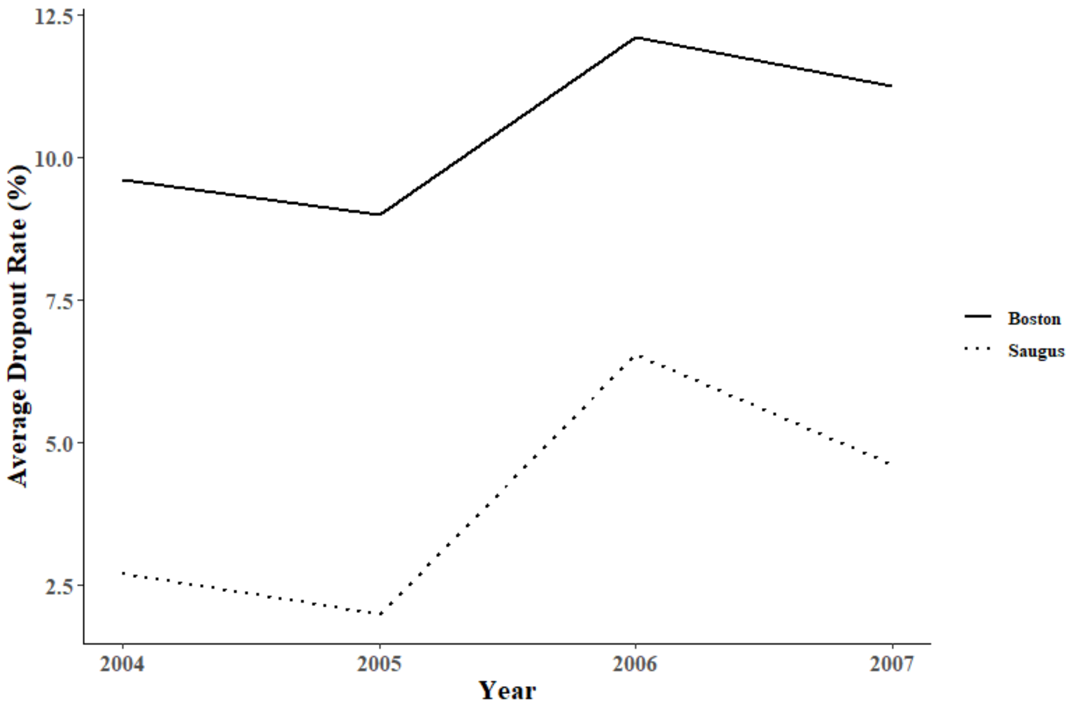


Figure 32: Average Dropout Rates in Boston, Pioneer Valley and Quaboag Regional



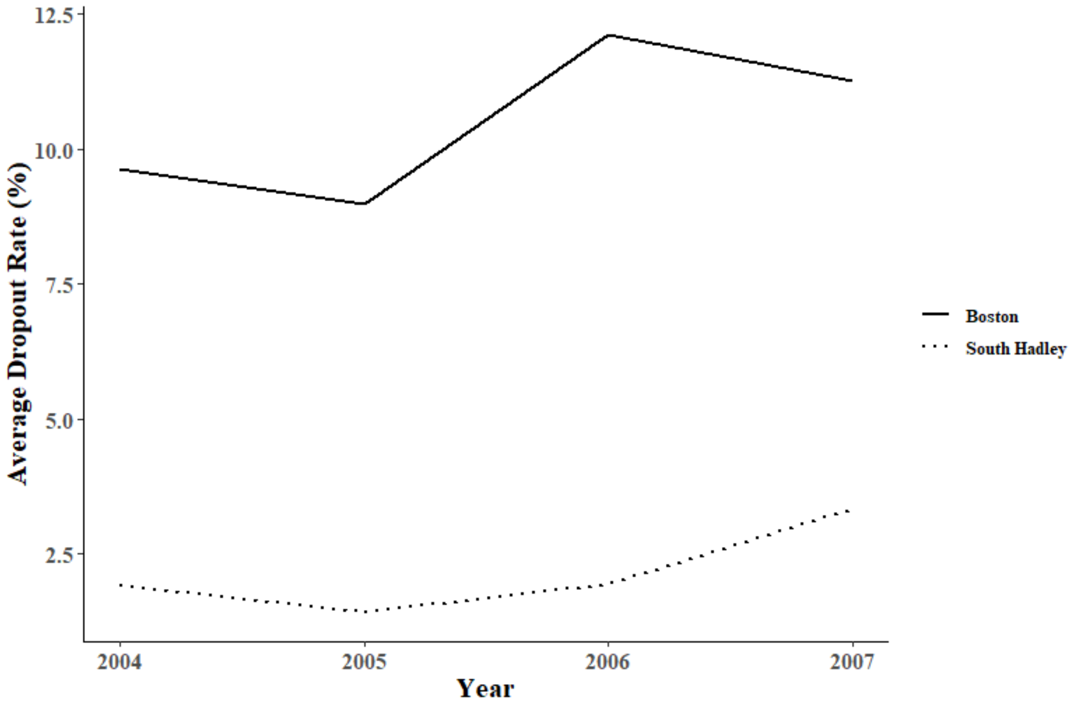
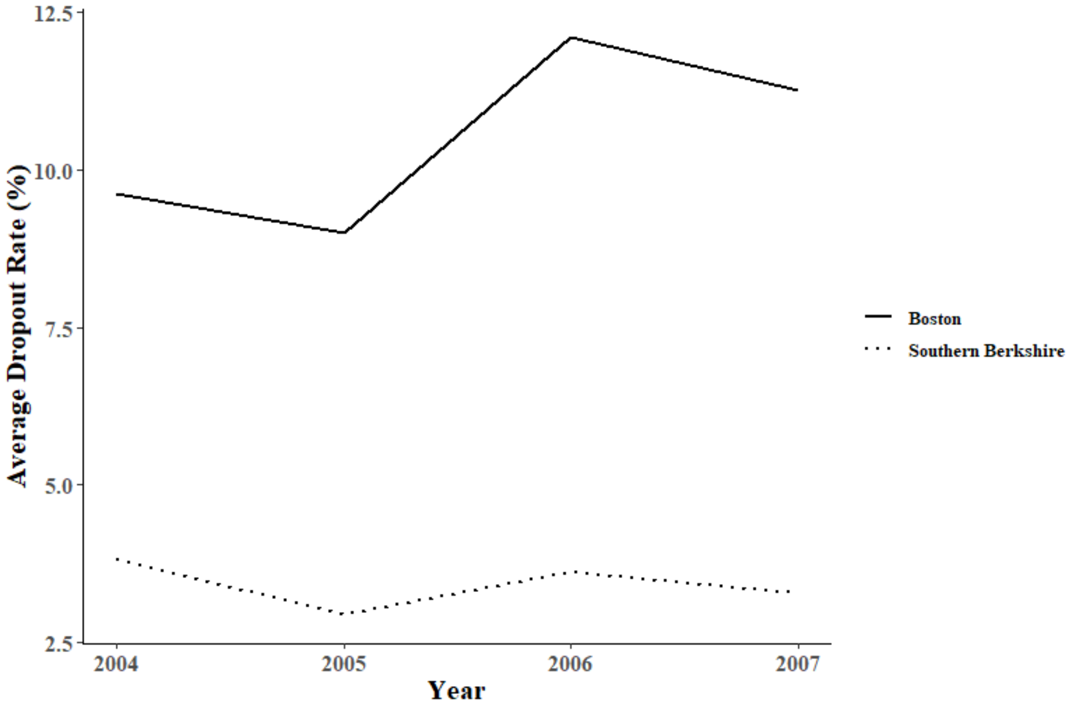


Figure 33: Average Dropout Rates in Boston, Saugus and South Hadley



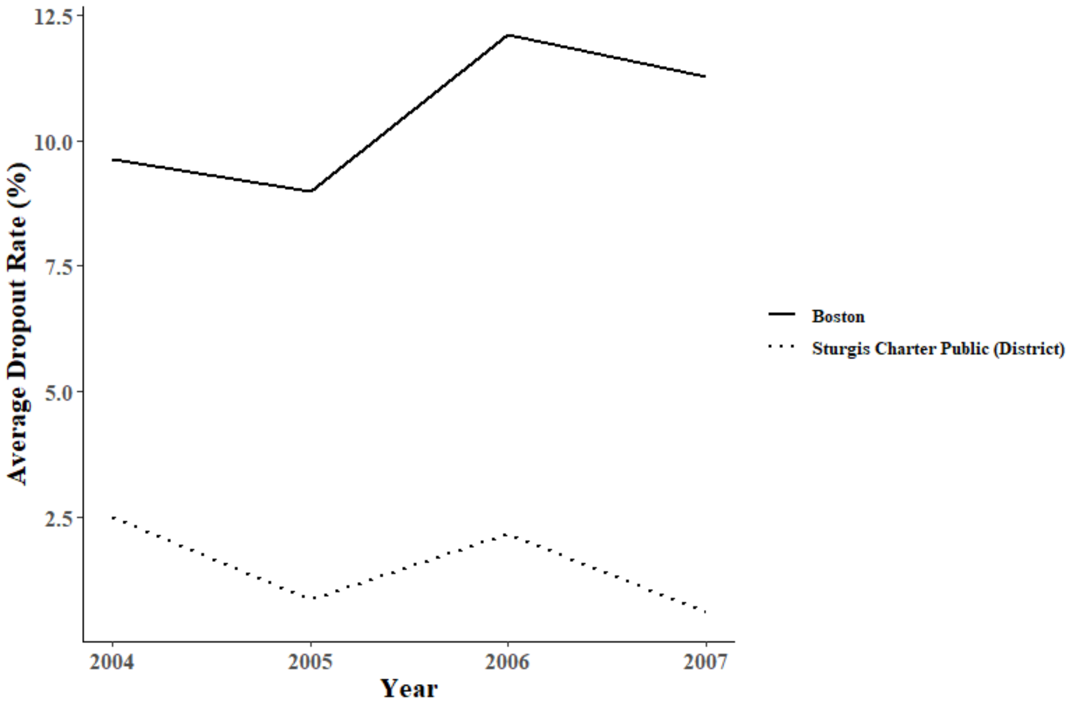


Figure 34: Average Dropout Rates in Boston, Southern Berkshire and Sturgis Charter Public (District)

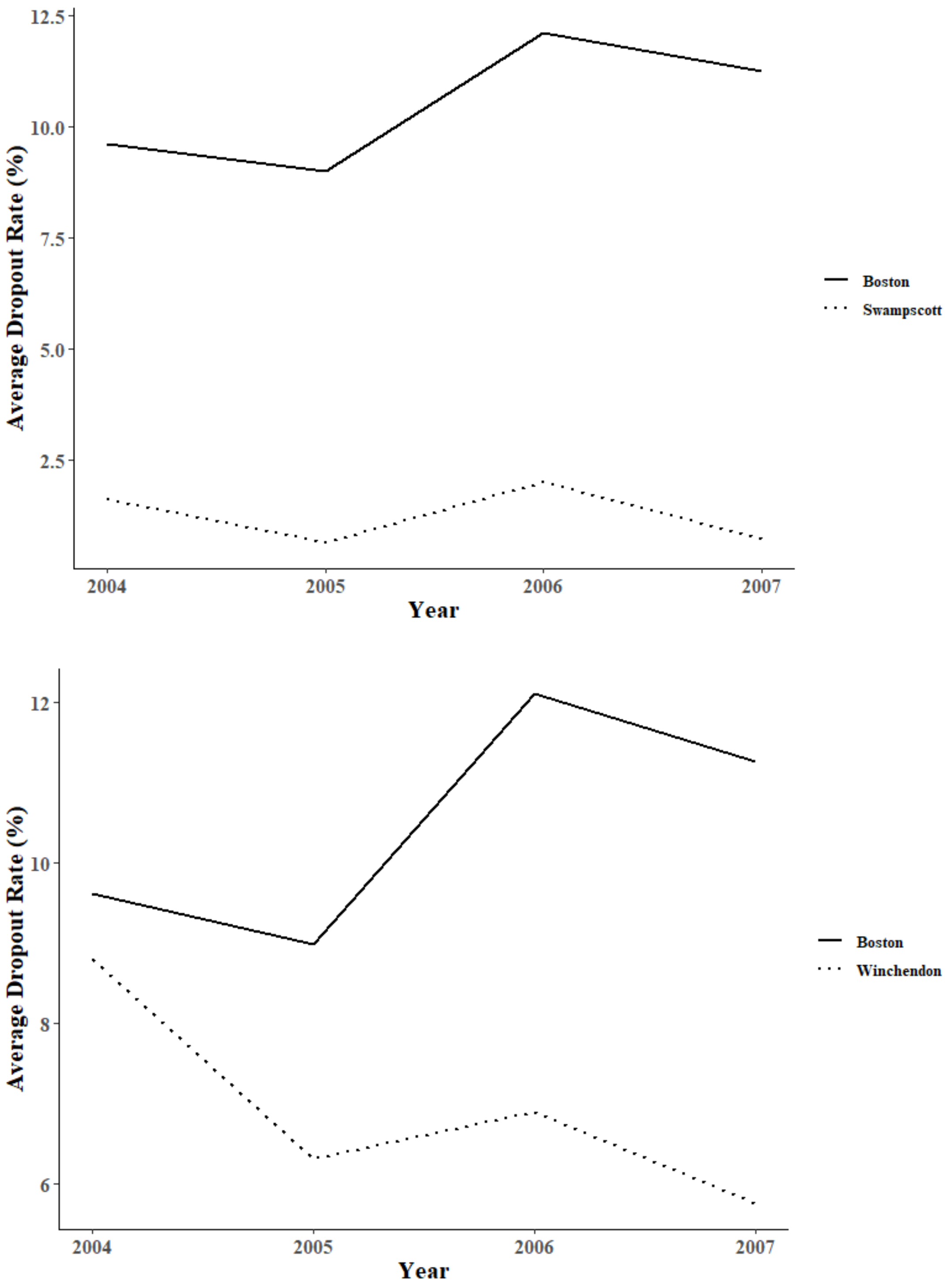


Figure 35: Average Dropout Rates in Boston, Swampscott and Winchendon

D. TURKISH SUMMARY / TÜRKÇE ÖZET

Amerika'da liseyi bırakma oranlarının yüksekliği ciddi bir problem olarak güncelliğini korumaktadır. Amerika Ulusal Eğitim İstatistikleri Merkezi (National Center for Education Statistics - NCES) tarafından kullanılan olay bırakma oranı (event dropout rate) metriğine göre 2016-2017 okul yılında Amerika'da okulu bırakma oranı yüzde 4,7 olarak gerçekleşmiştir. Söz konusu olay bırakma oranı, 15-24 yaş aralığında ve 10.-12.sınıflar arasında kayıtlı öğrencilerden okulu bırakanların yüzdesi olarak tanımlanmakta olup okulu bırakma eylemi ise bir lise diploması ya da GED gibi alternatif bir yeterlilik belgesi almadan okuldan ayrılmak olarak tanımlanmaktadır (McFarland et al. (2020)). Belirli öğrenci grupları düşünüldüğünde ise bu oranlar daha endişe uyandırıcı hâle gelmektedir. Örneğin 2017 yılında, olay bırakma oranı metriğine göre, bırakma oranı siyahi öğrenciler için yüzde 5,5, Hispanik öğrenciler için yüzde 6,5 ve engelli öğrenciler için yüzde 6,2 olarak gerçekleşmiştir.

NCES tarafından kullanılan diğer bir metrik olan statüs bırakma oranı (status dropout rate), 16-24 yaş aralığında olup herhangi bir okula kayıt olmamış ve bir lise diploması ya da GED gibi alternatif bir yeterlilik belgesi almamış sivil ve herhangi bir kurumla ilişkisi olmayan bireylerin yüzdesi olarak tanımlanmaktadır. Bu metriğe göre ise, 2017 yılında Amerika'da bırakma oranı yüzde 5,8 olarak ölçülmüş olup bu metriğe göre de belirli öğrenci gruplarında artış gözlenmektedir.

1992 ve 2017 yılları arasında bu metriklerin değerleri incelendiğinde, statüs bırakma oranının 1995'teki pik noktası yüzde 12'den 2017'de yüzde 5,8'e olmak üzere yüzde 50'nin üzerinde bir düşüş gerçekleştirdiği görülmektedir. Fakat olay bırakma oranı'nın yüzde 4 ve 6 seviyeleri arasında seyrettiği ve 2014-2017 yılları arasında da yüzde 6 seviyesinde yaklaşık olarak sabit kaldığı görülmektedir.

Liselerdeki bırakma oranlarını üzerinde düşünmeye değecek bir problem yapan bir diğer neden ise yüksek seviyelerdeki bırakma oranlarının sonuçlarıdır. Amerika Çalışma Bakanlığı İşgücü İstatistikleri Bürosu'na göre, 2019'un üçüncü çeyreğinde tam zamanlı çalışan 25 yaş üstü bireylerin lise diploması olmayanlarının haftalık medyan kazançları 606 $ olurken lise diploması olup herhangi bir lisans diploması bulunmayanların ise 749 $ olarak gerçekleşmiştir. (Belfield and Levin (2007)), liseyi bırakanların işsiz kalma oranlarının daha fazla olduğunu ve iş buldukları durumda da lise mezunlarına kıyasla maaşlarının daha düşük olduğunu gözlemlemektedir. Bir başka çalışmada ise, liseyi bırakanların lise mezunlarına kıyaslandığında sağlık sorunlarına ve suç içerikli eylemlere daha yatkın olduğu belirtilmektedir (Rumberger and Lim (2008)). Bunların yanı sıra, sosyal maliyetlerinden dolayı yüksek seviyede bir lise bırakma oranı genel anlamıyla ekonomiyi de etkilemektedir. Lise mezunlarına kıyasla daha az kazandıkları için, liseyi bırakanlar daha az vergi ödeyip gelirlerini sağlayabilmek için devlet programlarına daha fazla ihtiyaç duymaktadırlar (Rumberger and Lim (2008)).

Literatür taramasında açıklanacağı üzere, öğrencilerin neden liseyi terk ettikleri ve bırakma oranlarının prediktörlerinin ve ilişkilerinin neler olduğuna dair geniş bir literatür bulunmakla birlikte, literatürde sıklıkla gözden kaçan bir husus, liseyi terk etme oranları ile okul seçim mekanizmaları arasındaki bağlantıdır. Bir çocuğun gideceği okulu seçme yeteneği olarak tanımlanan okul seçimi, çocukları okul çağına geldiğinde ebeveynlerin karşılaştığı en önemli kararlardan biridir. Mekanizma tasarımı açısından okul seçim problemi; sonlu bir öğrenci kümesi, sonlu sayıda yere sahip bir okul kümesi, her bir öğrenci için okullar açısından tercih edilme sırası ve her bir okul için belli bir kritere göre belirlenen öğrencilerin öncelik sıralamasından oluşmaktadır (Vulkan et al. (2013)). Bu unsurlar dikkate alınarak, öğrenciler okullara belirli bir algoritma/mekanizma eşliğinde yerleştirilirler. Bu kapsamda dünya üzerinde en çok kullanılan algoritma Boston Mekanizmasıdır (Boston Mechanism-BM). Boston şehrinde ise, bu mekanizma 2006-2007 okul yılında Abdulkadiroğlu and Sönmez (2003) tarafından önerilen Öğrenci Optimal Kararlı Mekanizma (Student Optimal Stable Mechanism-SOSM) ile değiştirilmiştir. Bu değişimin en önemli sebebi ise, bazı ebeveynlerin BM algoritmasının doğasında yer alan dezavantajları kullanarak okullar açısından gerçek tercihlerini yanlış beyan etmek suretiyle tercihlerini doğru beyan eden ebeveynlerin zarar görmesine sebebiyet verebilecek olmasıdır. Bu açıdan daha adil olan mekanizmanın lise bırakma oranlarında düşüşe yol açtığının gösterilmesi durumunda, politika üreticilerinin Massachusetts eyaletindeki diğer okul bölgelerini de daha adil olan uygulamaya geçirmesinin önü açılacaktır.

Dolayısıyla, bu tezin iki amacı bulunmaktadır. İlk olarak, okullarla ilgili özelliklerin ne derecede liseyi bırakma oranlarını etkilediği, Massachusetts eyaletindeki okul düzeyinde veri kullanılarak incelenmektedir. Bu kapsamda iki farklı ekonometrik model sunulmaktadır. İkinci olarak ise, Boston şehrinde 2006-2007 okul yılında gerçekleşen liselere geçişte kullanılan okul seçim mekanizmasındaki değişikliğin liseleri bırakma oranlarını etkileyip etkilemediği incelenmektedir. Bu kapsamda ise, farkların farkı metodu kullanılmaktadır.

Literatür incelendiğinde, insanların neden okula gittiğini anlamak için ortaya atılan fikirlerin başında kabiliyet kavramı gelmektedir. Sen (1980) tarafından ortaya atılan yaklaşım, kabiliyeti bireylerin olmak veya yapmak istediği şeylere erişebilme özgürlüğü olarak tanımlamaktadır. Örneğin, okumak ve bilgiyi işleyebilmek normal bir hayat sürmek için kritik kabiliyetler olarak düşünülebilir. Chechhi (2006), normal bir hayat sürmek için gerekli olan minimal kabiliyetlerin kazanılmasını eğitim talebinin basit bir açıklayıcısı olarak düşünmektedir. Etkili bir başlangıç noktası olmasına rağmen, bu yaklaşım insanların söz konusu minimal kabiliyetleri kazandıktan sonra da eğitimlerine devam etmelerini açıklayamamaktadır. Eğitim talebini açıklayan en çok kabul görmüş yaklaşımlardan birisi de Becker (1962) tarafından formalize edilen beşerî sermaye teorisidir. Beşerî sermaye; işgücünün sahip olduğu, mal ve hizmetlerin üretimi sürecinde yardımcı olan bilgi ve beceriler olarak tanımlanmaktadır (Goode (1959)). Öncelikle, Mincer (1958) eğitim seviyesinin bireysel gelir farklılıklarını önemli derecede açıkladığını ve eğitim seviyesinin artmasıyla ücretlerin arttığını gözlemlemiştir. Becker (1962) ise bunun üzerine bugünkü anlamıyla beşerî sermaye teorisini formalize etmiştir.

Bu teoriye göre, her bir birey parasal ve parasal olmayan getirilerin bugünkü net değeri eğitim masrafları toplamının bugünkü net değerine eşitlenene kadar eğitime yatırım yapmaktadır. Fakat, eğitim seviyesinin getirileri çok yüksek olmasına rağmen insanlar okumayı bırakmaktadırlar. Bu durumu genel olarak iki farklı yaklaşım açıklamaktadır. Birincisi, bireyler eğitim almayı çoğunlukla maddi nedenlerle bırakmaktadır. Son yıllarda eğitim masraflarının önemli ölçüde artması ve borçlanma olanaklarının giderek kısıtlanması sonrası eğitim almanın masrafları oldukça yükselmiştir (Lochner and Monge-Naranjo (2012)). İkinci olarak ise, söz konusu teori bireylerin bir sene daha ilave eğitim almanın getirilerini ve maliyetlerini doğru bir şekilde hesaplayabildiğini varsaymaktadır. Fakat, getirilerin belirsiz ve eğitim yatırımı yapıldıktan uzun bir süre sonra gerçekleşeceği göz önüne alındığında, bu varsayım durumu tam olarak kapsayamamaktadır. Bu kapsamda, insanların getirileri ve maliyetleri doğru bir şekilde hesaplayamaması da eğitim almayı bırakmalarında etkili olmaktadır.

Eğitim bilimleri literatürü incelendiğinde, Hanushek (1986)'da ortaya atılan "Öğretimin Ekonomik Modeli" teorisine göre, öğretim sürecinin girdi, süreç ve çıktı olmak üzere üç ana bileşeni bulunmaktadır. Öğrenci, öğretmen ve diğer kaynakları girdi olarak alan okullar, tanımlanan eğitim süreçleri içerisinde mezuniyet/okulu terk etme çıktılarını üretmektedir. Bu modele göre, öğretim süreci üç hiyerarşik seviyede incelenmektedir; öğrenci seviyesi, sınıf seviyesi ve okul seviyesi. Her üç seviyede de girdiler ilgili süreçlerle çıktılara dönüştürülmektedir. Girdiler, çoğunlukla okullar için değiştirilemez faktörler olmakta olup okullar, çıktıları etkilemede akademin/sosyal süreçler kapsamında etkili olmaktadır.

Lise bırakma oranlarını tahmin etmede kullanılan oldukça fazla sayıda okulla ilgili değişken yer almaktadır. Bir okulun kaynaklarını belirttiği düşünülen öğretmen başına düşen öğrenci sayısı bu faktörlerden birisidir. McNeal Jr (1997), öğretmen başına düşen öğrenci sayısındaki artışın öğrencilerin okulu bırakma olasılıklarında kayda değer bir artışa yol açtığını gözlemlemiştir. Rumberger (1995) de makalesinde, sosyal sınıf farklılıkları ve öğrencilerin dağılımı gibi faktörler göz önüne alındığında dâhi öğrenci-öğretmen oranındaki artışın önemli derecede okulu bırakma riskini artırdığına işaret etmektedir.

Literatürde incelenen bir diğer okulla ilgili faktör ise öğrenci kompozisyonudur. Öğrencilerin sosyal karakteristiklerinin yanı sıra okulu oluşturan öğrencilerin sosyal özelliklerinin birleşimi de öğrencilerin başarısında etkili olmaktadır (Gamoran (1992)). Sander (2001), bir okuldaki İngilizce yeterliliği kısıtlı öğrencilerin ve düşük gelirli öğrencilerin yüzdesindeki artışın okul bırakma oranlarını artırdığını belirtmektedir. Bunun yanı sıra, Asyalı öğrencilerin yüzdesindeki artışın da okulu bırakma oranlarını düşürdüğünü, fakat Hispanik ve Siyahi öğrenciler için herhangi bir kayda değer etki gözlemlenmediğini göstermiştir. McNeal Jr (1997) de azınlık öğrencilerin yüzdesindeki değişimin okulu bırakma üzerinde etkili olduğunu belirtmiştir. Buna karşılık, Rumberger and Thomas (2000) ise öğrencilerin sosyal yapısının okulu bırakma oranlarında etkili olduğunu gözlemlese de öğrencilerin ırksal dağılımı için herhangi bir etkisinin olmadığını belirtmektedir.

Bir okulun sosyal ve akademik ortamının da okulu bırakma oranlarını tahmin etmede önemli bir etken olduğu bilinmektedir. Literatürde okul iklimini sayısallaştırmak ve ölçmek için yaygın olarak kullanılan göstergelerden birisi genel katılım oranıdır. Literatürde, katılım oranı yüksek okulların düşük okullara göre daha az okul bırakma oranlarına sahip olduğu kanıtlanmıştır (Rumberger and Thomas (2000), Christle et al. (2007)).

Tezin ikinci amacı doğrultusunda okul bırakma oranları ile okul seçim mekanizmaları arasındaki ilişkiyi inceleyen literatürün oldukça kısıtlı olduğu görülmüştür. Deming et al. (2014) makalesinde Charlotte-Mecklenburg devlet okullarına kura ile yerleştirme işleminin lisans yerleşimi ve lisenin tamamlanması üzerindeki etkisini çalışmış ve kurayı kazanarak ilk tercihlerine yerleşen öğrencilerin başarı oranında kayda değer bir artış gözlemlemiştir. Lavy (2010) ise okul seçiminde yaşanılan bölgeye dair zorunlulukların kaldırıldığı ve belirli bir derecede serbest seçimin uygulandığı bir liselere yerleştirme uygulamasının daha düşük okul bırakma oranlarına vesile olduğunu gözlemlemiştir.

Amerika'daki her eyalet çeşitli okul bölgelerine ayrılmıştır. Okul bölgeleri, belirli bir coğrafi alanda ulusal eğitim politikalarını uygulaması amacıyla kurulmuş idari birimler olarak tanımlanmaktadır. Okul bölgelerinin coğrafi alanları, çoğunlukla idari bölge sınırlarına göre çizilmekte olup eyalet eğitim birimleri tarafından belirlenmekte olup devlet okulları, okul bölgeleri altında hizmet vermektedir. 2019-2020 öğretim yılı itibariyle Massachusetts eyaletinde 401 adet aktif okul bölgesi bulunmaktadır. Massachusetts eyaletindeki okul bölgelerinde anasınıfı, 1. sınıf, 6. sınıf ve 9. sınıfa yeni geçecek öğrenciler, her okul yılının bahar döneminde kendi okul bölgesi ofislerine okul tercihlerini sıralı olarak bildirmekle yükümlüdür. Başka bir okula transfer olma durumu dışında, diğer sınıflarda bulunan öğrenciler mevcut okullarında eğitimlerine devam edebilmektedir. Öğrencilerin belirttikleri tercih listesine göre okullara tahsis edilmesini sağlayan algoritma ise o bölgenin okul komitesi tarafından belirlenmektedir. Boston hariç Massachusetts eyaletinin diğer tüm okul bölgelerinde Boston Mekanizması (BM) kullanılmaktadır.

BM'ye göre, öncelikle, her okul sadece kendisini ilk sırada tercih eden öğrencileri hesaba katarak kendi sıralamasına göre öğrencilerin okula kayıtlarını gerçekleştirir. İkinci aşamada, ilk aşamada herhangi bir okula kaydedilemeyen öğrenciler ile söz konusu okulu ikinci sırada tercih etmiş öğrenciler arasından seçim yapılır. Bu prosedür, tüm öğrenciler okullara yerleştirilinceye kadar devam eder. Oldukça sık kullanılmasına rağmen BM'de bazı içsel sorunlar bulunmaktadır. Abdulkadiroğlu and Sönmez (2003) makalesinde bazı ebeveynlerin okullar üzerindeki tercihlerini doğru olmayan biçimde sergilemelerinin daha karlı olabileceğini ve dolayısıyla BM'nin adil-dağılım koşullarına (strategy-proofness) uygun olmadığını belirtmektedir. Örneğin, tercih sıralamasında çok talep olan okulları daha aşağıda ve görece daha güvenli okulları daha yukarıda belirten ebeveynler, bu okullara yerleştirmeyi garanti edebilmektedir. Bu durum, gerçek tercihlerini belirten ebeveynlerin çok talep olan okulları ilk sıralara ve güvenli okulları arka sıralara koydukları durumda çocuklarının her iki tercihlerine de yerleşmemelerine sebebiyet verebilmektedir. Bunun nedeni ise çok tercih edilen okul için öncelik sıralamasında geride olması ve güvenli okulun ise doğru tercihlerini sergilemeyen ailelerin çocuklarını kabul etmesi olarak karşımıza çıkmaktadır. Dolayısıyla, BM içsel olarak tercihlerini doğru belirten ailelerin tercihlerini doğru belirtmeyen aileler tarafından zarar görmesi ihtimalini barındırmaktadır. Çalışmasında Abdulkadiroğlu et al. (2006), ailelerin %19'unun bu bağlamda stratejik davranmadığını ve sistem tarafından zarar görebileceğini vurgulamıştır. Bu çerçevede, BM içsel olarak bazı aileleri sistemin açıklarını kullanabilmek adına tercihleri konusunda yanıltıcı olmaya teşvik etmektedir.

Belirtilen sebeplerden ötürü, Boston Okul Komitesi (Boston School Committee) 1999'dan beri kullanılmakta olan BM'yi uygulamayı bırakarak 2006 yılında Abdulkadiroğlu and Sönmez (2003) 'in makalesinde önerilen ve adil-dağılım koşullarına uygun olan Öğrenci Optimal Kararlı Mekanizmasını (Student Optimal Stable Mechanism- SOSM) kullanmaya başlamıştır. SOSM algoritmasına göre öncelikle her okul sadece kendisini ilk sırada tercih eden öğrencileri hesaba katarak kendi sıralamasına göre öğrencilerin okula kayıtlarını "geçici olarak" gerçekleştirir ve diğer öğrencileri reddeder. İkinci aşamada ise, okulların geçici öğrenci seçimi yaptığı set BM'e göre değişiklik göstermektedir. Her okul ilk aşamada reddedilen, tercih listesinde ikinci sırada yer veren ve mevcut geçici kayıtlı öğrencilerini düşünerek geçici kayıtlarını gerçekleştirir ve bir önceki aşamada geçici olarak kaydettiği öğrenci de reddedilebilir durumdadır. Bu süreç tüm öğrenciler nihai okullarına yerleştirilinceye kadar devam eder.

Bu çalışmada, Massachusetts eyaletinden okul seviyesinde veri kullanılmaktadır. Söz konusu veri, Massachusetts İlk ve Orta Öğretim Departmanından (Massachusetts Department of Elementary and Secondary Education- MDoE) alınmıştır. MDoE tarafından yayınlanan çoğu veri okul seviyesinde olup bunun dışındakiler okul bölgesi seviyesindedir. Veri seti, 2003-2004 öğretim yılından 2017-2018 öğretim yılına kadar yıllık olarak verilmiştir. Tez boyunca bir okul yılı ikinci kısmı belirtilerek kullanılmaktadır. Örneğin, 2007 yılı denildiğinde 2006-2007 öğretim yılı belirtilmektedir.

Eğitim teorilerinin belirttiği gibi, bir öğrencinin liseden mezun olma ya da liseyi bırakmasında öğrenci kompozisyonu, okul yapısı, okulun kaynakları ve genel/akademik iklim ve süreçler etkili olmaktadır. Bu kapsamda, MDoE'nin ilgili verileri incelenmiştir. Öncelikle, MDoE liseyi bırakma oranını 9-12. sınıflardaki öğrencilerden mezuniyeti gelmeden ve başka bir okula transfer olmadan bir sonraki sene için kaydını yenilemeyen öğrencilerin toplam 9-12. sınıf popülasyonuna oranı olarak tanımlamaktadır. Beyaz öğrenciler oranı, 9-12. sınıflardaki toplam öğrencilerin içerisinde köken olarak Avrupa, Orta Doğu veya Kuzey Afrika olan öğrencilerin oranı olarak tanımlanmakta olup öğrencilerin ırksal kompozisyonu hakkında bilgi vermektedir. Aynı şekilde, kısıtlı İngilizceye sahip öğrencilerin oranı değişkeni ana dili İngilizce olmayan ve sıradan sınıf çalışmalarını icra edemeyecek olarak belirtilmiş öğrencilerin oranını vermektedir. Öğrenci kompozisyonu hakkında fikir veren bir diğer değişken de ekonomik zorluk çeken öğrenci oranı olarak belirtilmekte olup söz konusu öğrenci aşağıdaki kriterlere sahip olmalıdır;

- Ücretsiz veya düşük fiyatlı öğle yemeği hakkı olmak,
- Aileye yardım anlamında burs almak,
- Yemek pulu hakkı olmak.

Öğretmen başına düşen öğrenci sayısı okulun kaynaklarını belirtmesi anlamında iyi bir temsili değişken olarak MDoE tarafından raporlanmaktadır. Bu değişkenlerin yanında, bir okulun genel/akademik ortamı direkt olarak ölçülemeyen bir değişken olduğundan, ortalama katılım oranı uygulanan yöntemlerin öğrencileri okula gelmeye ne kadar motive ettiğinin bir göstergesi olarak temsili bir değişken olarak kullanılmaktadır. MDoE, katılım oranını en az 20 gün boyunca okula kayıtlı olan öğrencilerin ortalama okula katılım yüzdesi olarak tanımlamaktadır.

Bu değişkenler kapsamında oluşturulmuş veri seti öncelikli olarak 5887 gözlemden oluşmaktadır. Bu veri setinden, yukarıda verilen değişkenlerden en az birinde eksik bilgi bulunan gözlemler çıkarılmış olup çıkarılan toplam gözlem sayısı 266'dır (%4,5). Buna ilave olarak, kalan 5621 gözlem içerisinden, herhangi bir yılda aşağıda belirtilen özelliklere sahip okullar aykırı değer olduklarından çıkartılmıştır;

- Bırakma oranı %70'in üzerinde olan okullar (5 gözlem)
- Öğrenci-öğretmen oranı 100'ün üzerinde olan okullar (5 gözlem)

Belirtilen değişkenlerle bırakma oranları arasındaki ilişkiyi anlamak adına ilk başta yıllar arasındaki farklılıklar göz ardı edilerek sadece 2018 verisine odaklanılmıştır. 2018 verisine ilaveten sonuçların tutarlılığını anlamak için 2004 verisi de kullanılarak sonuçlar karşılaştırılmıştır. Bu kapsamda, aşağıdaki regresyon modeli En Küçük Kareler (EKK) yöntemi ile tahmin edilmiştir;

$$DRP_j = \beta_0 + \beta_1 ATT_j + \beta_2 STR_j + \beta_3 WHI_j + \beta_4 ECO_j + \beta_5 LEP_j + \beta_6 COUNTY_j + u_j \qquad \text{(D.2)}$$

Bu denklemde kullanılan değişkenler aşağıdaki şekilde tanımlanmaktadır;

- DRP_j: Okul j için bırakma oranı (%),
- ATT_j: Okul j için katılım oranı (%),
- STR_j: Okul j için öğrenci-öğretmen oranı,
- WHI_j: Okul j için beyaz öğrencilerin oranı (%),
- ECO_j: Okul j için ekonomik zorluk çeken öğrenci oranı (%),
- LEP_j: Okul j için kısıtlı İngilizceye sahip öğrenci oranı (%),
- $COUNTY_j$: Okul j'nin içinde bulunduğu ilçe.

Tek yıllı model, bırakma oranları ve tahmin edici faktörler hakkında önemli öngörüler sunsa da okullar arası farklılıklar göz ardı edildiği için çeşitli kısıtlamalar içermektedir. Bir okulun bırakma oranlarını etkileyen birçok faktör olacağı düşünüldüğünde, tek yıllı modelin atlanan değişken hatasına sahip olma ihtimali oldukça fazladır. Bu kısıtlardan dolayı, 2004-2018 yıllarını kapsayan bir panel veri regresyon modeli geliştirilmiştir. Sahip olunan veri seti, T= [2004,2018] yılları arasında n=453 okula ait toplam N=5.611 gözlemden oluşmaktadır.

Temel lineer panel veri modeli aşağıdaki gibidir;

$$DRP_{jt} = \alpha_{jt} + \beta_{1jt}ATT_{jt} + \beta_{2jt}STR_{jt} + \beta_{3jt}WHI_{jt} + \beta_{4jt}ECO_{jt} + \beta_{5jt}LEP_{jt} + u_{jt} \quad \text{(D.2)}$$

Burada j=1, ..., n okul indeksini, t=1, ..., T zaman indeksini ve u_{jt} ise ortalaması 0 varsayılan hata terimini göstermektedir. Genellikle hata terimleri, değişkenlerin bağımsızlığı ve katsayılar hakkında bazı varsayımlar yapılmakta olup bu varsayımlara göre çeşitli panel veri modelleri ortaya çıkmaktadır. Yapılan en temel varsayım, katsayıların tüm bireyler ve zamanlar için sabit olmasıdır;

$$\alpha_{jt} = \alpha,\ \beta_{kjt} = \beta_k\ \forall j \in [1, n], t \in [1, T], k \in [1,5] \quad \text{(D.3)}$$

Bu varsayım D.2'deki denklemi aşağıdaki şekilde değiştirmektedir;

$$\begin{aligned} DRP_{jt} &= \alpha + \beta_1 ATT_{jt} + \beta_2 STR_{jt} + \beta_3 WHI_{jt} + \beta_4 ECO_{jt} \\ &+ \beta_5 LEP_{jt} + u_{jt} \end{aligned} \quad \text{(D.4)}$$

D.4'teki denklem ise, bir önceki tek yıllı model ile aynı olmaktadır. Verinin okullar ve zamanlar arası değişkenliğinin göz ardı edildiğinden dolayı, bu modele havuzlanmış (pooled) model denilmektedir.

Okullar ve yıllar arası farklılıkları dikkate alabilmek için, genellikle hata teriminin üç ayrı bileşenden oluştuğu varsayılır, ilk bileşen her bir bireye spesifik olup zamana göre değişmeyen ve ikinci bileşen zamana spesifik olup bireye göre değişmeyen değişkenlerken üçüncü bileşen ise rassal hata bileşenidir;

$$u_{jt} = \mu_j + \lambda_t + \varepsilon_{jt} \quad \text{(D.5)}$$

Bu varsayım ise D.4'teki denklemi aşağıdaki şekilde değiştirmektedir;

$$DRP_{jt} = \beta_j + \lambda_t + \beta_1 ATT_{jt} + \beta_2 STR_{jt} + \beta_3 WHI_{jt} + \beta_4 ECO_{jt} + \beta_5 LEP_{jt} + \varepsilon_{jt} \qquad \text{(D.6)}$$

ve burada aşağıdaki denklem geçerlidir;

$$\beta_j = \alpha + \mu_j \qquad \text{(D.7)}$$

Dolayısıyla en son denklemde, her bir bireyin (okulun) ve her bir zamanın (yılın) kendi sabit etkileri nedeniyle farklı olduğu, farklı β_j ve λ_t değerleri vasıtasıyla dikkate alınmaktadır.

Eğer birey spesifik etlilerin değişkenlerden bağımsız rassal değerler olduğu varsayımı yapılırsa, model Rassal Etkiler (Random Effects) (RE) olarak isimlendirilir, bu varsayım yapılmadığı taktirde ise model Sabit Etkiler (Fixed Effects) (FE) olarak isimlendirilir. Bu tezde lise bırakma oranları, okul-spesifik etkilerin katılım oranı gibi değişkenlerden bağımsız olmadığı düşünüldüğünden, sabit etkiler modeli kullanılarak analiz edilmiştir. Fakat, bu seçimin formal metotlarla test edilmesi gerekmektedir.

Öncelikle havuzlanma testi yapılmıştır, yani D.3'ün doğruluğu test edilmiştir. Test sonuçları güçlü bireysel sabit etkilerin olduğunu göstermektedir. Dolayısıyla, havuzlanmış model sabit etkiler modeli alternatif hipotezine göre reddedilmiştir. Daha sonra, doğru modelin sabit etkiler modeli ya da rassal etkiler modeli olduğunu test etmek için ise Hausman testi kullanılmaktadır. Bu testte hem sabit etkiler hem de rassal etkiler modeli kurulup tahmin edildikten sonra bireysel etkilerin değişkenlerden bağımsız olduğu hipotezi kurulmaktadır. Test sonuçlarına göre hipotez bireysel etkilerin bağımsız olmadığı yönünde reddedilmektedir. Dolayısıyla, veri setindeki lise bırakma oranları için sabit etkiler modeli uygun görülmektedir.

Model sonuçları verilmeden önce, sonuçların anlamlı olması adına model üç ayrı amaç ile test edilmelidir. Öncelikle, aynı bölgedeki okullar için bırakma oranlarını etkileyen ortak faktörler olabileceği düşünüldüğünde, okullar arasında kesitsel ilişki (cross-sectional dependence) beklenmektedir. İkinci olarak, birden çok yılı kapsayan bir veri ile çalışıldığından serisel korelasyon, yani özilinti (serial correlation) beklenmektedir. Son olarak ise, tek yıllı modelde olduğu gibi, farklıserpilimsellik (heteroscedasticity) beklenmektedir. Test sonuçlarına göre bu tezde kullanılan verinin güçlü bir şekilde okullar arası bağlantılı olduğu ve serisel korelasyon (serial correlation) probleminin var olduğu gözlenmiştir. Driscoll and Kraay (1998)'in makalesinden ortaya çıkan mekansal korelasyon tutarlı (spatial correlation consistent) (SCC) kovaryans tahmincilerinin heteroskedastisite, özilinti ve kesitsel ilişki sorunları için sağlam ölçünlü hata terimleri (robust standard errors) verdiği literatürde belirtilmektedir (Hoechle (2007)). Bu nedenle, model sonuçları SCC hata terimlerine göre analiz edilmiştir.

Tezin ikinci amacı kapsamında, Boston'daki ortalama lise bırakma oranlarının mekanizma değişikliğine nasıl tepki verdiğini incelemek için ise, oranlarda gerçekleşen değişimin sadece mekanizma değişikliğinden kaynaklandığından olabildiğince emin olmak amacıyla yapılan analizde sadece 2006 ve 2007 yılları kullanılmıştır. Farkların farkı metodolojisi kapsamında, 2005-2006 okul yılı değişimden önceki yıl ve 2006-2007 okul yılı değişimden sonraki yıl olarak düşünülmektedir. Boston'da olup yeni mekanizmayı uygulayan okullar deney grubu, diğer okullar ise kontrol grubu olarak adlandırılmıştır. Farkların farkı metodolojisinde sağlanması gereken diğer bir varsayım da paralel eğilim varsayımıdır (parallel trend assumption). Temel olarak bu varsayım, uygulama grubu ile kontrol grubu arasındaki farkın herhangi bir uygulama olmaması halinde aynı kalacağını belirtir, yani bu varsayım, herhangi bir değişimin yaşanmadığı durumda bu iki grubun paralel bir eğilim (trend) izleyeceklerini söylemektedir. Bu varsayım için formal istatistiksel bir test olmadığı için, eğilimlerin görsel bir sunumu herhangi bir farkların farkı analizi için gerekli olmaktadır. Bu sebeple, Boston şehri ile paralel eğilim varsayımı kapsamında uyumlu olan okul bölgeleri bulunmuştur. Bu kapsamda, aşağıdaki kestirim yapılmıştır;

$$DRP_{it} = \beta X_{it} + \beta_1 BOSTON_i + \beta_2 2007_t + \beta_3 (BOSTON_i \times 2007_t) \tag{D.8}$$

Burada X_{it} okul i ve yıl t için bağımsız değişkenlerin oluşturduğu matrisi, $BOSTON_i$ okul Boston'da ise 1 değerini alan kukla değişkeni ve 2007_t yıl 2007 ise 1 değerini alan kukla değişkeni göstermekte olup ilgilendiğimiz farkların farkı tahmincisi $\widehat{\beta_3}$'tür.

Sonuçlar incelendiğinde, öncelikle tek yıllı regresyon model sonuçlarına göre katılım oranı bırakma oranlarının güçlü bir tahmin edicisi olarak karşımıza çıkmaktadır. Hem

2018 verisi regresyonu hem de 2004 verisi regresyonu, diğer faktörler kontrol edildiğinde katılım oranlarında meydana gelecek %1'lik bir artışın liseleri bırakma oranlarında yaklaşık %0,75-0,85 seviyesinde bir düşüşe yol açacağını ön görmektedir. Buna ilave olarak tek yıllı model sonuçları, ekonomik zorluk çeken öğrenci oranında gerçekleşecek yüzde 1'lik bir artışın bırakma oranlarını %0,1 oranında artıracağını tahmin etmektedir. Bırakma oranlarına etki anlamında öğrencilerin ırksal dağılımı da istatistiksel olarak anlamlı çıkmış olup öğrenci-öğretmen oranı için 2018 ve 2004 regresyon sonuçları farklılık göstermektedir. Her iki model de kısıtlı İngilizceye sahip öğrenci oranının bırakma oranları üzerinde istatistiksel olarak anlamlı bir etkisinin olmadığını saptamaktadır. Daha önceden de belirtildiği gibi, tek yıllı modeller okullar arasındaki farklılıkları hesaba katmadığı için sonuçlar çok yıllı model ile birlikte ele alınmalıdır.

Çok yıllı sabit etkiler panel veri modeli incelendiğinde ise, katılım oranlarının diğer değişkenlerin yanında okul ve yıl sabit etkileri kontrol edildiğinde dâhi liseleri bırakma oranlarında istatistiksel olarak anlamlı negatif bir etkisinin olduğu gözlenmektedir. Fakat, bu etkinin büyüklüğü tek yıllı modellere göre düşük çıkmaktadır. Çok yıllı model, katılım oranlarında %1'lik artışın lise bırakma oranlarında yaklaşık %0,17 oranında bir düşüşe yol açtığını ön görmektedir. Bunun nedeni olarak ise tek yıllı modellerde okulların farklılığını göz önüne alınmadığı için ortaya çıkan ihmal edilmiş değişkenlerin yol açtığı sapma olarak değerlendirilmektedir. Bunun yanında, öğrencilerin ırksal dağılımı ve ekonomik zorluk çeken öğrencilerin oranı okul sabit etkileri göz önüne alındığında istatistiksel anlamlılığını yitirmektedir. Kısıtlı İngilizceye sahip öğrenci oranı ile öğretmen başına düşen öğrenci sayısı değişkenleri ise tek yıllı modellerde olduğu gibi lise bırakma oranlarında istatistiksel olarak etkili görünmemektedir.

Tezin ikinci amacı kapsamında yapılan farkların farkı model sonuçları analiz edilmeden önce, paralel eğilim varsayımını sağlayan okul bölgeleri grafiksel olarak incelenmiştir. Bu kapsamda, 27 adet okul bölgesinin Boston okullarındaki bırakma oranları ile paralel bir eğilime sahip olduğu saptanmıştır. Bu kapsamda, bu 27 okul bölgesindeki okulların kontrol grubunu oluşturduğu veri seti kullanılarak D.8'deki model tahmin edilmiştir. Model sonuçlarına göre ilgilenilen $\widehat{\beta_3}$ katsayısı beklentilere paralel olarak -0,98 çıkmaktadır. Bu da söz konusu mekanizma değişikliğinin

Boston'daki liselerdeki bırakma oranlarında yaklaşık %0,98 puan düşüşe yol açtığını söylemektedir. Fakat, bu değer istatistiksel olarak anlamlı bulunmamıştır. Çıkan bu sonuçta, paralel eğilim varsayımını sağlayan sadece 27 adet okul bölgesinin olması ve toplam 128 gözlemden 70 adetinin Boston, kalan 58 adedinin de kontrol grubuna ait olması etkili olmuştur. Dolayısıyla numune büyüklüğünün görece düşük ve dengesiz olması istatistiksel olarak anlamlı bir sonuç bulunmasını zorlaştırmıştır. Fakat sonuç olarak, kurulan model mekanizma değişikliğinin lise bırakma oranlarına istatistiksel olarak anlamlı bir etkisinin olmadığını ön görmektedir.

Özetle, bu tez çalışmasında, Massachusetts'teki lise bırakma oranları üzerine çalışılmıştır. Lise bırakma oranlarını etkileyen okullarla ilgili faktörler araştırılmış ve Boston şehrinde uygulanan mekanizma değişikliğinin lise bırakma oranlarında etkili olup olmadığı irdelenmiştir. Bu kapsamda, 2003-2004 ve 2017-2018 yılları arası okul seviyesinde veri kullanılmış olup verilen bir yıldaki herhangi bir okul için katılım oranı, öğrenci öğretmen oranı, beyaz öğrenci oranı, ekonomik zorluk çeken öğrenci oranı ve kısıtlı İngilizceye sahip öğrenci oranı açıklayıcı değişkenler olarak kullanılmıştır. Tezin ilk amacı doğrultusunda, tek yıllı EKK ve çok yıllı sabit etkiler modelleri geliştirilmiştir. Tek yıllı model sonuçlarına göre katılım oranı, ekonomik zorluk çeken öğrenci oranı ve beyaz öğrencilerin oranı, bırakma oranlarını etkilemede istatistiksel olarak anlamlı çıkmıştır. Çok yıllı model sonuçları dikkate alındığında ise, ekonomik zorluk çeken öğrenci oranı ve beyaz öğrencilerin oranı değişkenleri okul sabit etkileri dikkate alındığında etkisini yitirmekte olup katılım oranlarının etkisi de ihmal edilmiş değişkenlerden dolayı oluşan sapmanın etkisini yitirmesi sonucu görece daha mantıklı bir seviyeye düşmüştür. Tezin ikinci amacı doğrultusunda ise, 2006 ve 2007 yılları kullanılarak farkların farkı metodolojisinden faydalanılmıştır. Ayrıca mevcut okul bölgelerinden Boston ile paralel eğilime sahip olanlar kontrol grubunu oluşturacak şekilde kestirim yapılmıştır. Model sonuçları, her ne kadar mekanizma değişikliğinin bırakma oranları üzerinde yaklaşık %1'lik bir düşüş ön görse de bu etkinin istatistiksel olarak anlamlı olmadığı gözlenmiştir.

Bu tez çalışması, literatüre iki farklı yönden katkı sunmaktadır. İlk olarak bu tez, katılım oranlarının bırakma oranları üzerindeki negatif etkisi olduğu savını güçlendirmekte olup okul ve zaman sabit etkileri kontrol edildiğinde diğer değişkenlerin etkileri üzerine kuşku doğurmaktadır. İkinci olarak ise, Boston şehrinde gerçekleşen mekanizma değişikliğinin bırakma oranlarında istatistiksel olarak herhangi bir etkisinin olmadığını saptamaktadır. Her ne kadar bu tezde verilen modeller lise bırakma oranlarına ilişkin var olan bilgi setini ilerletse de tezde sınırlamalar ve iyileştirme potansiyelleri bulunmaktadır. Öncelikle, öğrenci-öğretmen oranı ve katılım oranı değişkenleri temsili değişkenler olarak ele alınmaktadır. Dolayısıyla, her ne kadar bu değişkenlerin ekstra bilgiye sahip olma sorunu çok yıllı sabit etkiler modeli ile belli bir seviyeye kadar çözülmüş olsa da yanlı tahmin olasılığı bulunmaktadır. Bunun yanında, farkların farkı analizinde örneklem büyüklüğünün yeterince yüksek olmaması sonucu işaret olarak doğru çıkan sonuç istatistiksel olarak anlamlı bulunmamıştır. Varsayımları sağlayan ve yeterli seviyede daha çok veri olması durumunda istatistiksel olarak anlamlı bir etki gözlenmesi daha mümkün olacaktır.

E. THESIS PERMISSION FORM / TEZ İZİN FORMU

<u>**ENSTİTÜ /** INSTITUTE</u>

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences ☐

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences ■

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics ☐

Enformatik Enstitüsü / Graduate School of Informatics ☐

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences ☐

<u>**YAZARIN /** AUTHOR</u>

Soyadı / Surname : AKBULUT
Adı / Name : ORHUN
Bölümü / Department : ECONOMICS

<u>**TEZİN ADI /** TITLE OF THE THESIS</u> (**İngilizce** / English) : HIGH SCHOOL DROPOUT RATES IN MASSACHUSETTS: CONTRIBUTING SCHOOL FACTORS AND EFFECTIVENESS OF A CHANGE IN SCHOOL CHOICE MECHANISM

<u>**TEZİN TÜRÜ**</u> **/** <u>DEGREE**:**</u> **Yüksek Lisans** / Master ■ **Doktora** / PhD ☐

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ■

2. **Tez <u>iki yıl</u> süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of <u>**two years**</u>**. *** ☐

3. **Tez <u>altı ay</u> süreyle erişime kapalı olacaktır.** / Secure the entire work for period of <u>**six months**</u>. ***** ☐

***** *Enstitü Yönetim Kurulu kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir.*
A copy of the decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.

Yazarın imzası / Signature .... **Tarih** / Date ………………