The Effect of a Desirable School Zone on Housing Prices: Evidence from a Quasi-Natural Experiment in New Zealand

Peng Sun

TABLE OF CONTENTS

Abstract

Most research on the effect of school quality on housing prices has taken place at city or country level. However, the heterogeneity of each individual school zone or level of schools makes the estimation of a causal relationship difficult since individual houses might face a diverse combination of alternative school choices. In this study, I analyse a quasi-experiment where a desirable public secondary school with no close substitutes unexpectedly reduced its enrolment zone twice over a three-year period. This enables me to conduct a difference-in-differences study to determine the impact of school zone on housing prices. I estimate effects using a variety of specifications, including using all housing sales transactions with controls for housing characteristics vs. housing fixed effects for repeat sales only, and with or without year/group interaction terms that test for parallel trends. My estimation shows the first downsizing decreases housing prices between 2.33% and 15.32%, with the smaller estimates not statistically significant. However, I find the second downsizing decreased prices less, by between +0.48% to -7.76%, with more estimates not significant. The above estimations are robust to changes of pre- and post-treatment periods identification and to clustering standard errors house-level.

Tests in the pre-treatment period suggest parallel trends cannot be rejected for the first downsizing most of time, but could for some comparisons for the second downsizing. As always, such tests cannot prove parallel trends do hold even for the first downsizing. I conclude that the high school's first downsizing very likely had a negative effect on housing price, but the size would be small. Given that this natural experiment should identify an "upper bound" effect, my estimate range tends to be smaller than those found in the previous literature.

Key Words

School housing price premium, school zone, difference-in-differences (DID, DiD, or DD), hedonic house pricing models, valuation of school quality, quasi-experiment

1. Introduction

1.1 Motivation

Access to public schools in most public education systems is dictated by each school's school enrolment zone[1] (OECD 2019). Therefore, a desirable school zone implies access to high-quality education resources. To secure better education opportunities for their children, parents have an incentive to choose to live in the school zone of high-quality public schools. If student places in high-quality schools are scarce, "in-zone" houses might command higher prices than "out-of-zone" houses. Access rights to specific public schools thus can become an essential attribute of each house.

Beyond individual parents, policymakers are often concerned with public schools' enrolment systems as well. Policymakers often seek to achieve equal or equitable access to high-quality education (Maguire 2019; Holsinger et al. 2009; Checchi 2003). However, if there are substantial housing price premiums for access, only rich parents can gain access to good schools, which goes against the policymakers' intention to assure education equality via enrolment zone systems. As a result, the size of school price premiums becomes an indicator of how much government policies on school zoning to pursue equity will be offset by people's sorting reactions into more desirable school zones.

1.2 Research Question

Unlike private goods, access to publicly funded schools is generally rationed by non-price means, such as geographic proximity, queuing, or balloting. For public schools perceived by parents as higher quality, rationing access by geographic enrolment zone naturally leads to higher demand for housing in such zones. Therefore, the scarcity of school places available in those high-performing public schools motivates parents to pay more to acquire such enrolment rights. As a result, the market-

[1] School zones are also called enrolment zones, home zones or catchment areas. In this paper, I will use these terms interchangeably.

clearing price of housing in "good school" zones rises as access rights become capitalized as part of housing value.

Although this "rationing by house price" prediction is theoretically plausible, how does it hold empirically? That is, to what extent do access rights to desirable schools raise the price of housing? This research aims to estimate how much of a premium a “good” school zone adds to housing prices by analysing how the housing market responds to a reassignment of school zones.

More specifically, I try to estimate the price premium for housing in the school zone of Cashmere High School (CHS), a much sought-after public school without close substitutes in the city of Christchurch, New Zealand. From simple supply and demand theory, I expect to see a decrease in housing sales prices for those houses losing access rights to CHS since the alternative high schools they have been reassigned to are less sought-after, with lower average student academic achievement and lower socio-economic status.

1.3 Empirical Approach

In this chapter, I use CHS's recent school rezoning as a quasi-experiment, in an attempt to estimate the causal effect of the loss of access rights to a desirable high school on housing sales prices. I employ a difference-in-differences approach to a hedonic house pricing model to compare the changes in housing prices before and after the school rezoning between houses staying in-zone versus those being excluded.

1.4 Summary of Findings

I find that compared to the base group of houses always remaining inside the CHS zone, losing access to this school zone from the first school zone downsizing has likely caused a decrease in housing prices within the second period of my study (between April 4th, 2018 to November 19th, 2019). This effect

from the first downsizing ranged between 2.33% and 15.32% with the smaller estimates not statistically significant. In other words, the empirical results reveal that the upper bound effect size of a desirable school zone on housing prices could be as high as 15.32%, though not consistent in different specifications. The direction of such effect size is consistent with conclusions from mainstream studies about the valuation of school quality through housing prices. However, no significant effect is found from the second school zone downsizing in my research.

1.5 Main Contribution

This chapter provides a case study of the impact of reductions in the school zone of a much sought-after secondary school on urban residential housing prices in Christchurch, New Zealand. Its contribution is to cleanly identify housing price premiums from school zone using a quasi-experiment and difference-in-differences analysis based on two unexpected school zone downsizings. In addition, this is the first such DID causal inference study to be applied to education and housing in New Zealand. It provides a rare estimate for New Zealand and serves as a point of comparison for researchers interested in school-related house price premiums internationally.

My findings extend the current understanding of the impact of school zone on housing prices. It offers further evidence about current demand-side determinants of housing prices and financial barriers for lower-income families to access sought-after public schools. In addition, for policymakers in the Ministry of Education (MoE), Ministry of Business, Innovation and Employment (MBIE), and the Cabinet, the findings provide insights on how the housing market responds to changes in the supply of educational resources. Such understanding might benefit future policy decisions on housing price management, enrolment scheme reform, and enhancing education equality. Furthermore, this research will also help parents with their decision-making regarding the trade-offs between the school

access rights associated with a house and its other attributes, both for child education and family investment purposes.

1.6 Roadmap of the Paper

This chapter consists of eight sections, including the introduction. The second section reviews the literature about housing price determinants, school quality measurement, school choice models, and previous studies on how school quality affects housing prices. The third section introduces the institutional background regarding school access in New Zealand and the CHS's quasi-experiment context. The fourth section presents the data used in my study. The empirical methodology of difference-in-differences and model identification is illustrated in the fifth section. The sixth section provides my estimation results, while robustness checks are reported in the seventh section. The final section summarizes what I find about the causal effects of school zone on housing price premiums, discusses some limitations of this research, and proposes some ideas about future studies and policy implications.

2. Literature Review

2.1 Review of Housing Price Determinants and Hedonic House Pricing Models

Researchers have been striving to explore the exact determinants of housing prices for decades. In theory, the valuation of housing assets consists of recognizing both tangible and intangible features that collectively determine price. Such features include housing structure, the presence of amenities and disamenities (both existing and expected), socio-economic characteristics of the neighbourhood, as well as the spatial correlation of houses and amenities. Other more indirect factors may also affect housing market participants' expectations, such as monetary and taxation policies, population, education, crime, transportation, and urban planning policy. Empirically, researchers have confirmed that housing prices are associated with a comprehensive list of house characteristics, including but

not limited to house structure, location, natural environment, and neighbourhood (Sirmans et al., 2005).

In urban economics, hedonic models and repeated sales models, and a hybrid of the two are the three most common housing price models (Cho, 1996). In this subsection, I will review hedonic house pricing models as they have been most widely adopted. Lancaster (1966) used hedonic pricing models to decompose the price of a comprehensive good into its components by attaching an implicit price to each element. Rosen (1974) further illustrated that the market equilibrium is reached through individual consumer and producer's heterogeneous valuation of each characteristic embodied in each good. This provides a framework for the hedonic pricing model even though Rosen does not propose a specification of the function.

By regressing house sales price on attributes, researchers can estimate the marginal contribution of each feature of the house to its overall housing sales price. Typically, the hedonic model regresses housing sales price on the physical characteristics of each house, the features of its neighbourhood, the attributes of spatially lagged houses weighted by distance, proximity to amenities, and the attributes of such amenities, such as environmental quality, water resources, air quality, access to public services, etc.

Although hedonic pricing models have become the predominant method for establishing housing valuation, the method faces certain limitations in application. First, as always in econometric analysis, results can heavily rely on model assumptions, specification, and measurement manipulation. For example, the hedonic housing price model is developed on the basis of efficient market assumptions. In addition, hedonic housing price models assume that value measurement for different housing characteristics and amenities is the same for all agents, which does not hold in reality.

As the hedonic price model became the standard method of valuing housing characteristics, there emerged a vast literature employing it (Malpezzi, 2002). For instance, researchers have used this approach to estimate housing price premiums for urban location (Ottensmann, Payton and Man, 2008), land and structure components (Diewert, Haan and Hendriks, 2015), transportation facilities (Seo, Golub and Kuby 2014; Armstrong and Rodriguez 2006), air quality (Anselin and Lozano-Gracia, 2008), the crime rate (Dubin and Goodman, 1982), and proximity to parks and green spaces (Kovacs 2012; Panduro and Veie 2013).

The precise implementation of hedonic housing price models has also evolved with the development of econometric methodologies. Even though Halvorsen and Pallakowski (1981) indicate that the true hedonic function form is unknown, researchers have continued trying to find the most appropriate parametric, semi-parametric or even non-parametric hedonic function form. For example, Follain and Malpezzi (1980b) confirm that semi-log hedonic models are preferable to linear hedonic models. Similarly, Thibodeau (2002) claims an improvement of 20% in the fit of his hedonic model when he incorporates spatial autocorrelation.

In New Zealand in particular, researchers have used hedonic models to study the effects of environmental and location-related amenities and disamenities on housing prices. They have illustrated housing price premiums associated with attributes such as access to views (Rohani 2012, Plimmer 2014, Fillippora 2009, Bourassa et al. 2004 and 2005; Samarasinghe and Sharp 2008), scope and quality of views (Krausse 2015), sunlight exposure (Fleming et al. 2018), and housing price discounts associated with disamenities like proximity to high voltage overhead transmission lines (Bond and Hopkins, 2000), cellular phone base stations (Bond 2007), and noise or congestion from busy roads (Fillippova and Rehm 2009) or Christian churches (Bade et al. 2008).

Relevant to my research, the valuation of access to local public goods such as high-quality schools has also been widely studied by economists. Theoretically, Tiebout (1956) extended the Musgrave-Samuelson public finance theory to local public goods, which predicts that the effect of education services ("better schools") on housing demand ought to be positive. Subsequently, numerous researchers have used house pricing models to understand the role that schools play in property values. Examples include Oates, 1969; Hamilton, 1976; Rosen and Fullerton, 1977; Gurwitz, 1980; Jud and Watts, 1981; Dubin and Goodman, 1982; Hayes and Taylor, 1996; Black, 1999; Downes and Zabel, 2002; Brasington and Haurin, 2006; Clapp et al., 2008; Chiodo et al., 2010. I move now to review the theory and findings of this literature in greater detail.

2.2 Determinants of School Choice

At its broadest level, economics is a study of choice under the assumption that all agents follow the utility maximization principle. Parents choose schools for their children to maximize not just their current utility, but also to offer the best investment in their children's human capital, as a way of facilitating higher future utility. As evidence of this, the 2009 survey result of the OECD's Programme for International Student Assessment (PISA)[2] finds that school quality is vital to all parents surveyed from 11 countries. According to this survey, the top five factors parents consider when assessing school quality are academic achievement, reputation, pleasant environment, safety, and distance from home to school[3]. Potentially, parents' school choice decisions could be based on a comprehensive and thorough consideration of a series of factors of school type (public, private, religious, etc.), school education quality, school transport, teacher quality, curriculum, class size, student scores, publicity of school reputation, school value-added, socio-economic status of peer families, cost of attending, fit of a student with school, etc. The weighting of the above factors would vary due to heterogeneity in parents' preferences.

[2] PISA is a test measuring 15-year-olds' ability to use their reading, mathematics and science knowledge and skills to meet real-life challenges.
[3] Source: http://dx.doi.org/10.1787/888932957498.

Other researchers have also tried to identify those factors that enter into parents' school choice. Using the 1999 National Household Expenditure Survey and 2001 micro-data on SAT test-takers, Belfield (2004) explores determinants of school choice between broad categories of schooling. He posits household utility functions over educational outcomes from four types of schooling: public, private-independent, private-religious and home schooling. Belfield finds that preference over schooling type is highly correlated with a family's economic status. Bifulco et al. (2008) also study the effect on school choice of parents' preferences regarding different aspects of school quality, such as instructional quality, rigor of the curriculum, peer composition, etc. They find that geographic student assignment policies (i.e., school enrolment zones) have an influence on households' residential location decisions. They also argue that school choice tends to increase segregation by race, class and achievement between students from advantaged vs. disadvantaged backgrounds as measured by socio-economic status and parental education.

However, since many of a school's attributes may be unobservable or hard to quantify, parents may simplify their decision-making by prioritizing one or two measurable elements such as school's overall (average) graduation or achievement rates, or the socio-economic status of students enrolled. Ericson (2017) finds in his survey paper that parents consider holistic factors of academic rigor, safety, extracurricular activities, teacher quality, religious or moral instructions, class size and overall fit for their children.

2.3 School Quality Measures and Their Effect on Housing Prices

Applied researchers have used numerous ways of proxying for school quality when examining its effect on housing prices. Before the 1990s, researchers mainly used per student expenditure at the school or district level (e.g., Oates, 1969; Hyman and Pasour, 1973; Rosen and Fullerton, 1977). Occasionally, researchers used a subjective survey indicator of school quality (Ridker and Henning, 1967). More

recently, researchers have focused instead on output measures, such as widely-adopted test scores (Li and Brown, 1980) from different education levels, such as national standardized exams, or state-level math exams for particular grades (Black, 1999). Some have tried to convert student gross test score measures into value-added by controlling for prior achievements (Brasington1999; Gibbons et al. 2013). Though less frequent, other researchers have employed other indicators such as student-teacher ratio (Card and Kreuge, 1992), peer effect (Wang et al., 2020), number of graduates entering a specific elite university (Bae and Chung, 2013), school ratings (Figlio and Page, 2003), or each school's academic performance index (Ye, 2017). Overall, researchers have found that school output indicators outperform school input indicators in explaining housing prices in general (Hayes and Taylor, 1996; Downes and Zabel, 2002; Crone, 2006; Seo and Simons, 2009). However, as parents' evaluation of school quality is holistic, using the measurement of certain aspect of school characteristics to proxy for overall school quality or parental preferences may omit relevant information in quantitative analysis.

Changes in measures commonly used to proxy for school quality are also reflected in the research on school quality effects on prices. For example, Jud and Watts (1981) study the relationship between average scores of 3rd-grade students on state standardized reading tests and housing prices in one school district in Charlotte, North Carolina, USA with a limited number of neighbourhood characteristics controlled for. Hayes and Taylor (1996) explore the relationship between the value added of a school and house values with a sample of 288 houses. Brasington and Haurin (2006) test the alternative effect of school expenditures vs. test scores on housing sales prices. By regressing 77,578 house sales covering 310 school districts in seven urban areas of Ohio in 2000, they find that school value-added to student achievements has little effect on housing sales price compared to the effect of both school expenditure and proficiency test scores.

2.4 Limitations of Early Papers Testing Effects of School Quality on Housing Prices

In early research about the effect of school quality on housing prices, economists tended to overestimate schools' effect on house prices due to a lack of proper control for neighbourhood characteristics. Among those initial efforts of isolating school effect on housing prices, there are a few common issues in data used, model setup and estimation methods. First, instead of actual market value of houses, some research used appraisal prices or insurance valuation as proxy for house sales price (Oates 1969, Atkinson and Crocker 1987). Second, due to the lack of panel data, researchers mainly used cross-sectional regression (Ind and Watts 1981, Haves and Taylor 1996). For example, Oates (1969) concludes that a $50 increase in expenditure per pupil increased house prices by roughly $1200 by regressing housing appraisal values on the weighted per pupil expenditure of the nearest public school based on a cross-sectional dataset from the American state of New Jersey. Third, the size of housing sales datasets used was usually quite small. For example, Walden (1990) finds that school districts with magnet schools have less capitalization in housing value, but based this estimation on 598 house sales records in the year of 1987 from a single county in North Carolina.

2.5 Development in Estimation Methods in Papers Testing Effects of School Quality on Housing Prices

Among the vast body of studies about the effect of school quality on housing prices, researchers have most often employed the framework of hedonic pricing models, and their estimation approaches have gradually evolved from simple cross-section OLS regression to panel fixed-effects models incorporating spatial autocorrelation. One drawback of simple cross section hedonic pricing models is that they cannot address the potential endogeneity of school quality measures. This is because different neighbourhoods can differ in unobserved ways and parents self-select into neighbourhoods. More recent studies have made use of boundary discontinuities, or difference-in-differences, or instrumental variable methods, or have followed event interventions such as education reforms or school zoning changes, to try to untangle the pure effect of school quality from numerous other

factors affecting housing prices. The aim of these newer identification strategies has been to establish a clear causal connection between school quality and housing prices.

Black (1999) pioneered the use of a boundary fixed effects model in the estimation of school quality impact on housing price by comparing houses on both sides of school district boundaries. She bypasses the neighbourhood heterogeneity of houses by examining the prices differences of suburban Boston house prices within .15 miles of elementary school attendance district boundaries. Compared to traditional hedonic house pricing models, Black's hedonic regression with boundary fixed effects reduces the school effect to half of its previously estimated size. In particular, she finds that a one standard deviation increase in test scores of elementary schools is associated with a 2.1% percent increase in house prices.

Since the publishing of Black's (1999) paper, many economists have followed the trend of incorporating a fixed effect from individual school zone or school district boundary discontinuities into their school quality-housing prices studies. Also in the last twenty years, with a focus on causal inference of school quality effect on housing prices, many advanced econometric research methods have been widely applied in this field, such as incorporating spatial autocorrelation, structural modelling, instrumental variables, mixed models, regression discontinuity, difference-in-differences, etc. For instance, research from Brasington and his co-authors started to include distance to CBD in their school quality and housing price study in their 1996 paper under a hedonic pricing framework, and advanced to include spatial autocorrelation in their 2000 paper. Based on a school catchment area adjustment of the Vancouver School Board in Canada in 2000, Ries and Somerville (2004) find significant effects of secondary school quality on housing prices using a difference-in-differences method with both cross-sectional hedonic regressions and repeat sales analysis. Similarly, using 1,173 house prices from metropolitan Chicago, Downes and Zabel (2002) conclude that higher average levels of school achievement raise house values and confirm that households value achievement test

outcomes. Similar difference-in-differences analyses have been applied by Haisken-DeNew et al. (2017), Wen, Xiao and Zhang (2017), Huang et al. (2020), and Han et al. (2021) in their school quality capitalization into housing prices study based on various quasi natural experiments of education policy changes.

2.6 Review of Results on the Effect of School Quality on Housing Prices

Most research has found that access to high-performing schools has a positive effect on housing prices (see for example Barrow 2002, Figlio and Lucas 2004, and Bayer et. al 2007). Clapp et. al (2007) discover a 1.3-1.4% increase in housing prices for a one standard deviation increase in student performance by using a 10-year dataset of 8th grade math scores in Connecticut, USA. Dougherty et al. (2009) find that, after controlling for neighbourhood characteristics, proximity to attendance boundary, and student minorities, house purchasers are willing to pay a 1.9% higher price for an increase of one standard deviation in test scores. Other papers have found larger magnitudes of effects.

Outside the United States, research from Europe has also revealed a positive effect from school quality on housing prices, with the average effect size found ranging between 2 to 7 percent for a one standard deviation increase in a school quality proxy factor (Gibbons and Machin, 2006; Brasington and Haurin, 2006; Fack and Grenet, 2007; Machin and Salvanes, 2007; and Davidoff and Leigh, 2007). For example, Machin and Salvanes (2007) find that housing prices fall by 2-4 percent when an enrolment policy changed from zone-based enrolment to open enrolment, based on secondary test score data from Oslo, Norway.

Among other relevant studies, Leech and Campos (2003) identify a 16 to 20 percent housing premium for access to two elite secondary schools on housing prices in Coventry, UK with neighbourhood characteristics controlled. Their findings represent an upper bound of the school quality effect size on

housing prices in the literature I have identified. Brehm (2017) arrives at the opposite conclusion that access to charter school does not affect house price after studying house sales in Los Angeles County, California, from 2008 to 2011.

2.7 Meta-Analysis and Literature Survey Results about School Quality Effects on Housing Prices

As the number of individual studies of school house price effects has grown, survey papers and meta-analysis studies have emerged in an attempt to provide a general picture of the literature's findings. Meta-analyses differ from survey papers in using the key estimated effects and characteristics of each individual study to run a regression to calculate overall average effects.

Nguyen-Joang and Yinger (1999) have surveyed 50 papers on the capitalization of school quality into house values covering the period between 1999 to 2010, while Black and Machin (2010) have surveyed 54 papers on the same topic but covering a wider time span. Among the literature covered in both survey papers, only less than five papers find no effect or non-conclusive results. To sum up, most research has identified a strong positive association between school quality and house price.

As for meta-analysis studies, Turnbull and Zheng (2019), Yadavalli and Florax (2020), and Zhang et. al (2020) cover a majority of school quality capitalization studies carried out in the United States and China.

Turnbull and Zheng (2019) conduct a meta-analysis based on 56 papers. Their empirical conclusion is that variation in school quality measures used by individual studies is the key cause for mixed results regarding the size of the effect of school quality on housing prices. Among the major quality measures, peer effect and value-added test scores show lower level of capitalization effect compared to other output- or input-based measures. They do not argue that any single measure dominates. They also finds that region matters in the estimation results as school quality effects in the South of the United

States are weaker than in other areas of the country. Interestingly, Turnbull and Zheng also find that by comparing results from both boundary fixed effects and neighbourhood fixed effects meta regression, that econometric methods do not appear to be driving results. However, control for neighbourhood amenities consistently reduces the capitalization significance from school quality.

Yadavalli and Florax (2020) also carry out a meta-analysis based on 48 underlying studies from the United States. They find that measures of primary/secondary school test score and expenditures per pupil stand out among a bundle of eight school quality proxies as having significant effects on house prices. The other five measures considered are value-added, peer racial and socio-economic composition, and pupil/teacher ratio. They also find that model specification and number of school quality variables embedded in the model can affect the magnitude and direction of the effect size. Overall, Yadavalli and Florax find that primary test scores, secondary test scores and expenditures all have significant effect on housing prices. But surprisingly, the effect direction and magnitude were not conclusive because the authors believe that the primary studies used may have been mis-specified, and that the inclusion of multiple school quality factors complicates model interpretation. However, results from their Fischer's Z-transformation model do indicate that secondary test scores have a more significant school quality effect on housing prices than primary test scores.

Finally, by synthesizing 38 empirical studies on education capitalization from 2006 to 2017 in China, Zhang et al. (2020) confirm that based on a random effects meta regression model, the effect of schools on housing premium in the compulsory education stage (Year 1-9) is larger than in the non-compulsory education stage (Year 10-12). School quality and whether the house lies within a school zone boundary have significant impacts on housing premiums beyond distance between house and school, or how many schools a particular house is close to. Zhang et al. also conclude the heterogeneity in individual city housing markets and the distribution of schools in each city leads to the differences in school effect size in different studies, as each study focuses typically on one city.

2. 8 New Zealand Papers on School Quality Impact on Housing Prices

In New Zealand, the effect of school zones on housing markets has frequently been discussed in the mass media (Botting, 2018; Wilkes, 2020). Economists have also made quantitative efforts to identify the impact of school quality on housing prices in the past two decades (McClay and Harrison, 2003; Rehm and Filippova, 2008).

Media reports tend to exaggerate the housing price premium associated with desirable school zones. For example, Edmunds (2017) reports that economists from Homes.co.nz, a local real estate website, have calculated school zone premiums in housing prices by comparing the mean housing price of in sought-after school zones with that of houses in the same city outside of such zones. The premiums estimated for 22 prestigious high schools in 6 cities of New Zealand range from 1.01% to 90.52%[4]. With a premium of 90.5% over housing prices in the rest of Auckland, Epsom Girls' Grammar School was estimated to have the highest premium.

However stunning this result, simply comparing house prices in one school zone with those outside it cannot disentangle the impact of school zone on house prices, as in-zone and out-of-zone neighbourhoods can differ in many other ways such as house conditions, community quality, neighbourhood safety, the convenience of transportation, and quality and quantity of many other amenities. Particularly if these factors correlate positively with school desirability, overlooking them could tremendously exaggerate the effect of school zone on housing prices.

[4] The specific premiums were: Epsom Girls' Grammar(90.52%), Glendowie College (63.43%), Takapuna Grammar School (61.29%), Macleans College(58.75%), Selwyn Collge (49%), Rototuna Junior High School (41.55%), Rangitoto College (41.55%), Auckland Grammar (41.21%), Westlake Girl's High School (37.72%), Albany Senor High School (37.31%), and Christchurch Boys' High School (35.85%).

Unfortunately, despite widespread interest, as of the time of writing, only four academic papers have studied the impact of school quality on housing prices in New Zealand. McClay and Harrison (2003) employ OLS regressions to estimate the housing price premium of popular state secondary school zones with sales data for 514 houses in Christchurch, New Zealand. They find that the premium from the desirable Christchurch Girls High School zone could be as high as $203,000, or 53 percent of the mean price. This research gained extensive publicity as an early study, but subsequent researchers noted the likelihood of omitted variable bias as it lacked controls for neighbourhood characteristics or spatial autocorrelation. Gibson et al. (2005) conduct another study using a similar dataset to McClay and Harrison, but add control for spatial autocorrelation using spatial lags, and find the premium for the same Christchurch school falls to $77,000. As with McClay and Harrison, however, Gibson et al. omit controls for environmental and other public amenities that could be affecting housing prices. The effect of geographical location of houses relative to schools has been further explored in New Zealand by Rehm and Filippova (2008), using a larger dataset of housing transactions from downtown Auckland. Their research focuses on identifying the heterogeneous effect of housing price premiums for different suburbs over different time periods. Using suburb dummy variables in their hedonic models, Rehm and Filippova find that the effect of school zoning on housing prices is not uniform. In particular, premiums of houses at the peripheral areas of a school zone could be diminished due to uncertainty regarding potential boundary changes.

Finally, based on a 12-month (10/2004 to 10/2005) housing transaction dataset in Christchurch, Gibson, and Boe-Gibson (2014) use NCEA pass rates for the measure of school quality a house has access to. They find that a one percentage point increase in the weighted average Level 1 NCEA pass rates is associated with an increase of housing prices by 0.7%, all else constant. Alternatively, a standard deviation increase in the mean of Level 1-3 pass rates is associated with 6.4% higher housing price after controlling for spatial autocorrelation of houses, housing characteristics, and neighbourhood demographic features.

There is an institutional advantage in using New Zealand data to study the school quality effect on housing prices. Unlike school funding systems of the United States and many other European countries, public schools of New Zealand are almost entirely central government-funded. Therefore, the study of school quality effect on housing prices faces less endogeneity risk caused by the socio-economic status of households in the region of a school, as schools receive no direct funding from them (Gibson and Boe-Gibson, 2014).

2.9 My Contribution to the Literature

Judging by the number of papers I have reviewed, it seems there have been fewer papers on this topic over the last ten years. Thus, one motivation for this paper is to provide more recent empirical evidence about how school quality affects housing prices, while still using a strong causal inference approach. My research adds to the literature by looking at a case study with a clear causal identification as a desirable school zone has been downsized, and there is a sharp gap between the quality of that school and the alternative state school for affected households. The strength of this research lies in its bypass of the complexity in gauging school quality and neat identification of a pure causal effect of school zone change to the change of housing sale prices by using the econometrics of a difference-in-differences setup for an arguably exogenous event. It also provides a plausible upper bound estimation of the school zone effect on housing prices focusing with data collected from a residential area of over 100km^2 where many possible confounding factors in the housing market can be excluded. In addition, this paper provides rare look at school effects on housing price in New Zealand.

3. Housing and Education Provision in New Zealand

3.1 The New Zealand Residential Housing Market

Housing is a central part of the New Zealand economy and accounts for around half of the assets of New Zealand households (calculated with data from the Household Balance Sheet, compiled by the Reserve Bank of New Zealand)[5]. On March 31st, 2021, Stats NZ reported an estimate of 1.95 million private dwellings and 1.87 million households in the country.[6] With the Covid-related fall in tourism, the property industry has recently been recognized as the largest industry of New Zealand, directly contributing $41.2 billion to its economy and accounting for 15% of its national GDP (Property Industry Impact Report 2021[7]).

Due to the increase in housing demand and decrease in housing affordability, the census shows that the home ownership rate of New Zealanders has dropped from its relatively high peak of 73.8% in the 1990's to a recent low record of 64.5% in 2018. Meanwhile, there has been a trend away from building single family houses to multi-unit dwellings (either terrace houses or apartments). About 40 percent of all new consented dwellings are multi-unit buildings as of mid-2019 (*Housing in Aotearoa: 2020*[8]*).* Generally speaking, the driving force of housing demand comes mainly from first-time homebuyers, holiday homebuyers and property investors. Overall housing demand has also demonstrated a trend of an increase in number of households but a decrease in number of people per household (*Housing in Aotearoa: 2020)*. On the housing supply side, there is a consistent recognition of a shortage, but there are conflicting opinions about whether it is specifically a comparative shortage of affordable entry-level houses or simply an overall absolute shortage of supply. Johnson et al. (2018) identify an overall shortage of 28,000 dwellings in Auckland despite high rates of housing construction, while Tookey (2017) argues the real shortage lies in affordable housing which enables more people to become homeowners. However, data from StatsNZ show that the number of new homes consented

[5] https://www.rbnz.govt.nz/statistics/c22.
[6] https://www.stats.govt.nz/information-releases/dwelling-and-household-estimates-march-2021-quarter.
[7] https://www.propertynz.co.nz/industry-impact-report.
[8] https://www.stats.govt.nz/reports/housing-in-aotearoa-2020.

nationwide in the year ended October 2021 was a historical peak of 47,715, 26% more than that from the previous year, or an equivalent of 9.3 homes consented per 1000 residents of New Zealand[9]. Once completed, the increase of new homes should significantly solve the supply shortage crisis.

There are two major housing price indices in New Zealand: the Real Estate Institute of New Zealand's Housing Price Index (REINZ HPI) and the QV House Price Index[10]. REINZ's Residential Statistics Report for October 2021 indicates that housing market values nationwide have lifted substantially over the previous twelve months, by 25.7% in Auckland, and by 33.3% in the rest of the country. As of October 2021, the median housing price in New Zealand was $890,500, and that of the country except Auckland was $753,000. In contrast, the QV House Price Indicator reports a national average housing price of $1,002,153, with a 27% national growth rate over the past 12 months as of October 2021.

3.2 The New Zealand Education System and Enrolment Scheme

In New Zealand, the pre-tertiary education system consists of thirteen grades or year levels divided into primary for Years 1-8 and secondary for Years 9-13. As of July 1st, 2020, there were a total of 2536 schools in New Zealand, among which 463 offered Years 9 to 13 secondary education (MoE, 2019). According to the Organisation for Economic Cooperation and Development (OECD), New Zealand has above-average education attainment and labour market outcomes compared to the rest of the world (OECD, 2021). The most recent Programme for International Student Assessment (PISA) results indicate that New Zealand students have higher performance in reading, mathematics and science compared to the average of OECD countries (PISA 2018).

[9] https://www.stats.govt.nz/news/new-home-consents-remain-high-in-april.

[10]REINZ HPI is developed in partnership with the Reserve Bank of New Zealand based on up-to-date unconditional housing transaction data submitted every month by its over 14,000 real estate members, most of whom are holders of the Real Estate Agents Authority (REAA) licenses. QV House Price Index is a NZ house price movement measurement index released by Quotable Value (QV), which provides property valuation and rating service nationwide.

3.2.1 Schooling Choices for New Zealand Parents

Parents in New Zealand have four education options for their children: state schools, state-integrated schools, private schools or homeschooling. State schools are fully owned and funded by the government. State-integrated schools are also partially government-funded but with a special character, such as Catholic schools. State integrated schools teach the national curriculum plus elements in keeping with their character, charge moderate fees to cover some costs, and can select students in keeping with their character. Private schools have greater autonomy in curriculum but receive very limited public funding and charge substantial fees.

In practice, only a comparatively small percentage of student are enrolled in state-integrated, or private schools, or home-schooled. As of July 2020, the share of students going to state, state-integrated, and private schools or home schooling was 84%, 10%, 5% and 1%, respectively[11]. Virtually all private schools and most state-integrated schools have no enrolment scheme, so parents usually do not need to consider their home location as a condition for enrolling into such schools as long as they decide to accept their higher tuition cost. However, the MoE has a maximum roll control on state-integrated schools so many of them operate waiting lists for interested families.

3.2.2 Rationale for School Enrolment Schemes

The setting of an enrolment zone for a desirable public school is a common practice across the world, even though the execution and administration of enrolment zone rules differ across jurisdictions. The purpose of school zones is to manage and adjust aggregate demand for the education resources provided by individual public schools.

In New Zealand, the Education and Training Act 2020, a national-level legislation, requires that "[e]very domestic student is entitled to free enrolment at any state school during the period beginning

[11] https://www.educationcounts.govt.nz/statistics/.

on the student's fifth birthday and ending on January 1st after the student's 19th birthday" (Education and Training Act 2020, S33(1)). In order to ensure every student has the right to attend a state school, the Secretary of the MoE has the responsibility to develop enrolment schemes for popular schools with the potential for overcrowding. This is done to ensure fair and transparent enrolment selection when necessary, and to optimize the usage of the existing networks of state schools (Education and Training Act 2020, S71.1). According to Schedule 71.3, a state school's enrolment scheme must define a "home zone" for the school and identify any special program offered by the school and the criteria by which students are to be accepted into any special program that bypasses the enrolment scheme (Education and Training Act 2020).

According to Schedule 20.1 of the Education and Training Act 2020, a state school's home zone, also called school zone or catchment zone, is defined by concrete geographic boundaries that allocate each address either within or outside the home zone.[12] Section 74.1 regulates that a person who lives in the home zone of a state school is entitled to enrol at that school. Accordingly, if excess capacity exists, students who live outside a school's home zone can be offered places in a state school in a nationally-defined order of priority. Siblings of current and previous students are to have priority over other applicants. If available places are less than demand within the priority group, selection must be done through a ballot (Schedule 20.3 Education and Training Act 2020). Thus, to guarantee enrolment in state or state-integrated schools, students need to reside within the home zone of the school. If parents reside outside a school's enrolment zone, they may also enter their child in a ballot to enrol in that school, if additional spaces exist.

Statistics from the MoE indicate that among a total of 1484 state and state-integrated schools, 934 schools had enrolment schemes in 2018 (New Zealand Schools 2019, MoE). Among them the earliest

[12] Note that the zones of different schools of the same education levels are generally mutually exclusive with rare overlapping. Exceptions of overlap can be found in single-gender schools and state-integrated schools, which are not included in the scope of this research.

batch of enrolment schemes were set up in 1999. A majority of these 934 schools have adjacent schools with enrolment schemes, suggesting excess demand for state schools is spatially correlated. According to the National Education Growth Plan (NEGP) developed by the MoE in 2018, the Government is planning for 100,000 additional learner places in high-growth areas to 2030.

In order to deal with changing school capacity pressures caused by uneven growth, the MoE generally adopts one of three responses. It establishes enrolment schemes for yet more existing schools, adds classrooms to schools with existing enrolment schemes, or builds new schools (New Zealand Schools 2019, MoE). The MoE has also gradually taken control of enrolment schemes on a regional level from individual schools, over such issues as determining excess capacity that can be made available via ballots. This has been justified on the basis of achieving a more equitable access to schooling across racial and socio-economic divides (Wiles, 2020), though it also encourages enrolment in publicly funded under-utilized schools. This has resulted in a recent increase in the once-rare adjustments to enrolment schemes. For example, the secondary school that will be the focus of my study, Cashmere High School, had an enrolment scheme from 2004 that did not change until 2018. According to MoE's Enrolment Scheme Master, a data tool updating changes of Enrolment Scheme Home Zones and geocode data of new zones, just under 5 percent of schools have had changes in their enrolment schemes between 2018 and 2020.

3.2.3 Decile Ratings of New Zealand State and State-Integrated Schools [13]

Since the 1990's, the New Zealand MoE has assigned public schools a "decile" from 1 to 10 based on the relative socio-economic status of the census meshblocks [14] of their student families. The calculation of decile is based on five socio-economic disadvantages indicators: low household income, lowest skill occupation, household crowding, lack of adult education qualification and whether

[13] For a detailed background about the decile system, see https://www.education.govt.nz/school/funding-and-financials/resourcing/operational-funding/school-decile-ratings/.

[14] The meshblock is a minimum aggregated census statistics unit in New Zealand, each consisting of roughly 50 households.

families receive income support. These are taken from the meshblocks of children attending a school, weighted by the number of students from each meshblock. The MoE uses the decile rating system to target modest additional fundings to state and state-integrated schools who serve a greater portion of disadvantaged students. However, though designed to address disadvantage, parents often use a school's decile ranking as an imperfect indicator for school quality. Parents typically prefer to enrol their children into high decile schools rather than low decile schools. As evidence for this, the number of student enrolments in higher decile (Decile 6-10) schools grew continuously from 2013 to 2019 while it fell at lower decile schools (Education Counts[15]).

3.2.4 NCEA Graduating Qualifications

First introduced in 2002, New Zealand's National Certificates of Educational Achievement (NCEA) are nationally set qualifications for senior secondary school students. There are three levels of NCEA certificates—Level 1, Level 2 and Level 3—corresponding with student achievement in Years 11 to 13 respectively. In New Zealand, general high school graduation requires NCEA Level 2, while direct advancement to university requires NCEA Level 3 or equivalent. Students' NCEA results are used for university and polytechnics admission and by employers. Therefore, along with decile ranking, parents in New Zealand tend to use each school's average NCEA performance as a proxy for its quality.

3.3 Christchurch and Cashmere High School

3.3.1 Christchurch

With a population of 369,000 (2018 Census), Christchurch is the second largest city in New Zealand and the largest city on the country's South Island. Compared to other areas of New Zealand, housing prices in Christchurch have tended to be more stable and risen less rapidly before mid-2021. In the ten years prior to this, the housing supply in the city has vastly increased following its loss of housing capacity caused by the 2010/2011 earthquakes. In October 2021, the median house price in

[15] https://www.educationcounts.govt.nz/statistics/6028.

Christchurch City was $668,000. There are 18 state or state-integrated secondary schools in the urban area of Christchurch City.[16] In addition, there are another 8 state or state-integrated composite schools providing some or whole grades of secondary education in Christchurch[17].

3.3.2 Cashmere High School (CHS)

Located in the affluent southwest, Cashmere High School (CHS) is the second largest secondary state school in Christchurch. CHS is a Decile 9 school based on the current decile ratings from the MoE. As a Decile 9 school, CHS has enjoyed a high reputation in the city. The school has a total student roll of 1845 as of July 1st, 2020. [18] Figure 3.1 shows CHS has a consistently growing student roll. Its Year 9 in-zone enrolment and all-grade out-of-zone application numbers are illustrated in Figures 3.2 and 3.3 respectively. According to the Education Review Office's (ERO) Education Review Report, CHS "consistently meets national expectations for at least 85% of students to leave with NCEA Level 2 or above" (CHS Education Review Report 2019). The ERO's overall evaluation judgement of CHS's performance in achieving valued outcomes for its students is "[w]ell placed". In recognition of its high-performance, the ERO only reviews CHS every six years.

Despite the common positive spatial correlation of decile ratings of schools, it is relevant to note that adjacent state or state-integrated schools to CHS have lower decile ratings. Other secondary schools close to CHS are state schools Hillmorton High School (5.4km away, Decile 4), Linwood College(7.7km away, Decile 3), and state-integrated Hillview Christian School (3.9km away, Decile 6).

---

[16] Avonside Girls' High School, Burnside High School, Cashmere High School, Catholic Cathedral College, Christchurch Boys' High School, Christchurch Girls' High School, Hagley Community College, Hillmorton High School, Hornby High School, Linwood College, Mairehau High School, Marian College, Papanui High School, Riccarton High School, Shirley Boys' High School, St Bedes College, St Thomas of Canterbury College, and Villa Maria College.

[17] Hillview Christian School, Emmanuel Christian School, Middleton Grange School, Haeata Community Campus, Rudolf Steiner School (Chch), Te Pa o Rakaihautu, TKKM o Te Whanau Tahi, and TKKM o Waitaha.

[18] Source: Education Counts (www.educationcounts.govt.nz), which is the public education data and statistics sharing website provided by the Ministry of Education, New Zealand.

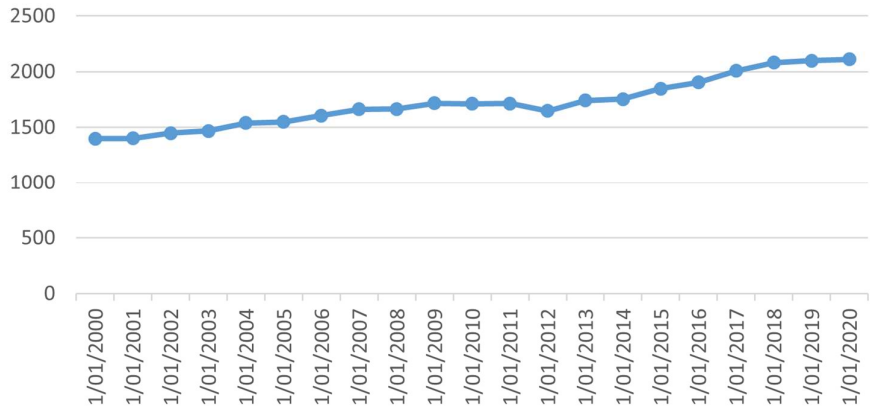


Figure 3.1 2000-2020 Student Roll of Cashmere High School[19]

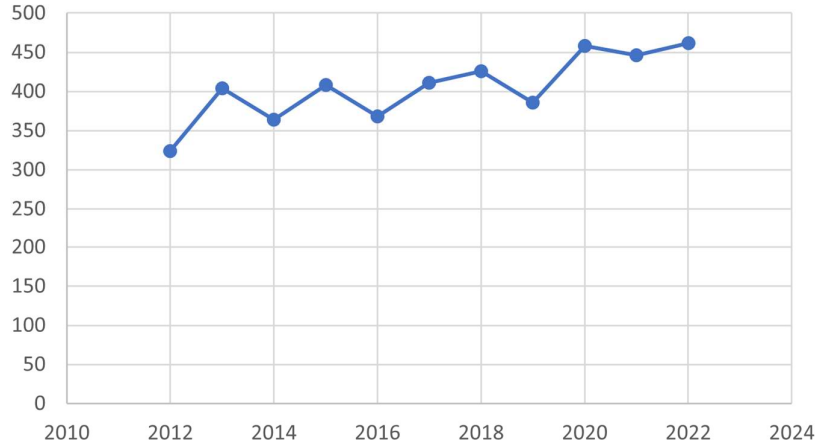


Figure 3.2 2012-2022 Cashmere High School Year 9 In-Zone Student Numbers[20]

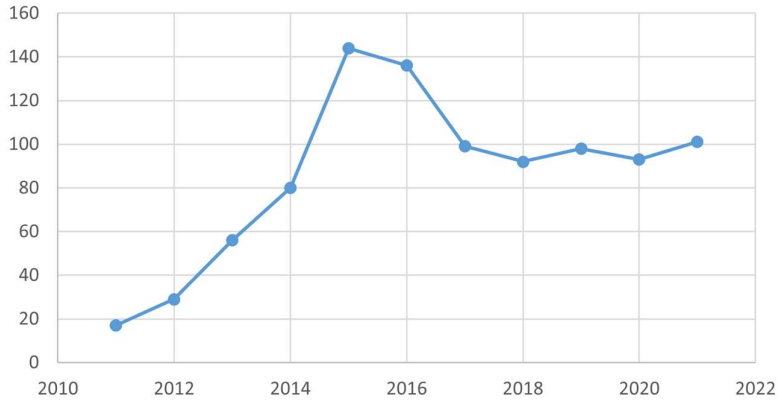


Figure 3.3 2011-2021 Cashmere High School Out of Zone Application Numbers[21]

[19] Source: Education Counts.
[20] Source: Cashmere High School Office.
[21] Source: Cashmere High School Office.

3.3.3 Timeline of CHS School Zone Downsizings

The previous school enrolment scheme of CHS before its recent changes was confirmed by the MoE on July 26th, 2004, and has remained unchanged until April 2018. Its school zone is adjacent to Hillmorton High School in the west and Linwood College in the east. Thus, the households cut out of the CHS zone tended to be reassigned to the former in the first downsizing, and to the latter in the second downsizing. (See Figure 3.4.)

In October 2017, KPMG and the MoE released a report entitled *Greater Christchurch Secondary School Enrolment Review: A Case for Change*, which suggested a number of changes to the secondary schooling network of Christchurch to ensure the optimal use of the network in the future and to prevent overcrowding at more popular schools. The report identified that CHS would need to amend its school zone because of overcrowding. Based on this suggestion, the MoE proposed an enrolment scheme change to the Board of Trustees (BoT) at CHS on March 3rd, 2018, following which the BoT notified its contributing schools and community boards of its student body about the potential changes and timeframe of public consultation, on April 4th, 2018. Thereafter, a public notice was printed in the local newspaper *The Star* on April 12th, 2018, releasing the upcoming change to CHS zone to the public. The BoT decided on a new school zone on April 26th, 2018 after incorporating community feedback, though the final zone was almost the same as the proposed changes given in the public notice. Then an approval checklist was issued by the MoE on April 27th, 2018. The first change to the CHS Zone came into effect on January 1st, 2019.

Slightly over one year after its first downsizing, on September 26th, 2019, CHS received another official notification from the MoE about reducing its enrolment scheme further due to the continued increase of the school roll. However, this unexpected second change proposal aroused dramatic objection among parents and house owners affected by the potential downsizing. After the proposal was released on November 20th, 2019, several community meetings were held during the consultation

period which attracted parents of current and prospective students and house owners in the current zone. A typical comment received during the consultation was that the houses that would be no longer in-zone for CHS would drop in property value. The BoT's Response to that comment was "[h]ouse prices are not a consideration for the school or the Ministry when considering an amendment to an enrolment scheme." The new CHS Enrolment Scheme was released to the public on April 17th, 2020, and came into effect on January 1st, 2021. Several Facebook groups were organized to challenge the new zoning, with the largest one “Reset the Rezone” having over 100 members. As a result of such reset efforts, a petition to review the proposed CHS rezoning was submitted to Parliament on April 28th, 2020, but that effort did not change the rezoning decision. [22] Table 3.1 illustrates a detailed timeline for both school zone downsizing developments. Figure 3.4 presents the map of CHS zones before and after the two changes.

In both downsizing cases, households from the excluded areas were reassigned to Hillmorton High School or Linwood College, Decile 4 and Decile 3 respectively. There are several strands of evidence that the desirability of the alternative school zones was far lower than CHS, such as their lower attendance rate, school leavers trends, and NCEA achievement rates. The ERO's Linwood College 2020 Review Report states, "one third of students are leaving without attaining NCEA Level 2", and ERO's overall evaluation judgement of the school was “[d]eveloping”. As reported in its 2017 Education Review Report, the NCEA Level 2 achievement rate at Hillmorton High School was at a similar level as at Linwood College. Similarly, both Linwood College and Hillmorton High School need to be reviewed by the Education Review Office every three years. Regarding attendance rates, Figure 3.5 indicates the comparison of unadjusted absence rates of Decile 1-10 public schools in New Zealand since 2011, and lower decile schools clearly have a higher rate of unadjusted absences. Table 3.2 similarly illustrates the comparison of student retention rates of Cashmere High School, Linwood College, and

[22] See https://www.parliament.nz/en/pb/sc/reports/document/SCR_114554/petition-of-jo-bethell-review-of-proposed-rezone-of-cashmere?fbclid=IwAR0mDMzx5gZcark0UCN4AjWhSYyMPQj-0SRqUx7t4q4b6LIf6P3mM-vnsTA.

Table 3.1 Key Dates of CHS Home Zone Downsizing between 2018 and 2021

Key Dates	First Downsizing Events
26/07/2004	Setup of the original CHS school zone
October 2017	KPMG report which suggested changes to the secondary schooling network in Christchurch
03/03/2018	Email from the MoE to CHS with proposed reduction for 2019
04/04/2018	CHS sent consultation emails to BoT members and principals of feeder schools; Email sent to media to publish the Public Notice
05/04/2018	Email sent to local communities
12/04/2018	Public Notice published in the Star
24/04/2018	Closing of consultation
27/04/2018	Approval checklist issued by MoE to CHS
12/05/2018	Approved by MoE
	Second Downsizing Events
September 2019	CHS board indicated the need to consider a further amendment to its zone
24/09/2019	MoE emailed CHS requiring it develop an amendment to its enrolment scheme
19/11/2019	Consultation started
10/02/2020	Consultation closed
25/02/2019	CHS sent proposed amendment to the MoE
17/04/2020	MoE approved
01/01/2021	New Home Zone comes into force

Table 3.2 Comparison of retention trend data (2018-2020)

Comparison Groups	Percentage staying until at least 17th birthday		
	2018	2019	2020
Cashmere High School	89.1	88.2	89.3
Linwood College	63.7	68.6	64.5
Hillmorton High School	68.1	73.1	73.6
Decile 04	79.3	77.8	78.7
Decile 03	78.1	75.4	76.5
New Zealand	83.8	82.8	83.5

(Source: Education Counts, MoE 2021)[23]

Table 3.3 Comparison of NCEA Level 1 Achievement Data (2018-2020)

Comparison Groups	Percentage with NCEA level 1 or above		
	2018	2019	2020
Cashmere High School	94.5	94.2	94.5
Linwood College	84.1	77.8	83.7
Hillmorton High School	80.6	77.1	82.2
Decile 04	88.6	87.4	86.5
Decile 03	86.3	83.6	83.8
New Zealand	89.6	88.3	88.4

(Source: Education Counts, MoE 2021)

[23] Source: Education Counts.

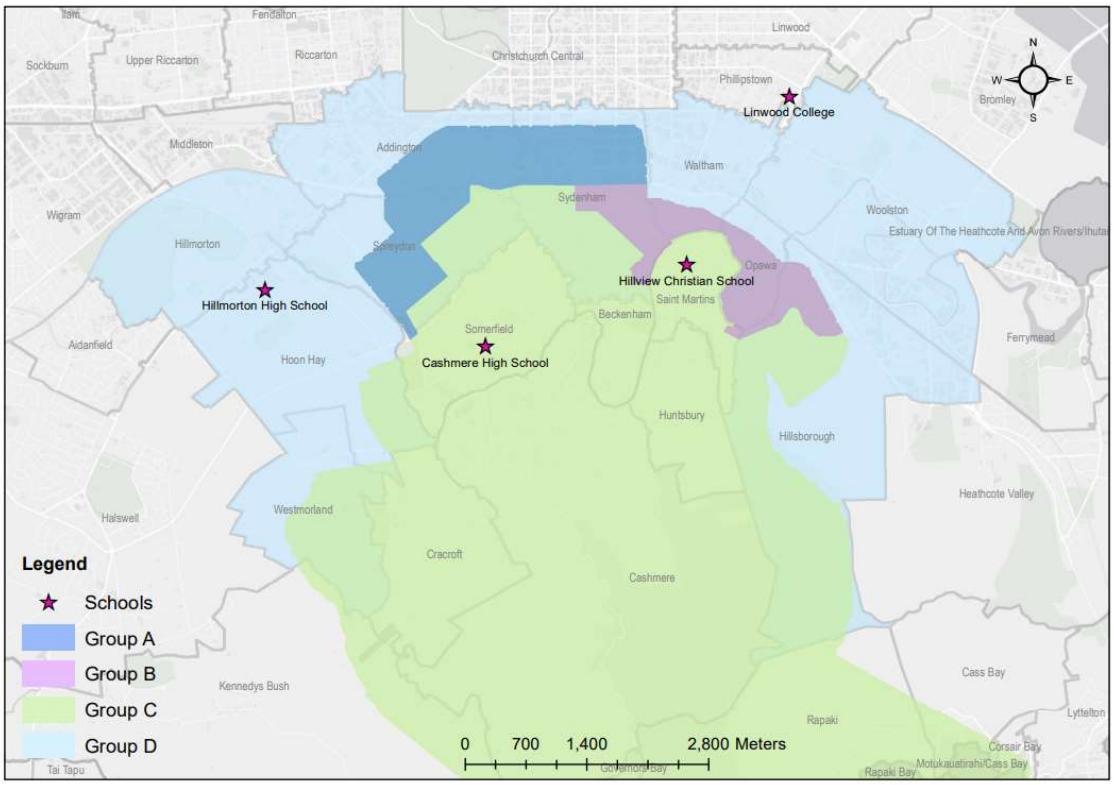


Figure 3.4 Map of CHS Zone Changes[24]

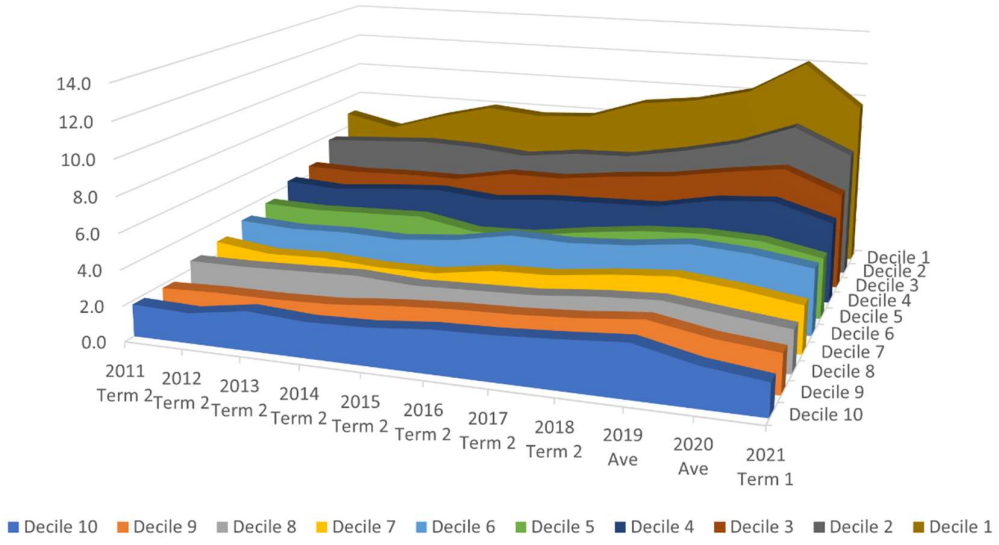


Figure 3.5 Unadjusted Attendance Rates of New Zealand Schools by Decile[25]

[24] This map is drawn according to the CHS school zone descriptions acquired from CHS website. See Appendix 1 for details.
[25] Source: Education Counts.

Hillmorton High School. Finally, Table 3.3 presents a comparison of students' NCEA Level 1 achievement rates at all three high schools along with the average of national, Decile level 3, and 4 schools. Each of these measures suggest that on academic grands CHS outperforms the other two schools (unjustified attendance rates, retention rates, and NCEA Level 1 achievement rates).

In sum, the above evidence might suggest that houses excluded from the first and second downsizing of the CHS enrolment zone were experiencing a substantial drop of state school access rights. For the first downsizing, affected households were reassigned to Hillmorton High School, and for the second, they were reassigned to Linwood College. Thus, if ever one might expect to find an effect of "good school" access right or "school quality effect" on the price of housing, it would be from this quasi-experimental change.

4. Data

4.1 School Zone Boundaries

The description of CHS's initial and subsequent enrolment zones was obtained from the school's official website and an archived record of previous versions from the internet archive of the Wayback Machine[26]. There are three versions of the enrolment zone used in this study. The earliest version was in force from April 26th 2004 until December 31st 2018. The second version was in force between January 1st 2019 and December 31st 2020, and the most recent version has been in effect since January 1st 2021 ((See Appendix 1 for details of school zone descriptions). The announcements of the latest two changes were made to the public by the CHS Board of Trustee on April 4th, 2018, and November 19th, 2019, respectively. In my research, I have chosen to use the announcement dates of school zone changes as cut-off dates to identify the three periods of pre-change, between the first and second changes, and post the second change. I assume the effect of school zone changes on

[26] https://web.archive.org/web/20180217072710/http://www.cashmere.school.nz/enrolment/CHS-zone.html.

housing prices should start immediately after the announced zone revisions are made public, instead of waiting until the effective date of such changes.

4.2 Housing Sales Data

Data for housing sales prices and characteristics were acquired from CoreLogic's PropertyGuru website (www.property-guru.co.nz). This website is the authorized source of housing transaction data in New Zealand and stores all types of property transaction records provided by government agencies in New Zealand. I constructed a dataset of houses in my region of interest by setting the search conditions to "residential" transactions "anytime" from 26 suburbs in and around the CHS Catchment. This dataset consists of 33,085 distinct houses. Based on the house valuation reference number and profile link of each house, I scraped the sales history records for all houses in the dataset and developed a house-sales event dataset with 111,822 observations (house sales records).

Housing characteristics captured in this dataset include *HouseId*, *SalesId*, LogHousePrices (transformed from *GrossSalesPrice*), *SalesDate, Address, Suburb*, *LandArea*, *FloorArea*, *BuildingAge*, *BedroomsNumber, GaragesNumber, OuterWallMaterial, RoofMaterial, Contour* (hilly or not), *Deck*, *SalesType, TenureType*, *SaleTenure*, and *bonafide*.[27, 28] The last three variables will be explained in further detail.

In my dataset, the value of *TenureType* consists of four types: 1) single or wholesale, 2) multi sale of two or more properties (units) sold in one transaction, 3) cross-reference sale where a property was sold in conjunction with a multi sale transaction, and 4) partial sale where only part of a property was

[27] SalesType denotes whether the sale is for the whole property or part of the property.
[28] Except variables explained in the text, the dataset also consists of property valuation record data, including gross price, net price, chattels, capital value, land value, improvement, valuation reference and valuation date. The sum of net sales price and chattels equals gross sales price. The dataset also includes estimated valuation data used to assess property taxes in New Zealand, including valuation reference, capital value, land value, improvement value, and most recent revaluation date, among which capital value is the sum of land value and improvement value. Here improvement indicates the property built on the land lot.

sold. In my study, I use only the first type of single or wholsale transactions where only one property was involved because I could not accurately allocate the transaction price per property in other categories.

The variable *SaleTenure* describes the ownership status of the property. "Freehold" means the property and land are both included in the title of the property; "Leasehold" means the ownership of land is not included in the property title; "Share in property" means the ownership only covers a particular percentage share in the property; and "Other" indicates the leaser's interest, air rights or mixed tenures. In other words, the prices reported for the last three types of sale tenures do not reflect the value of the property as a whole. For that reason, I only kept records for Freehold sales in data cleaning.

The variable *BonaFide* shows the price versus value relationship for a transaction. This is classified as an open market sale that can be roughly matched with the property's estimated capital value; or an open market sale which cannot be roughly matched with the property's capital value; or a non-market non-arm's length sale that cannot be matched with the property's capital value, such as a gift, public works or family sale. In my study, I excluded the latter two categories.

4.3 Data Cleaning and Imputation

4.3.1 Data Cleaning and Sample Reduction

Aside from dropping three sales records which appeared to be outside my 26 selected suburbs, my main data cleaning involved dropping duplicate sales records from the Property Guru data. This reduced my sample of house sales from 111,822 to 90,870. In order to make the dataset an annual panel, I next only kept the last sale of any house in a particular year that sold multiple times within the same year. This left 87,951 house sales observations. Next, I eliminated reported sales prices higher than 2 million or lower than $100,000 NZD, as well as non-freehold, non-whole sales, non-

bona-fide sales, and sales with 0sqm of floor area as these sales are extreme outliers in the sample. This resulted in a final sample size of 71,128 house sales. Finally, I considered another two factors in my sampling. As the school zone of CHS before the 2018 downsizing was set up in 2004, I decided to exclude any transaction before 2004 in my analysis. Similarly, as the housing market in Christchurch has gone through tremendous damage and uneven development due to the 2010 and 2011 earthquakes, I further excluded all sales before 01/01/2012, and arrived at a final dataset of 14,738 sales records from 11,093 houses between 2012 and 2021.

4.3.2 Data Imputation

Unfortunately, even among my final dataset of sales records, there are some missing values among the housing characteristics variables. These are particularly relevant for our pooled cross section regressions later. My imputation strategy is to replace all missing values for the categorical variables (BuildingAge, Deck, Contour, GarageNumber, OuterWallMaterial and RoofMaterial) with "Unreported", while replacing missing values for both FloorArea and LandArea with their means from the summary statistic results in my dataset.

5 Empirical Estimation Strategy

5.1 Canonical DID Models

Since its first use in Snow's (1855) study on the Mode of Communication of Cholera, the Difference-in-differences method has been extensively applied as a key tool of causality analysis in many branches of economics, such as Urban Economics (Baum-Snow and Ferreira 2015), Labor Economics (Angrist and Krueger, 1999; Blundell and Macurdy 1999), Education Economics (McEwan 2010; Schwerdt and Woessmann 2020), Finance Economics(Roberts and Whited, 2013), Health Economics (Umapathi, 2014), Environment Economics(Milliment, 2013), and so on.

To illustrate the canonical DID model, suppose there were only one treatment event in my study, namely the first school zone downsizing of CHS. I denote this "treatment" event as a binary variable $D_i \in \{0,1\}$. The outcome of interest in my study is LogHousePrices[29], denoted as Y_i. The main question I am asking is whether losing access to CHS affects the housing sales prices for those houses treated. Let's assume the decision of whether a house is affected by the school zone downsizing (treatment event) is randomly made, and no anticipation (or pre-treatment effect) exists. The treatment effect τ_i of losing access would be the difference

$$\tau_i = (Y_i|D_i = 1) - (Y_i|D_i = 0). \quad (1)$$

Under the potential outcomes framework of analysis, the ideal scenario in a treatment effect (intervention effect) study would be that the potential outcomes of Y_i both when D_i=1 and D_i=0 are observable. However, since it is impossible to observe Y_i in both statuses of treated ($D_i = 1$) and not treated ($D_i = 0$), the difference-in-differences method generally estimates the average treatment effect of the treatment group by comparing the difference in the group mean of outcomes over two periods of pre-treatment and post-treatment

$$\tau_{ATT} = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]. \quad (2)$$

Sometimes, this average treatment effect estimator is also denoted as

ATE= $(\overline{y}_{T,2} - \overline{y}_{T,1}) - (\overline{y}_{C,2} - \overline{y}_{C,1})$, where, ATE indicates average treatment effect, T and C indicate treatment group and control group respectively, and 1 or 2 indicates pre-treatment and post-treatment periods respectively.

The main idea of difference-in-differences is to detect the effect of treatment by comparing the difference between two groups, the treatment and control, before and after the treatment. Its basic formula without and then with controls are as follows:

Two-way Fixed Effect DID Model without controls

[29] LogHousePrices is transformed from gross house sales prices.

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 T_i + \beta_3 D_i * T_i + \varepsilon_{it}, \quad (3)$$

Two-way Fixed Effect DID Model with controls of covariates

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 T_i + \beta_3 D_i * T_i + \beta_4 X_i + \varepsilon_{it}, \quad (4)$$

Under the potential outcomes framework, Y_{it} indicates the outcome of unit i at time t; D_i indicates which group the unit is in, 1 for treatment group, 0 for control group. T_i indicates which period the unit belongs to, either 1 for post-treatment, or 0 for pre-treatment. X_i is a vector of control variables for unit i. $\beta_1$represents the fixed effect of being in the treatment group and is also called the group effect parameter. β_2 represents the effect of being in pre- or post-treatment period T and may also be called the period effect parameter. Of key interest, β_3 represents the singular coefficient on the interaction term $D_i T_i$, indicating the average treatment effect of units in the treatment group D_1. It is also called the parameter of average treatment effect on the treated group (ATET or ATT). Finally, β_4 denotes the effect of covariates of unit *i*.

The power of Difference-in-Differences (DID) strategy lies in its potential to eliminate confounding effects from both the time-invariant and time-varying unobserved characteristics of treatment and control groups. In a conventional setup, DID first eliminates the time-invariant unobserved group characteristics by differencing the mean outcome for the treatment and the control group over the two periods: pre-treatment and post-treatment, and then eliminates the time-varying confounders by differencing the two across-time differences of two groups. The above two steps could be swapped in order, and no matter which order is taken, the result would remain the same. See Appendix 2 for an illustration of this process.

5.1.2 Developing a General Multi-group Multi-period Multi-treatment DID Model

In this section, to better correspond with the two zone changes that have occurred to CHS, I present a General Multi-group Multi-period Multi-treatment Difference-in-Differences (DID) model, which

expands the canonical Single-Treatment, Two-Group, Two-period DID model. I begin without control variables or year dummies that further break up periods, then include them.

1) Expanded Model without Controls

$$Y_{it} = \alpha + \sum_1^j \beta_j\, Group_j + \sum_1^k \gamma_k\, Period_k + \sum_1^l \delta_l Treatment_l + \sum_1^j \sum_1^l \varphi_{jl}\, Group_j * Treatment_l + \sum_1^k \sum_1^l \omega_{kl}\, Period_k * Treatment_l + \sum_1^j \sum_1^k \rho_{jk}\, Group_j * Period_k + \sum_1^j \sum_1^k \sum_1^l \pi_{jkl}\, Group_j * Period_k * Treatment_l + \varepsilon_{it} \qquad \{i, t, j, k, l \in \mathrm{N}\}. \tag{5}$$

2) Expanded Model With Controls and Year Fixed Effects

$$Y_{it} = \alpha + \sum_1^j \beta_j\, Group_j + \sum_1^k \gamma_k\, Period_k + \sum_1^l \delta_l Treatment_l + \sum_1^i \theta_i X_i + \sum_i^t \eta_t Year_t + \sum_1^j \sum_1^l \varphi_{jl}\, Group_j * Treatment_l + \sum_1^k \sum_1^l \omega_{kl}\, Period_k * Treatment_l + \sum_1^j \sum_1^k \rho_{jk}\, Group_j * Period_k + \sum_1^j \sum_1^k \sum_1^l \pi_{jkl}\, Group_j * Period_k * Treatment_l + \varepsilon_{it} \qquad \{i, t, j, k, l \in \mathrm{N}\}. \tag{6}$$

In the above model, all coefficients but α are vectors. Here β_j estimates the time-invariant fixed effect of Group *j*; $\gamma_k$estimates the fixed effect of being in Period *k*, i.e. the time series change of each period. Next, δ_l estimates the time-invariant fixed effect of treatment event *l*; θ_i estimates the time invariant effects of control variables X_i; η_t estimates the year fixed effect; φ_{jl} estimates the time-invariant effect of treatment *l* for Group *j*; ω_{kl} estimates the group-invariant effect of treatment *l* in Period *k*; ρ_{jk} estimates the treatment-invariant fixed effect of Group *j* in Period *k*, and π_{jkl} estimates the effect of treatment *l* for Group *j* in Period *k*. Note that commonly, Treatment event *l* is dependent on Group *j* and Period *k*, and can usually be defined as $Treatment_l = Group_j * Period_k$. In that case, Model (5) would have fewer terms left due to collinearity, which I will explain in detail in Section 5.4.

5.2 Assumptions made under DID model

In the original two-way fixed effect DID model (with and without controls), a homogenous treatment effect is assumed implicitly for all units within the treatment group as proposed by Rubin (1980). Put another way, instead of identifying heterogeneous treatment effects for different units within the treatment group, both DID models estimate an average size of treatment effect across the treated group.

The main assumptions of a basic DID model include that the treatment group and the control group are systematically identical; all units in the treatment group get treated simultaneously, with no anticipation of treatment in the treatment group prior to the treatment. In addition, the status of groups and periods does not change over time. Among the above assumptions, the first one about systematic similarity between the treatment and the control groups is often called the "common trend" or "parallel trend" assumption. This is commonly viewed as a key assumption when setting up an effective DID empirical analysis. These assumptions carry over to the Two-treatment Four-group Three-period Expanded DID Model I will estimate.

In my study, I assume that the treatment event, school zone downsizing of CHS, is independent from the observed housing sales prices and observed housing characteristics data. As explained in Section 3.3.3, the initiative of school zone change was triggered by a predicted excess demand by in-zone students, as confirmed by a joint survey effort from the MoE and KPMG. A final implicit assumption of DID models is that all sales affected by the school zone downsizings in each period are treated simultaneously by ignoring the different gaps in time between the treatment occurrence and the exact timing of each house's market entrance.

5.3 Identification Strategy

In this quasi experiment, I want to reveal the impact of losing access to the CHS Zone on the sales prices of houses located in the affected areas. While there is a gap between the announcement of the CHS downsizing and the actual implementation, I assume that the effect should start immediately after the news is released to the public. At the same time, I assume there is no pre-announcement effect as the school zone downsizing was not much anticipated by the households affected. Based on the above assumption, I set the two downsizing treatment dates as April 4th, 2018 and November 19th, 2019 respectively. There are the dates the Board of Trustees of CHS released their public notice

of the upcoming school zone changes to its contributing primary schools, community boards and to the media. Hence, I have cut the timeline into three periods: t1 for the period before April 4th, 2018, t2 for the period between April 4th, 2018 and November 19th, 2019, and t3 for the period after November 19th, 2019.

Based on the school zone descriptions, I have identified whether each individual house is located within the zone or not over time, and divided all houses into four groups: Group A are houses whose loss of access was announced on April 4th 2018; Group B are houses whose loss was announced on November 19th, 2019; Group C are houses that always stayed in the CHS zone; and Group D are adjacent area houses that were never in the zone.

5.3 Identification

5.3.1 Three Periods Dummy Variables

I defined three period dummy variables t1, t2, and t3: t1 indicates the period before the first school change announcement made by the CHS Board of Trustees on April 4th, 2018; t2 indicates the period after the first zone change announcement but before the second, i.e. after April 14th, 2018, but on or before November 19th, 2019; and t3 indicates the period after the second school zone change announcement, i.e. from November 19th, 2019 (not inclusive) up to the close of my data collection on September 20th, 2021.

5.3.2 Year Dummy Variables

Notwithstanding my creation of three period dummies, in order to better control for time fixed effects within the longer first period, I also created year dummies from Year2012 to Year2017, and 2018JanApr(which equals 1 for sales taking place between January 1st, 2018 and April 4th, 2018), and combine these seven dummy variables together with t2 and t3 as a comprehensive set of time dummies.

5.3.4 Group Dummy Variables of Cashmere Enrolment Zone Status

To create my four Group dummy variables, I first created three dummy variables to describe the school zone status of each house: In Zone in Period t1, In Zone in Period t2, and In Zone in Period t3. The value 1 indicates that at the time of a sale, the house is located within the CHS Zone, while the value 0 means it is not. I manually constructed these three dummy variables based on the house location, date of sale, and the spatial descriptions of CHS's catchment zone as it changed twice in the period of my study.

I next created four school zone groups from the above three school zone status dummy variables, denoted as Groups A, B, C, and D. To create these variables, I used the difference between In Zone Period t1 and In Zone Period t2 to define whether a given house had lost access during the first school zone change. When the difference equals 1, I assigned the house to Group "A", meaning the house lost access in Period t2. Similarly, I used the difference between In Zone Period t2 and In Zone Period t3 to define whether the house lost its access in the second zone change. I assigned such houses to Group "B". I define a house to be in Group C if all its values of In Zone for all periods equalled 1, indicating that it has remained within the CHS Zone throughout the sample. Group C is the reference group in the upcoming regression analysis. Conversely, I assigned a house to Group "D" when all values of In Zone Period t1, In Zone Period t2 and In Zone Period t3 were 0, indicating that the house was never included in the Cashmere School Zone.

5.4 Two-treatment Four-group Three-period Expanded DID Model

I now fit my identification strategy into Model (5). As explained in Section 5.3, I have divided the whole sample into four groups, so I have j=4 groups, k=3 periods, and l=2 treatments. Also, as the treatment events can be expressed in terms of groups and periods, I have *Treatment 1=GroupA*t2*, and *Treatment 2=GroupB*t3*. Note that all interaction terms with *t1* and *Group C* are not included in my

model as *Group C* and Period *t1* are our omitted benchmarks. Table 5.1 thus provides a complete set of interaction terms retained in my base model.

Therefore, my final identified models without controls and year fixed effects is:

$$Y_{it} = \alpha_i + \beta_1 GroupA + \beta_2 GroupB + \beta_3 GroupD + \gamma_1 \mathrm{t2} + \gamma_2 \mathrm{t3} + \rho_1 GroupA * t2 + \rho_2 GroupA * t3 + \rho_3 GroupB * t2 + \rho_4 GroupB * t3 + \rho_5 GroupD * t2 + \rho_6 GroupD * t3 + \varepsilon_{it}. \quad (7)$$

Including either housing characteristics or house fixed effects and year fixed effects, this expands to either:

$$Y_{it} = \alpha + \sum_1^i \theta_i X_i + \sum_1^t \eta_t Year_t + \rho_1 GroupA * t2 + \rho_2 GroupA * t3 + \rho_3 GroupB * t2 + \rho_4 GroupB * t3 + \rho_5 GroupD * t2 + \rho_6 GroupD * t3 + \varepsilon_{it}, \quad (8)$$

or

$$Y_{it} = \alpha + \sum_1^i \theta_i H_i + \sum_1^t \eta_t Year_t + \rho_1 GroupA * t2 + \rho_2 GroupA * t3 + \rho_3 GroupB * t2 + \rho_4 GroupB * t3 + \rho_5 GroupD * t2 + \rho_6 GroupD * t3 + \varepsilon_{it}. \quad (9)$$

In all three models, Y_{it} represents the LogHousePrice. The results of Model (7) are to be reported in Column (1) of Table 6.4. Model (8), which includes controls for year fixed effects for each sale, and a vector of house characteristics X_i, will be reported in Column (2) of Table 6.4 Column 2. Model (9), which replaces the vector of house characteristics with a house fixed effect H_i, will be reported in Column (3) of Table 6.4. Note that due to collinearity between house and group dummies, the fixed effect of groups cannot be estimated in Models (8) and (9). This is because the groups are defined by house locations, and when all house dummies are included in the model, there will be perfect collinearity between each group dummy and the sum of the corresponding house dummies in that group. For the same reason, Period fixed effects will not be estimated with year dummies included in Models (8) and (9).

Table 5.1 Interaction terms between groups and periods

Periods\Groups	**A**	**B**	**C**	**D**
t1	*A*t1*	*B*t1*	*C*t1*	*D*t1*
t2	A*t2	B*t2	*C*t2*	D*t2
t3	A*t3	B*t3	*C*t3*	D*t3

(Note: The grey-colored, italic interactions terms with Group C and Period t1 are not included as both are baseline in my study)

Returning to the DID framework, my parameters of interest in Model (7) are defined as the average treatment effect on each treatment group in each treatment period of t1 and t2, expressed as:

$ATT_{Treatment1}$=$\left(\overline{y_{A,t2}} - \overline{y_{A,t1}}\right) - (\overline{y_{C,t2}} - \overline{y_{C,t1}})$= ρ_1 *in Model (7),* and

$ATT_{Treatment2}$=$\left(\overline{y_{B,t3}} - \overline{y_{B,t1}}\right) - (\overline{y_{C,t3}} - \overline{y_{C,t1}})$= ρ_4 *in Model (7).*

See Appendix 3 for a complete interpretation of all estimators included in Models (8) and (9).

6. Main Results

6.1 Descriptive Statistics

In my dataset, I have 14,738 house sales observations *(SalesId)* from 11,093 houses *(HouseId)*. The observations from Group C accounts for 44% of the total houses, Group A for 10.8%, Group B for 8.4% and Group D for 36.7%. In terms of period, 68.2% of housing sales occurred in the longer pre-treatment period *t1*, 16.9% after the first downsizing in *t2*, and 14.9% in *t3*. Table 6.1 provides a summary of group and period variables. The average LogHousePrices of all sales is 13.122 with a standard deviation of 0.398. Table 6.2(A) shows a summary of sales counts and mean of LogHousePrices for each group over all three periods. A more intuitive comparison of sales counts across the four groups over time is given in Figure 6.1 It's apparent that there are far fewer house sales in Periods *t2* and *t3* for both Group A and Group B compared with sales in Group C or D. The distribution of LogHousePrices of each group in all three periods is illustrated in Figure 6.2. As shown in the boxplots, the distribution pattern and upward trend in LogHousePrices among all four groups are very similar.

Table 6.1 Summary of Group, Period and Year Dummy Variables

Variable Name	N	Percentage	Meaning
HouseID	11093		ID of each unique house
LogHousePrices	14738		Log of each house sales price
Groups	14738		
-A	1598	10.8%	Houses leaving CHS zone in Period t1
-B	1236	8.4%	Houses leaving CHS zone in Period t2
-C	6492	44.0%	Houses within CHS Zone all three periods
-D	5412	36.7%	Houses never in CHS Zone
Periods	14738		
-t1	10057	68.2%	From Jan 1, 2012 to Apr 4, 2018
-t2	2484	16.9%	From Apr 5, 2018 to Nov 19, 2019
-t3	2197	14.9%	From Nov 20, 2019 to Sep. 20, 2021
Year	14738		
--Year2012	1488	10.1%	Year of 2012
--Year2013	1611	10.9%	Year of 2013
--Year2014	1596	10.8%	Year of 2014
--Year2015	1848	12.5%	Year of 2015
--Year2016	1674	11.4%	Year of 2016
--Year2017	1483	10.1%	Year of 2017
--Year2018JanApr	357	2.4%	From Jan 1, 2018 to Apr 4, 2018
--Year18to19Nov	2484	16.9%	From Apr 5, 2018 to Nov 19, 2019 (i.e. t2)
--Year19to21	2197	14.9%	From Nov 20, 2019 to Sep 20, 2021 (i.e. t3)

Table 6.2(A) Average of LogHousePrices and Sales Counts by Groups and Periods

Groups	Periods	Average of *LogHousePrices*	Count of *Sales*
C	t1	12.99	4399
C	t2	13.13	1110
C	t3	13.26	983
A	t1	12.76	1091
A	t2	12.85	257
A	t3	12.95	250
B	t1	12.79	857
B	t2	12.90	194
B	t3	12.99	185
D	t1	12.74	3710
D	t2	12.85	923
D	t3	12.95	779

For our main fixed effects analysis of repeat sales, of more particular interest are the number of houses that have been sold both before and after each of the two downsizings. For a better understanding of these, Table 6.2(B) shows the number of houses sold from each group in both pre and post periods of each downsizing. For our first downsizing, there were 92 houses in Group A that sold at least once in both periods of t1 and t2. This compares to 418 houses from Group C that sold in both those periods.

Table 6.2(B) Count of Houses Sold in Both Pre & Post Periods of Two Downsizings

Identification		Count of Houses Sold in both pre & post periods			
School Zone Changes	Pre Vs Post	A	B	C	D
1st Downsizing	T1 Vs T2	92	78	418	367
	T1 Vs T2+T3	203	155	782	706
2nd Downsizing	T1 Vs T3	117	82	388	361
	T1+T2 Vs T3	126	89	437	397
	T2 Vs T3	15	12	73	58

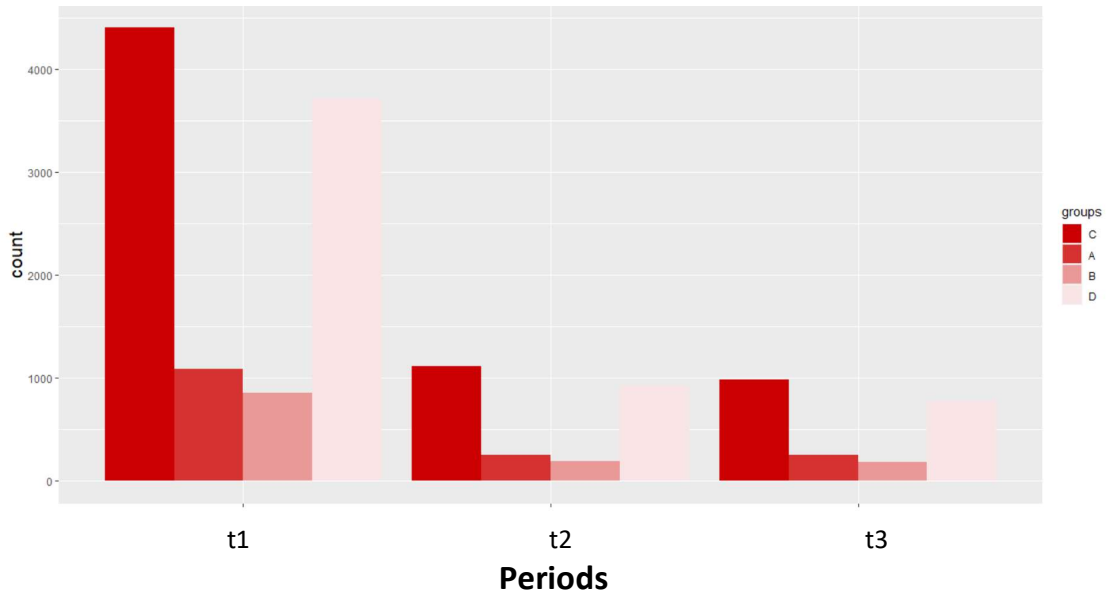
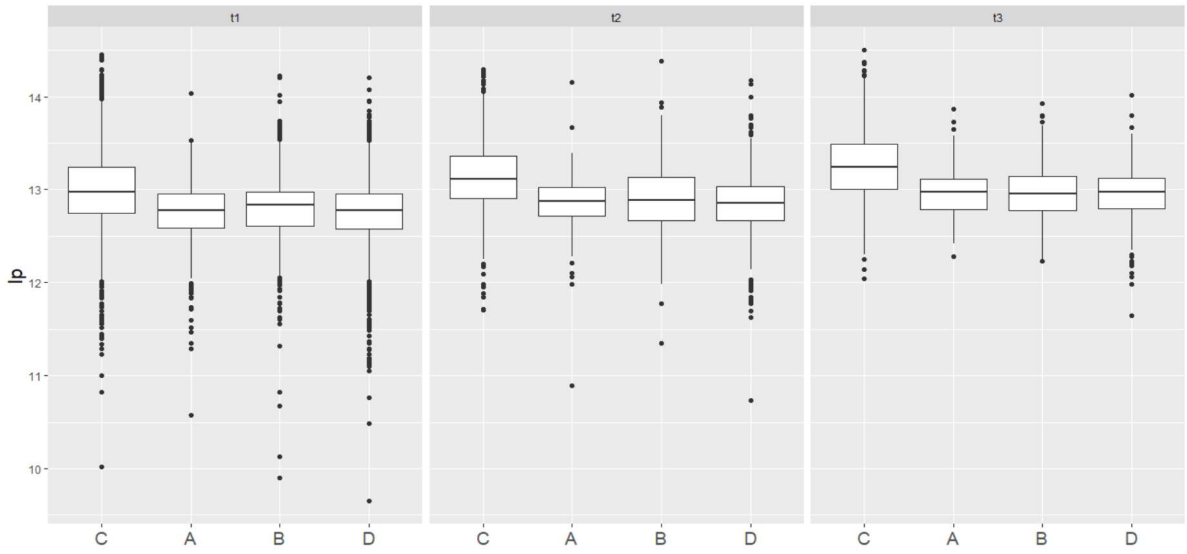


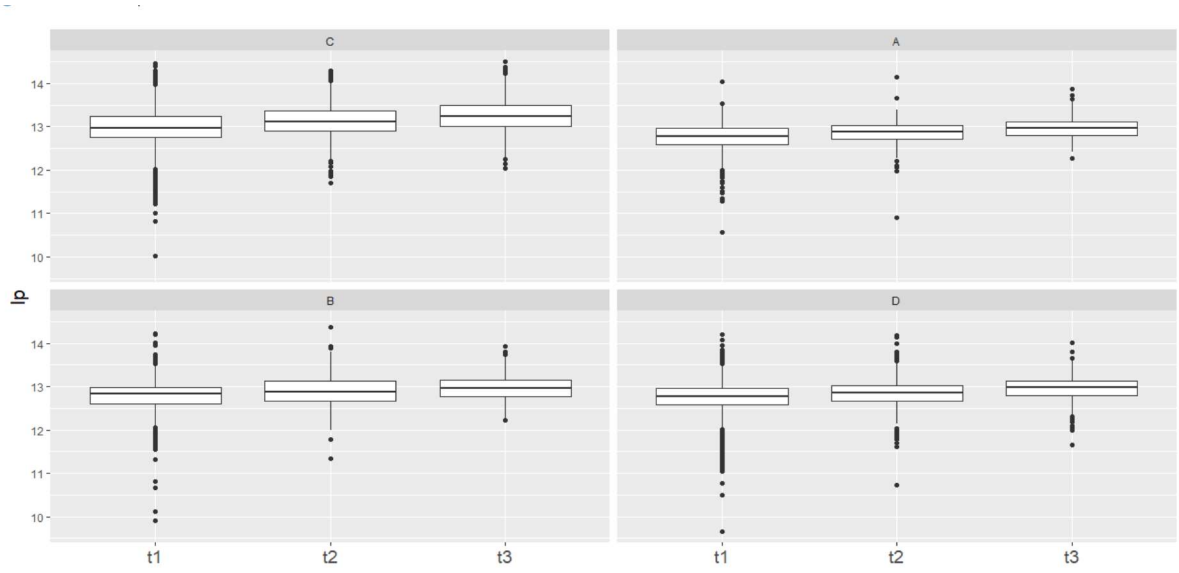
Figure 6.1 Comparison of Sales Counts by Groups across Periods

If we expand our post-treatment period to include both t2 and t3 (as in a later robustness check), this count increases to 203 houses selling before and after the first downsizing in Group A, and 782 houses in Group C. For our second downsizing, there were 82 houses from Group B that sold at least once in both periods of t1 and t3, and 388 houses in Group C. Again, this number climbs to 89 houses from Group B and 437 houses from Group C if we include both t1 and t2 into the pre-treatment period.

Furthermore, the distribution of LogHousePrices by groups and periods illustrated in Figure 6.2 indicates there are more outliers of house sales in Group B and D compared to Group A and C. An intuitive summary of what treatment effects may be uncovered is provided by the group mean of



(a) LogHousePrices Distribution by groups for each period



(b) LogHousePrices Distribution by periods for each group

Figure 6.2 Boxplot of Distribution of LogHousePrices by Groups and Periods

LogHousePrices over time, which is presented in Figure 6.3. Quite apart from treatments, it's apparent that the average sales price from houses in the "always enrolled" Group C is consistently higher than in the other three groups, while that of Group D is consistently lowest during the observation period. It's apparent that there is an approximate parallel trend in the increases in the mean of

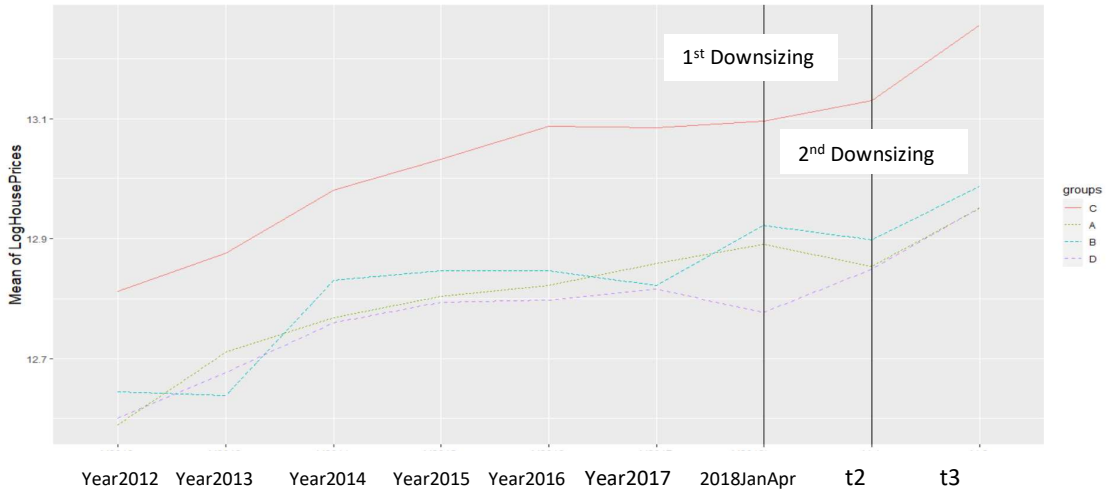


Figure 6.3 Line Graph of Mean of LogHousePrices of A, B, C, and D groups from 2012 to 2021

LogHousePrices between groups in the majority of years within the first treatment period t2 and in the second treatment period t3. The two vertical lines indicate two downsizings. Moving to the treatment effects, however, the magnitude of price increases in Groups C and D seem to remain similar over periods, while the increases of Groups A and B have been far less. This is the first suggesting evidence that the treatment of losing access to CHS negatively affected housing prices.

Finally, regarding descriptive statistics, I have collected house feature variables for each house sale. These are summarized in Table 6.3. Among all the housing features, the problem of missing values was severe for *GarageNumber, Contour, Deck* and *LandArea*. Nonetheless, from available data, my summary shows that an average residence from my dataset has a floor area of 136 square meters, 2.9 bedrooms and 1.7 parking spaces. My imputation strategy has been explained in Section 4.3.2.

6.2 Main Results

In the baseline Two-treatment Four-group Three-period DID Model (Model (7)), I regress LogHousePrices on the six interaction terms between periods t2 and t3, and location groups A, B and D. As illustrated in Section 5.4, I use Group C, the area that always stays within the CHS zone, and

Table 6.3 Summary of Descriptive Statistics on House Features

Panel (A) Continuous Variables							
Variable Name	N	Mean	Standard Deviation	Min	25%	75%	Max
LogHousePrices	14738	12.911	0.398	9.649	12.692	13.122	14.502
FloorArea	14738	136	60.578	28	100	157	740
LandArea	10269	0.069	0.059	0.006	0.05	0.081	3.196
BedroomsNumber	14708	2.934	0.791	1	2	3	16
GarageNumber	3409	1.664	0.662	0	1	2	5

Panel (B) Discrete Variables			
Variable Name	Count	Percentage	Meaning
Contour	14738		Contour levels of the house site
-Level	7999	54.27%	
-Easy to moderate fall	483	3.28%	
-Easy to moderate rise	683	4.63%	
-Steep fall	265	1.80%	
-Steep rise	287	1.95%	
-NA	5021	34.07%	Missing
Deck	14738		
-Yes	4483	30.42%	With Deck
-No	4759	32.29%	Without Deck
-NA	5486	37.22%	Missing
OuterWallMaterial			
-Brick	3813	25.87%	
-Weatherboard	3735	25.34%	
-Concrete	2875	19.51%	
-Roughcast, etc	2086	14.15%	
-Mixed Material	1472	9.99%	
-Fibre Cement	425	2.88%	
Others	318	2.15%	
-NA	14	0.09%	Missing
RoofMaterial			
-Steel / G-Iron	9510	64.53%	
-Tile Profile	4815	32.67%	
-Others	399	2.71%	
-NA	14	0.09%	Missing
BuildingAge			
1880 - 1889	9	0.06%	
1890 - 1899	4	0.03%	
1900 - 1909	227	1.54%	
1910 - 1919	778	5.28%	
1920 - 1929	1227	8.33%	
1930 - 1939	762	5.17%	
1940 - 1949	725	4.92%	
1950 - 1959	1130	7.67%	
1960 - 1969	1317	8.936%	
1970 - 1979	2005	13.60%	
1980 - 1989	884	6.00%	
1990 - 1999	1842	12.50%	
2000 - 2009	1683	11.42%	
2010 - 2019	1765	11.98%	
Mixed/Remod	93	0.63%	
NA	287	1.95%	Missing
Suburb	14738		
-Woolston	2214	15.02%	
-Spreydon	1793	12.17%	
-Hoon Hay	1572	10.67%	
-Sydenham	1515	10.28%	
-Somerfield	1403	9.52%	
-Cashmere	1234	8.37%	

-Addington	1063	7.21%	
-Waltham	817	5.54%	
-SaintMartins	638	4.33%	
-Westmorland	467	3.17%	
-Hillsborough	448	3.04%	
-Huntsbury	424	2.88%	
-Beckenham	339	2.3%	
-Opawa	270	1.83%	
-Others	538	3.64%	Including GovernorsBay, Cracroft, DiamondHarbour, CassBay, Purau, CharterisBay, CoarsairBay, Allandale, Rapaki, PortLevy, Teddington, TeRapakiOTeRakiwhatapuka

Period t1, the period before the first school zone change, as my reference in Model (7).[30] The results of Model (7) are shown in Column (1) of Table 6.4.

For Model (8), I include the housing characteristics in place of group fixed effects and add controls for year fixed effects from Year 2012 to Year 2017 from Period t1 in addition to Periods t2 and t3, with Year2018JanApr and Group C as references. This is shown in Column (2) of Table 6.4. I next replace house characteristics with house fixed effects with the same omitted time and group. This is shown in Column (3) of Table 6.4. The coefficients on the interaction terms *GroupA*t2* and *GroupB*t3, i.e.* ρ_1 and ρ_4, are my main parameters of interest, as these are the average treatment effects of the first and the second school zone reductions.

It seems intuitive to expect that the loss of guaranteed access to a desirable secondary school would negatively affect the sales prices of houses located in the affected areas, i.e. Groups A and B relative to Group C. I therefore expect houses in Group A will have lower sale prices compared to Group C in Period t2 when they are affected by the first school zone change. Similarly, I expect houses in Group B to have lower sales prices compared to Group C in Period t3.

[30] The estimations are conducted in R with the lm() function from the *base* package and the plm() function from the *plm* package.

Table 6.4 Empirical Results

	Column (1)	Column (2)	Column (3)	Column (4)	Column (5)
Dependent Variable:	*LogHousePrices*				
GroupA*t2	-0.047830*	-0.07561****	-0.0915505**	-0.14608**	-0.0316010
	(0.028298)	(0.02125)	(0.0415798)	(-.057096)	(0.1304383)
GroupB*t2	-0.031789	-0.08085****	0.0026474	-0.068390	-0.0490009
	0.031702	(0.02388)	(0.0447062)	(0.063771)	(0.1196120)
GroupD*t2	-0.038607**	-0.05338****	-0.0030742	0.008030	0.0685539
	(0.018296)	(0.01373)	(0.0257712)	(0.034135)	(0.0668812)
GroupA*t3	-0.077086***	-0.06293***	-0.1184989***	-0.13317**	-0.0607046
	(0.028830)	(0.02167)	(0.0382585)	(0.057232)	(0.1276822)
GroupB*t3	-0.068435**	-0.08083****	-0.0206929	-0.068272	-0.0704909
	(0.032474)	(0.02441)	(0.0439728)	(0.064052)	(0.1196519)
GroupD*t3	-0.060688***	-0.04890****	-0.0640739**	0.012535	0.0046656
	(0.019428)	(0.01471)	(0.0262021)	(0.034553)	(0.0667134)
N:	14738	14708	14738	14708	14738
R^2	0.1484	0.5247	0.27736	0.52662	0.28168
Adjusted R^2	0.1477	0.5215	-1.933	0.52288	-1.93
F-Statistics	233.2 on 11 and 14726 DF, p-value: < 2.2e-16	166.3 on 97 and 14610 DF, p-value: < 2.2e-16	99.5447 on 14 and 3631 DF, p-value: < 2.22e-16	141.155 on 115 and 14592 DF, p-value: < 2.22e-16	44.2743 on 32 and 3613 DF, p-value: < 2.22e-16
Fixed effects:					
Groups	Y	N	N	N	N
Periods	Y	N	N	N	N
Houseids	N	N	Y	N	Y
Years	N	Y	Y	Y	Y
HouseFeatures	N	Y	N	Y	N
Suburbs	N	Y	N	Y	N

Notes: 1. Standard errors are reported in parentheses.
*2. Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 ' ' 1*

Reassuringly, my baseline estimation results in Column (1) of Table 6.4 are indeed mostly significantly negative. Compared to the control group of houses in Group C in Period t1, housing prices in the treatment groups of Group A from Period t2 and Group B from Period t3 have experienced decreases of 4.67% and 6.61% respectively.[31] Interestingly, houses in Group A seem to suffer a delayed greater cumulative fall in sales price (7.42%) in the subsequent Period t3 on average.

When I include housing characteristics in Column (2), the effect of losing access to CHS zone becomes 7.28% for houses in Group A in Period t1, and that of Group B becomes 7.76%, both significant at the 0.001 level. Compared to the baseline of house prices from Group C in Year2018JanApr, the cumulative effect shrinks to a 6.1% drop in price for houses in Group A in Period t3, also with a lesser significance level of 0.01. These results indicate all three groups of A, B and D suffered a drop in housing prices in both school downsizings. Note that, while controlling for housing characteristics rather than housing fixed effects makes better use of data from the roughly 8000 houses that sold only once during my sample, this approach does not control as well for the effects of stable but unmeasured characteristics of houses that could affect their sales prices.

When I instead control for both house and year fixed effects in Column 3, the magnitude of the negative effect of losing CHS zone access on the sales prices for houses in Group A rises to 8.75% in Period t2 and a cumulative 11.17% in Period t3, compared to the same baseline as in Column 2. Both estimators are statistically significant at 0.05 and 0.01 level respectively. Overall, these Group A results are consistent with people's expectation that the school zone downsizing would have a negative impact on housing prices for those houses losing access to the zone. However, the fact that the effect grows in the subsequent period might suggest this "new information" about downsizing

[31] As the dependent variable is log-transformed, I exponentiate the coefficients, subtract one from the result, and multiply by 100 to get the effect size on housing price.

does not fully penetrate throughout the market in the early years of the change, though such an increase of treatment effect does not appear in the results from Column 2.

While the results using house fixed effects in Column (3) find stronger effects for the first downsizing, they find weaker effects for the second. In particular, the treatment effect of losing CHS access drops to 2.05% for houses in Group B. More importantly, this effect estimate is not significant. As we shall see, the insignificance and the near zero effect size of the second school downsizings on Group B might be a result of a potential violation of the parallel trends assumption between Group B and C before the second downsizing. Alternatively, there may be potential unobserved confounding factors affecting housing sales in Group B. I also explore the possibility that the failure of identifying any effect from the second downsizing might be the result of too few observations as the duration of t3 is short, and the geographical area and number of houses in Group B are smaller. From Table 6.2 (A), only 185 of the 1236 sales that took place in Group B occurred in Period t3. Similarly, Table 6.2(B) shows that while Group A had 117 houses sold both in Period t1 and t3, only 82 houses did so in Group B. However, given Group A shares the same issue of small repeat sale sample size, and its effect from the second downsizing is larger and significant, the small sample size for repeat sales in Group B might not be the reason.

6.3 Test of Parallel Trends

According to the experiment setup, since Group D is a never treated group, a parallel trend assumption holds if the housing prices differences between Group D and Group C remain stable over time. In other words, I can test for parallels trend by checking if the coefficients on the interaction terms GroupD*t2 and GroupD*t3 are not significantly different from zero. Taking my housing fixed effects results in Column (3) as most credible, the estimate of GroupD*t2 is -0.003 with no significance, but that of GroupD*t3 -0.064 is significant at the 0.05 level. Based on this estimation result, it seems the parallel trend assumption holds between Group C and Group D in Period t2 but not in Period t3. Similarly, as

Group B is not affected by the first downsizing, if a parallel trend assumption holds between Group B and Group C in Period t2, then the coefficient on GroupB*t2 should not be significantly different from zero. As shown in Column (3), the estimate I get is 0.0026, which is not significant. For a concrete view, I also provide a whisker graph showing the estimated effect sizes and their 95% confidence intervals in Figure 6.4.

An alternative approach to testing the parallel trends assumption is to focus on all four groups in t1, before the occurrence of either treatment event. To do this, I have added controls for interaction terms of Groups A, B and D with each of Year2012 to Year2017 in Model (8) and Model (9). Here I again use the housing prices of Group C in the period of Year2018JanApr as the baseline. The regression results are reported in Column (4) and (5) of Table 6.4, respectively.

The first thing to note in Column (4) is that the estimates of the two treatment effects, GroupA*t2, and GroupB*t3 still remain negative, but the results are less significant. However, the magnitude of the Group A treatment effect is still significant at the 0.05 level and rises to -13.59%. Moving to the parallel trends test, both GroupD*t2 and GroupD*t3 are not significantly different from 0 as expected. Similarly, while the coefficient on GroupB*t2 has a negative sign, it too is not far off 0. Next, all the coefficients of the newly included interaction terms between years within Period t1 and groups are illustrated in Figure 6.5. From the graph it's clear that in certain years within Period t1, Group D did not experience the same trends as Group C. In contrast, it seems the average gap between Group A and our base Group C in Year2018JanApr is smaller, staying within the range of -4% to -10% across the years from 2012 to 2017, and that between Group B and C is sometimes larger, staying within the range of 5% and -20%. However unlike for Group D, these coefficients for Group A and B are never significantly different from zero. Thus for the most part, parallel trends within Period t1 cannot be rejected, though the interaction term means appear far from zero.

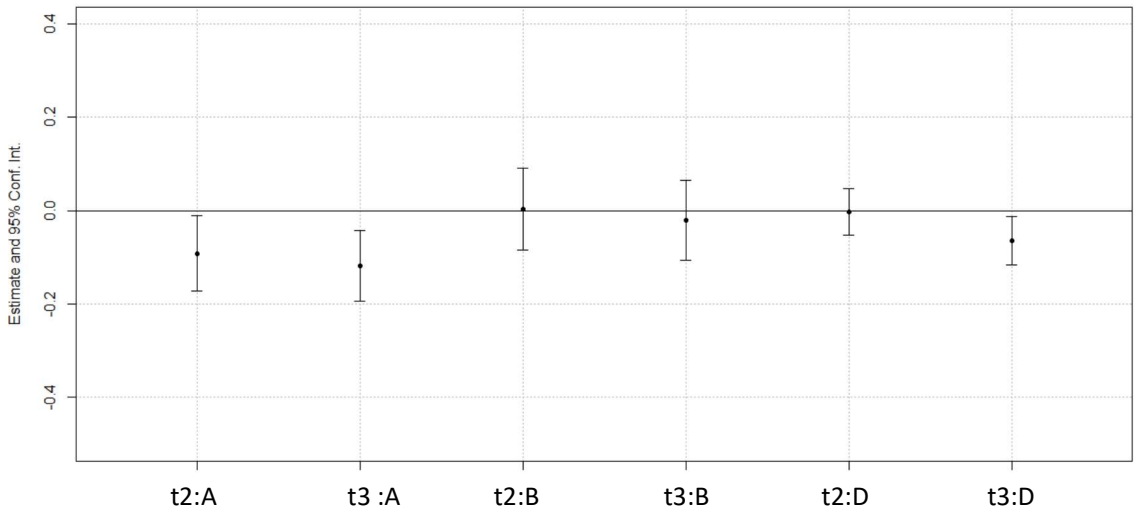
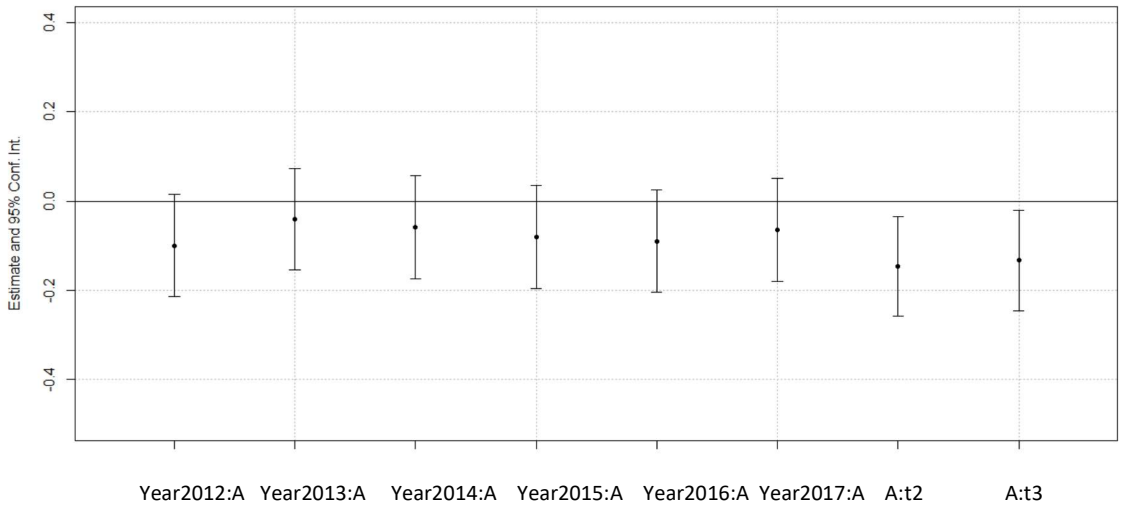


Figure 6.4 Graph of Estimates and 95% Confidence for Parallel Trend Test (Table 6.4 Column 3)



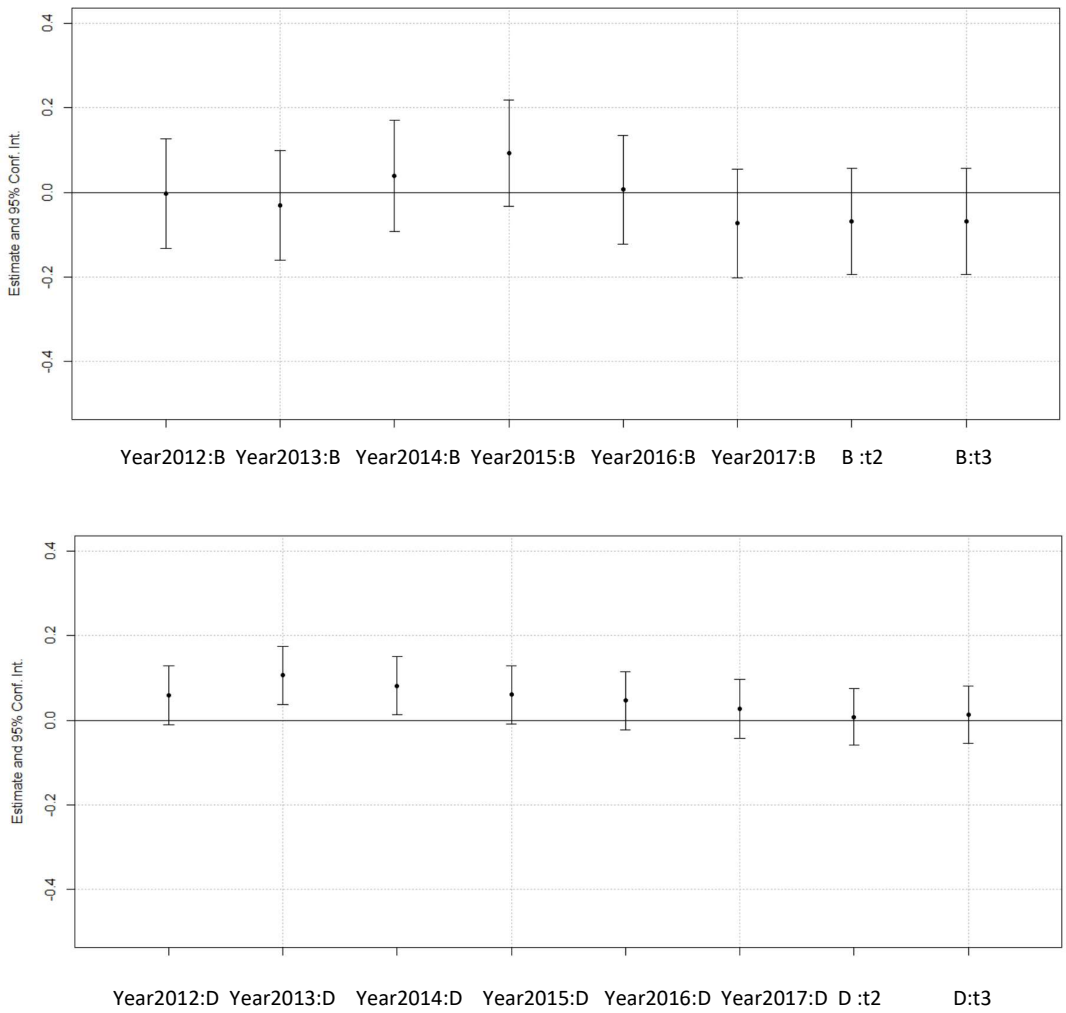


Figure 6.5 Coefficient Graphs for Parallel Trends Test based on estimation of Table 6.4 Column (4)

Moving to housing fixed effects in Column (5), it is interesting to note that both treatment effects lose statistical significance. Regarding parallel trends, all the coefficients of the newly included interaction terms are not significantly different from zero, which indicates that that the assumption of parallel trends roughly hold for all three groups of A, B and D compared with Group C throughout the pretreatment Period t1. Though again particularly for Group D, the interaction term means can appear far from zero. Figure 6.6 shows the coefficient graphs for these results. It seems the average gap between Group A and Group C throughout the pre-treatment period is in the range of 2% to 10%, and that between Group B and Group C is between 2.5% and 11% approximately.

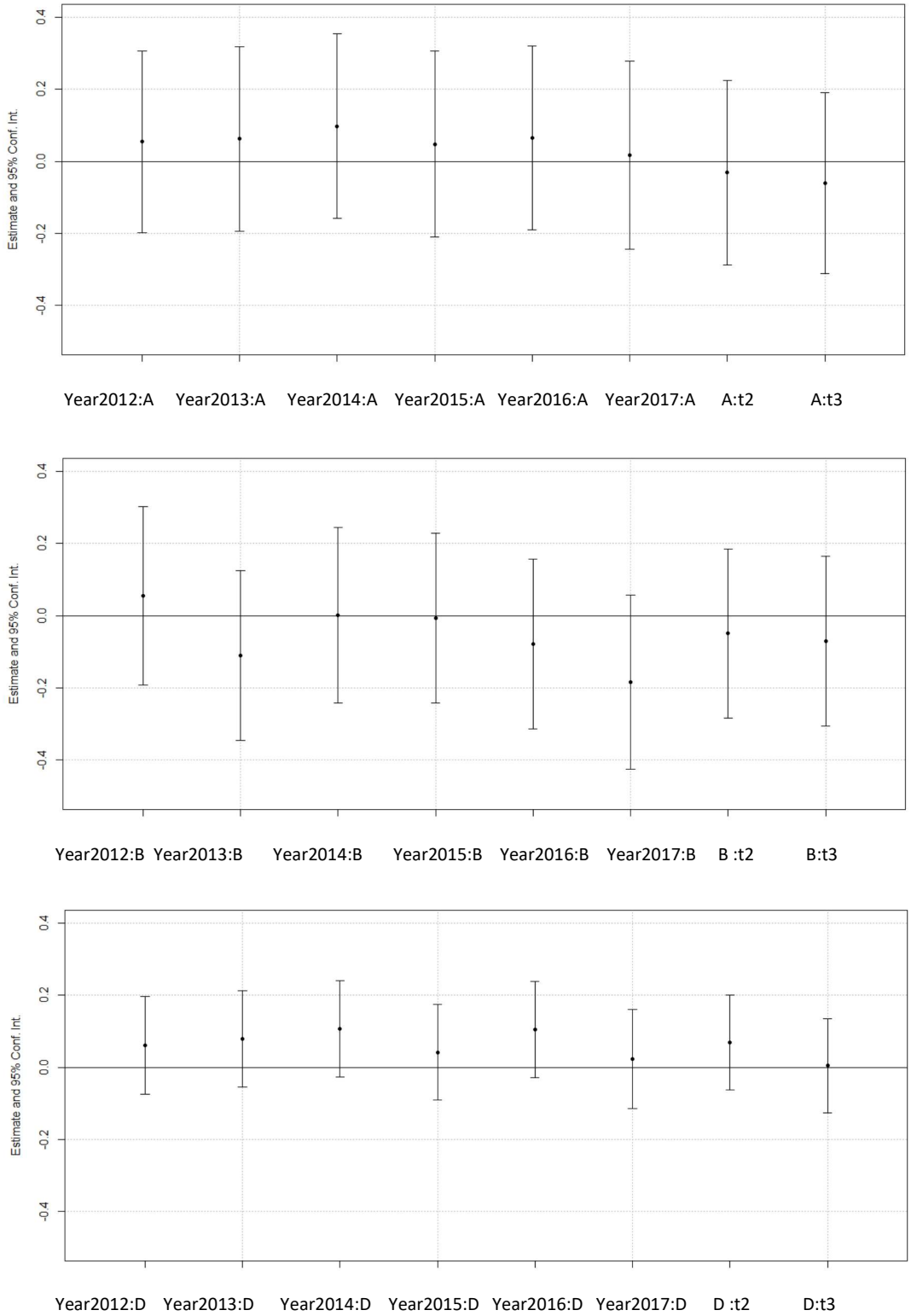


Figure 6.6 Coefficient Graphs for Parallel Trends Test based on estimation of Table 6.4 Column (5)

Based on the results of these parallel trends tests, I would claim that the estimators derived for the treatment effect for Group A seems to be a plausible causal effect of the first CHS zone downsizing on housing prices. These estimates range from not being significantly different from 0.0% in Column (5), to -13.59% in Column (4). With regard to Group B, even though the fluctuation gap between Group B and Group C is larger than between Group A and Group C, there is still some evidence of parallel trends holding. Here, all the point estimates of the second downsizing effect are negative, but generally not statistically significant, and never with housing fixed effects in Column (3) and (5).

6.4 Other Insights from the Regression Results

Results from Column (1) to (3) suggest an unexpected negative impact of both downsizings on sales prices in Group D, most of which are statistically significant. Since the area is always outside of the CHS Zone, I expect that Group D should have no significant increase or decrease in its prices compared to Group C in the baseline period. I note however, that the effect becomes insignificant once we include early year/group interaction terms in Column (4) and (5). In fact, Group D experienced similar negative effects in sales prices relative to Group C during the second treatment period also, dropping 6.21% though not in the subsequent columns. One possible explanation for any negative effects observed is that, as the CHS zone shrinks, houses in Group C gain a benefit of increased scarcity value. Thus, any loss in Group D's Sales value relative to Group C may reflect the latter's rise, rather than the former's fall. Another possible explanation is that there might be other potential confounding factors on housing prices going on in Group D in periods t2 and t3, which is unobserved in our data. For example, there might be more development activities of terrace houses or apartments in Group D, where the prices of such units are normally lower than those for detached family houses, or cause the prices of adjacent detached properties to fall.

7. Robustness Checks

In this section, I test the stability of my main results given in Table 6.4. To do so, I study the effect of each downsizing separately. First, I use only sales from Groups A and C to study the impact of the first school zone downsizing, and identify Period t1 as the pre-treatment period in comparison to the post-treatment period of both Periods t2 and t3. Following a similar logic, I create a subset of data for Groups B and C to estimate the effect of the second downsizing and identify the combination of Periods t1 and t2 as the pre-treatment period in comparison to the post-treatment period of Period t3. The results of two studies are reported in Panel (A) and Panel (B) of Table 7.1, respectively.

Panel A of Table 7.1 reports the estimation of the treatment effect size from the first treatment event. Most of the magnitude of the estimates increased ((1), (3) and (4)), as did their significance levels compared to those reported in Table 6.4. The effect size from the first downsizing now ranges from not significantly different from zero (Column (5)) to -15.32% (Column (4)) for houses in Group A, compared to houses from Group C in Period t1. Similarly, Panel B of Table 7.1 reports the results of estimations for the effect size from the second downsizing of CHS. Note that in this identification, houses from Group C in both Periods t1 and t2 are the baseline group. Here, all estimation results of the parameter of interest have not changed much in either magnitude or significance compared with those from Table 6.4. The baseline model loses significance (Column (1)), and with years of Period t1 parallel trends controlled for, the estimate changes from negative to positive in Column (4), but still insignificant. The range of the effect size of housing price drop in Group B due to the second treatment is between -5.84% to +0.5%, though with most not significant. In sum, the estimation results for Group A and Group B evaluated separately are fairly robust to changes in specification.

Regarding the parallel trends test with the robustness check for the first downsizing, I follow the same test strategy illustrated in Section 6.3. In Column (4) and (5) of Table 7.1, including the interaction terms between Group A (in Panel A) or Group B (in Panel B) with year dummies. Figure 7.1 (A) indicates

Table 7.1 Empirical Results of Robustness Check

Panel A:					
	Column (1)	**Column (2)**	**Column (3)**	**Column (4)**	**Column (5)**
Dependent Variable:	*LogHousePrices*				
GroupA*(t2+t3)	-0.0593024** (0.0232629)	-0.07038**** (0.01690)	-0.094430*** (0.029977)	-0.16627*** (0.057302)	-0.023587 (0.127597)
N:	8090	8072	8090	8072	8090
R^2	0.10599	0.5376	0.2738	0.52929	0.27444
Adjusted R^2	0.10566	0.5327	-2.0186	0.52399	-2.0253
F-Statistics	319.545 on 3 and 8086 DF, p-value: < 2.22e-16	109.2 on 85 and 7986 DF, p-value: < 2.2e-16	91.7127 on 8 and 1946 DF, p-value: < 2.22e-16	99.7154 on 90 and 7981 DF, p-value: < 2.22e-16	52.4136 on 14 and 1940 DF, p-value: < 2.22e-16
Panel B:					
	Column (1)	**Column (2)**	**Column (3)**	**Column (4)**	**Column (5)**
Dependent Variable:	*LogHousePrices*				
GroupB*t3	-0.060145*	-0.05893**	-0.020857	0.0048067	-0.021490
	(0.035214)	(0.0262)	(0.047925)	(0.033535)	(0.064769)
N:	7728	7715	7728	7715	7728
R^2	0.07519	0.5021	0.26713	0.50464	0.27145
Adjusted R^2	0.07483	0.4965	-2.076	0.49859	-2.0695
F-Statistics	209.3 on 3 and 7724 DF, p-value: < 2.2e-16	89.44 on 86 and 7628 DF, p-value: < 2.2e-16	74.5591 on 9 and 1841 DF, p-value: < 2.22e-16	83.4809 on 93 and 7621 DF, p-value: < 2.22e-16	42.7074 on 16 and 1834 DF, p-value: < 2.22e-16
Fixed effects:					
Groups	Y	N	N	N	N
Periods	Y	N	N	N	N
Houseid	N	N	Y	N	Y
Year	N	Y	Y	Y	Y
HouseFeatures	N	Y	N	Y	N
Suburb	N	Y	N	Y	N

Notes: 1. Standard errors are reported in parentheses.
*2. Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 ' ' 1*

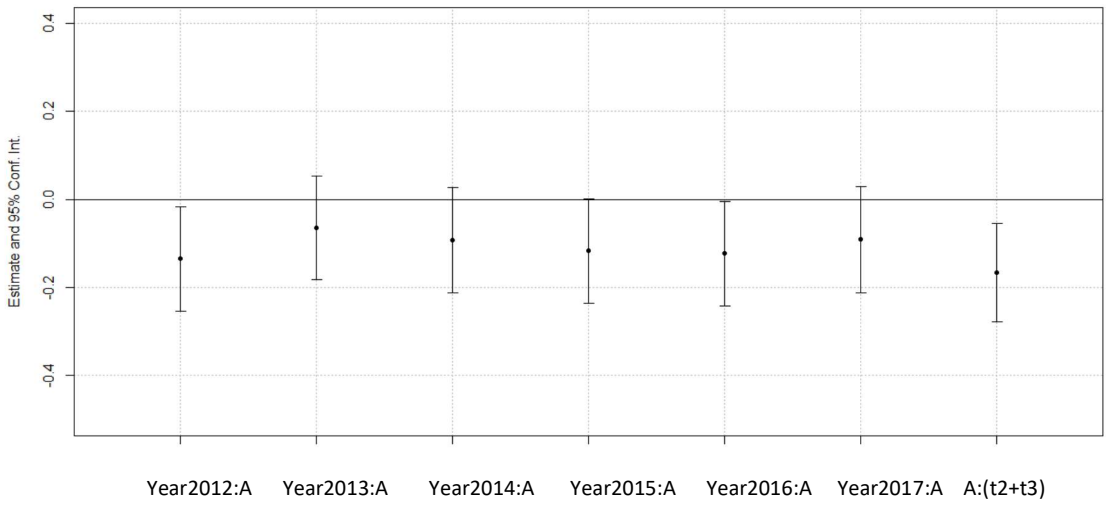


Figure 7.1 (A) Graph of Parallel Trends Test for Table 7.1 Panel A Column 4

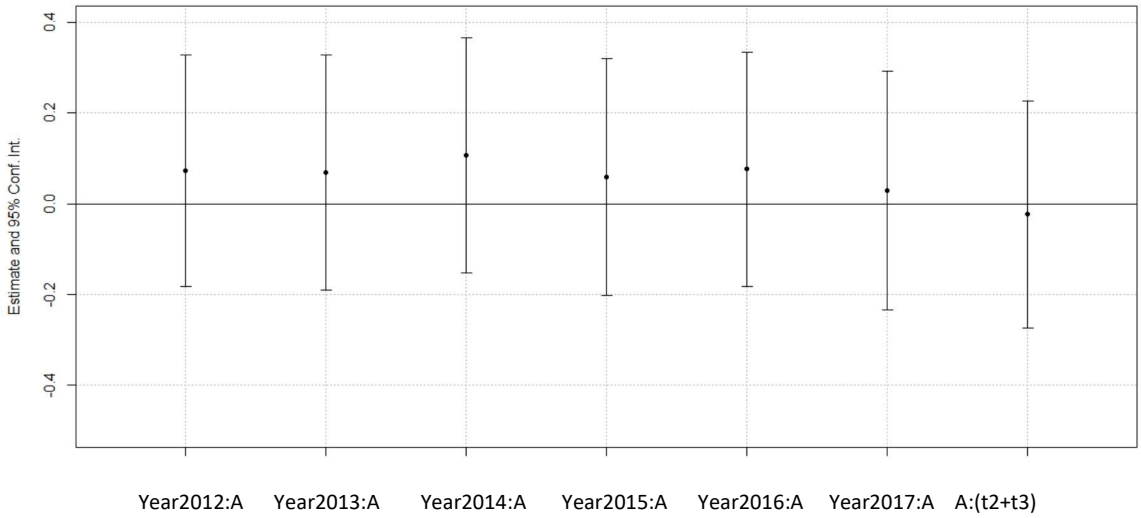


Figure 7.1 (B) Graph of Parallel Trends Test for Table 7.1 Panel A Column 5

that even though houses from Group A consistently have a lower price compared to houses from Group C in Year2018JanApr, the trends of price changes in both groups are not in parallel prior to treatment in 2012 or 2016 for Column (4). In contrast, Figure 7.1 (B) shows that the parallel trends assumption approximately holds between Group A and Group C for each year with the average gap of

prices between the range of 3% to 10% for Column (5). Therefore for the first treatment, I prefer the estimation result from fixed effects (Column (3) or (5)).

Here, on average, losing access to CHS has no significant effect on houses affected in Group A in Column (5), though it does in Column (3). Between these conflicting results, which one to believe would depend in part on the interpretation of the pre-treatment parallel trends results. Personally, I would put more confidence in Column (5) because it allows for linear trend differences among groups (by including interaction terms of pre-treatment years and groups). In other words, if the parallel trends assumption does not hold, the estimate of Column (3) which impose it might be biased; while if the parallel trends hold perfectly, the inclusion of those interaction terms will be dropped in the estimation. At the same time, it seems possible though unlikely that the inclusion of a number of additional interaction terms in Column (5) has reduced the precision of its estimates, which might account for the loss of significance of Group A sales price effects.

Let's move on to the parallel trends test for the second downsizing. Once again, the parallel trends assumption is rejected for Column (4) in Figure 7.2 for the years 2014, 2015 and 2016, but not so in Column (5) shown in Figure 7.2(B) This might again point to emphasizing fixed effects results for Group B in this robustness check, which again finds no significant treatment effect in either Column (3) or (5), just as in the original analysis.

As a final robustness check, I cluster standard errors in my fixed effects regressions to the level of HouseId. To do so, I calculate cluster-robust variance estimators for results from Model (9) in the main results (Table 6.4 Column (3)) and the robustness check with different identification of pre- and post-treatment periods (Table 7.1 Column (3)). The method I employed is "CR1" in the estimation of cluster robust standard errors, which multiplies the original form of the sandwich estimator developed by Liang and Zeger (1986) by m/(m-1), where m is the number of clusters. In a t-test, I choose to specify

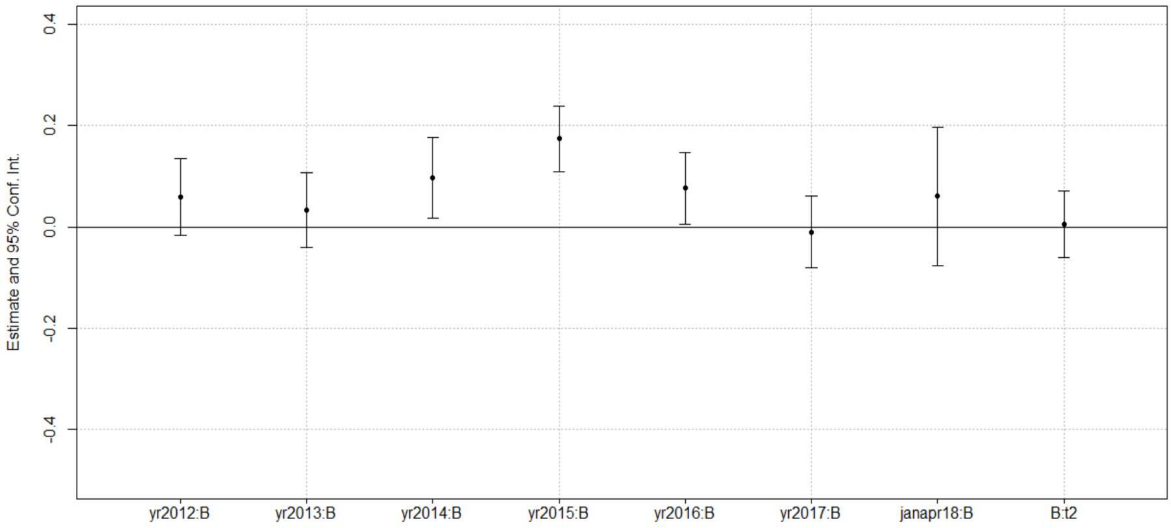


Figure 7.2 (A) Graph of Parallel Trends Test for Table 7.1 Panel B Column (4)

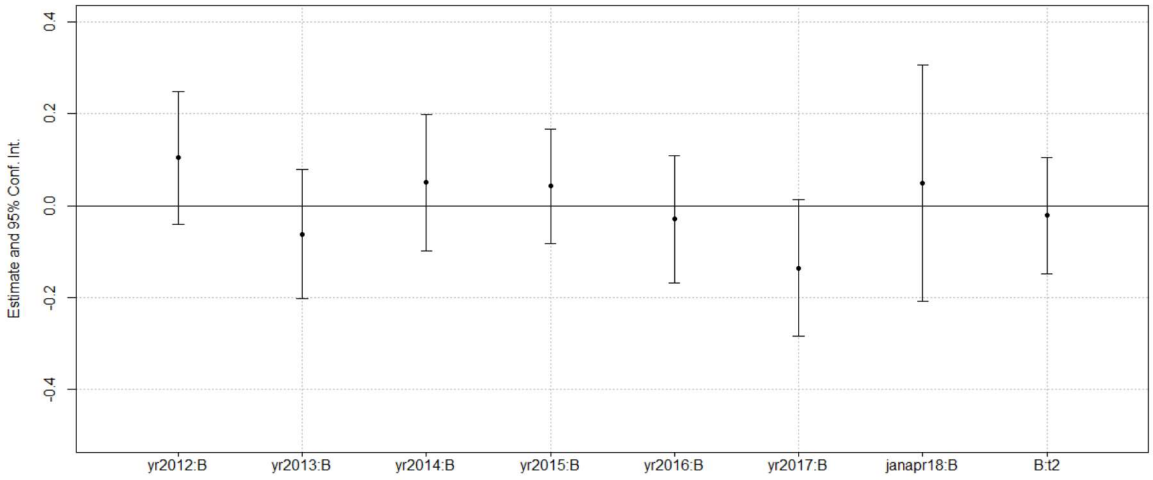


Figure 7.2 (B) Graph of Parallel Trends Test for Table 7.1 Panel B Column (5)

the degree of freedom as the difference of the number of unique HouseId minus one. Table 7.2 reports my estimators with standard errors clustered on HouseId in comparison to results reported without cluster adjustment32. After clustering, the significance of the first downsizing effect gets improved by

[32] The coef-test() function from the clubSandwich package in R.

Table 7.2 Robustness Check with Clustered Standard Errors

Coefficients	Results from	Estimator	Clustered on houseid
GroupA*t2	Table 6.4 Column (3)	-0.0915505**	-0.09155***
		(0.0415798)	(0.0352)
GroupB*t2		0.0026474	0.002645
		(0.0447062)	(0.0393)
GroupD*t2		-0.0030742	-0.00307
		(0.0257712)	(0.0271)
GroupA*t3		-0.1184989***	-0.1184989****
		(0.0382585)	(0.0289)
GroupB*t3		-0.0206929	-0.02069
		(0.0439728)	(0.0486)
GroupD*t3		-0.0640739**	-0.0640739***
		(0.0262021)	(0.0245)
GroupA*t2	Table 6.4 Column (5)	-0.0316010	-0.03160
		(0.1304383)	(0.1065)
GroupB*t2		-0.0490009	-0.04900
		(0.1196120)	(0.1107)
GroupD*t2		0.0685539	0.06855
		(0.0668812)	(0.0911)
GroupA*t3		-0.0607046	-0.6070
		(0.1276822)	(0.1017)
GroupB*t3		-0.0704909	-0.7049
		(0.1196519)	(0.1127)
GroupD*t3		0.0046656	0.00467
		(0.0667134)	(0.0851)
GroupA*(t2+t3)	Table 7.1 Panel A Column (3)	-0.094430***	-0.0944****
		(0.029977)	(0.0186)
GroupA*(t2+t3)	Table 7.1 Panel A Column (5)	-0.023587	-0.0236
		(0.127597)	(0.0789)
GroupB*t2	Table 7.1 Panel B Column (3)	-0.020857	-0.0209
		(0.047925)	(0.0407)
GroupB*t2	Table 7.1 Panel B Column (5)	-0.021490	-0.0215
		(0.064769)	(0.0516)

Notes: 1. Standard errors are reported in parentheses.
*2. Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 ' ' 1*

one star level (either from 0.05 to 0.01 or from 0.01 to 0.001), while the effect of the second downsizing still remains insignificant.

8. Discussion and Conclusion

8.1 Summary and Reflections

In this research, I have used a difference-in-differences (DID) approach to estimate the effect of school access right on housing sales prices through a quasi-experiment of two consecutive school zone downsizings of a much in-demand state high in Christchurch, New Zealand. This research identified a generally negative treatment effect in the housing prices among houses losing access to a desirable secondary school zone when the reassigned school zone was much less in demand, however the magnitude of the housing price drop is mixed. My estimation shows the first downsizing decreases housing prices between 2.33% and 15.32% with the smaller estimates not statistically significant. However, I find the second downsizing affected prices by between +0.48% to -7.76%, though with more estimates not significant. The above estimations are robust to changes of pre- and post-treatment periods identification and to house-level clustering standard errors.

Tests in the pre-treatment period suggest parallel trends cannot be rejected for the first downsizing in most cases, but could for some comparisons for the second downsizing. As always, such tests cannot prove parallel trends do hold even for the first downsizing. I conclude that the high school's first downsizing very likely had a negative effect on housing price, but the size would be small. Given that this natural experiment should identify an "upper bound" effect, my estimate range tends to be smaller than those found in the previous literature.

In the existing related literature, a positive impact of school quality on housing prices has been consistently found, though with different effect sizes due to differences in institution, data source and

measurement methods. Several previous papers have found similar results based on identical difference-in-differences methods but from different contexts. For example, Bogart and Cromwell (2000) also find that loss of access to a good school district is associated with a 9.9% drop in house prices based on data from a wealthy school district in Ohio, U.S.A. My research provides further evidence that there is a housing price premium from access rights to a quality school zone. I find a possible upper bound of such an effect of 15%, but the effect size is not constant across reasonable specifications. Hopefully, such findings could help policymakers have a better understanding of their efforts in promoting education equality through school zoning systems.

As reviewed in Section 2.3, studies on school quality capitalization face the challenges of systematic, consistent measures of school quality, and the endogeneity of school quality. In addition, as most previous researchers aim to disentangle the mystery between school quality and housing prices from a city or country level, the heterogeneity of each individual school zone or different levels of schools also make the estimation work of a causal relationship between school quality and housing prices harder since each individual house might face a totally diverse combination of alternative school choices, not to mention the differences in student enrolment schemes at different administration levels. The research design of my study, a near-random quasi-experiment of CHS zone downsizings where the school zone change event is exogenous to the housing sales, has avoided such issues.

8.2 Limitation of My Research

From an empirical analysis perspective, my research has several limitations, which could lead to a result that the estimates presented here might not accurately reflect the causal impact of the change in the school zone. First, there is the potential for undetected violation of the DID model assumption of no anticipation of the change. The context of school downsizing policy setting and the procedure of how the public notice was given does not hint at possible anticipation from householders, but I cannot prove that quantitatively. Similarly, for the parallel trends assumption, the results discussed in

Section 6. 3 have already shown that even though the assumption of parallel trends cannot be rejected in most cases, there are still a big range of gaps in group trends across years in the pre-treatment period of my study. In addition, there are apparent violations of parallel trends in certain years between Group B and C or Group D and C. Second, the existence of missing house characteristic variables that affect housing prices, such as demographic data regarding house purchase, may lead to other confounding factors uncontrolled in my analysis, which would affect both the direction and the size of the estimators. Third, the hedonic house pricing model I used implicitly assumes that the housing market is in equilibrium. However, the overall supply and demand of the housing market in the New Zealand city in which the downsizings occurred have both fluctuated in the years following the 2011 earthquake, and subsequent expansion of housing in new areas.

8.3 Ideas for Future Research

Given the lack of close substitutes to CHS, my research provides an empirical test of the upper bound estimate of school quality impacts on housing prices from one case study. Similar research could be repeated for more school zones experiencing sudden changes to get a better idea of the distribution of such effects, so as to get a more generalized knowledge of school zone impact on housing prices. Further studies could also be conducted to study the heterogeneous treatment effect on housing prices from school zone change if extensive demographics data from house buyers and sellers and transaction process data are accessible.

Appendix 1

Table 4.1 Description of Cashmere High School Zones Downsizings

Original CHS Zone	First Downsizing	Second Downsizing
From the junction of Frankleigh Street and Sparks Road	Starting from the junction of Frankleigh Street / Sparks Road / Lyttleton Street,	Starting from the junction of Frankleigh Street / Sparks Road / Lyttelton Street;
Along Sparks Road to the intersection of Sparks and Rydal Street,	West along Sparks Road to the intersection of Sparks and Rydal Street,	West along Sparks Road to the intersection of Sparks and Rydal Street
Along Rydal, into Northaw into Rollesby Street,	South along Rydal Street to Northaw Street • West on Northaw Street to Rollesby Street (including addresses to 16 Northaw Street), • South on Rollesby Street to Rydal Street	South along Rydal Street to Northaw Street ,West on Northaw Street to Rollesby Street (including addresses to 16 Northaw Street) ,South on Rollesby Street to Rydal Street
South along Rydal Street into Leistrella Road to the intersection with Hoon Hay Road.	South along Rydal Street to Leistrella Road	South along Rydal Street to Leistrella Road
South along Hoon Hay Road to Blakiston Street (including Barossa and Penmarc),	East on Leistrella Road to Hoon Hay Road (including all addresses on Leistrella Road up to number 35).	East on Leistrella Road to Hoon Hay Road (including all addresses on Leistrella Road up to number 35)
Along Kaiwara Street to Cashmere Road.	South along Hoon Hay Road to Blakiston Street (including Barossa Lane) • West on Blakiston Street to Kaiwara Street,	South along Hoon Hay Road to Blakiston Street (including Barossa Lane) West on Blakiston Street to Kaiwara Street
	North, then West and South on Kaiwara Street to Cashmere Road (including Penmarc Lane),	North, then West and South on Kaiwara Street to Cashmere Road (including Penmarc Lane)
West along Cashmere Road to Hoon Hay Valley Road,	West along Cashmere Road (including Boonwood Close) to Hoon Hay Valley Road,	West along Cashmere Road (including Boonwood Close) to Hoon Hay Valley Road
	Along Hoon Hay Valley Road to the southern end of the road, then	Along Hoon Hay Valley Road to the southern end of the road, then
Then across to the Summit Road and	Following a line directly across to the Summit Road / Worsleys Road intersection (including all Worsleys Road Addresses) and	Following a line directly across to the Summit Road / Worsleys Road intersection (including all Worsleys Road Addresses) and
Along the Summit Road to the junction with Gebbies Pass Road, including all residences with access from the Summit Road.	Along the Summit Road to the junction with Gebbies Pass Road (including all residences accessed from this portion of the Summit Road, including addresses 685, 703 and 706 Gebbies Pass Road),	Along the Summit Road to the junction with Gebbies Pass Road (including all residences accessed from this portion of the Summit Road, including addresses 685, 703 and 706 Gebbies Pass Road),

From the junction of Gebbies Pass and the Summit Roads to Port Levy, including all residences on the Port Levy Road and the Camp Bay Road and all other roads feeding to the south side of Lyttelton Harbour.	From the junction of Gebbies Pass and the Summit Road, in a straight line to Port Levy	From the junction of Gebbies Pass and the Summit Road, in a straight line to Port Levy
The remaining boundary shall run from the intersection of Frankleigh and Lyttelton Streets,		
Northwards along the west of Lyttelton Street to Neville Street.	Including all residences on Purau Port levy Road, Camp Bay Road.	Including all residences on Purau Port Levy Road, Camp Bay Road.
East along Neville Street to Barrington Street.	Including addresses in Port Levy; including Old Port Levy Road, Wharf Road, Fields Road, Fernlea Point Road, addresses down to 1251 Western Valley Road, Richfield Road, addresses down to 899 Port Levy-Pigeon Bay Road, Pa Road, Jetty Road, Puari Road, and Putiki Road.	Including addresses in Port Levy; including Old Port Levy Road, Wharf Road, Fields Road, Fernlea Point Road, addresses down to 1251 Western Valley Road, Richfield Road, addresses down to 899 Port Levy-Pigeon Bay Road, Pa Road, Jetty Road, Puari Road, and Putiki Road.
Northwest along Barrington Street to Lincoln Road,		
then along Lincoln Road to Moorhouse Avenue.	From Port Levy, West along the coast line (past Diamond Harbour) to Governors Bay, then north east along the coast line to a point on the coast due west of 20 Park Terrace (in Corsair Bay),	From Port Levy, West along the coast line (past Diamond Harbour) to Governors Bay, then north east along the coast line to a point on the coast due west of 20 Park Terrace (in Corsair Bay)
East along Moorhouse Avenue to the junction of Waltham Road.	North along Park Terrace to Governors Bay Road,	North along Park Terrace to Governors Bay Road
South along Waltham Road to the intersection with Brougham Street	West along Governors Bay Road to 90 Governors Bay Road, then	West along Governors Bay Road to 90 Governors Bay Road, then
and then along Brougham Street to the intersection with Opawa Road	Following a line from 90 Governors Bay Road, north west to the Summit Road / Rapaki Road (Track) intersection.	Following a line from 90 Governors Bay Road, north west to the Summit Road / Rapaki Road (Track) intersection
In a south easterly direction along Opawa Road to its junction with Port Hills Road.	North along Rapaki Road to 90 Rapaki Road, then	North along Rapaki Road to 90 Rapaki Road, then
West covering streets along the south side of the Port Hills Road to Rapaki Road.	Following a line north to the Port Hills Road / Opawa Road (SH76) intersection (including addresses 361-385 Port Hills Road)	Following a line north to the Port Hills Road / Opawa Road (SH76) intersection (including addresses 361-385 Port Hills Road)
South along Rapaki Road to its junction with Summit Road.		West along Port Hills Road to Centaurus Road (including Lucienne Place)
From the junction of Rapaki Road and the Summit Road in a straight line to the eastern end of Corsair Bay, to include the house at number 20 Park Terrace.	North on Opawa Road (SH76) to Opawa Road,	West along Centaurus Road to Aynsley Terrace

		North along Aynsley Terrace to address number 75-75A
		West to the Ōpāwaho / Heathcote River
		West along the South bank of the Ōpāwaho / Heathcote River to the Wilsons Road Bridge
		North west to the Wilsons Road / Waltham Road / Eastern Terrace intersection
		South west along Eastern Terrace to Tennyson Street
		West along Tennyson Street to Southampton Street
		North along Southampton Street to Croydon Street (including Ikamatua Lane)
		North east along Croydon Street to Huxley Street
		North west along Huxley Street to Colombo Street
	West on Opawa Road, across the river to Brougham Street,	North along Colombo Street to Brougham Street (SH76)
	West on Brougham Street to Selwyn Street,	West on Brougham Street (SH76) to Selwyn Street
	South along Selwyn Street to Coronation Street,	South along Selwyn Street to Coronation Street
	South west along Coronation Street to Barrington Street,	South west along Coronation Street to Barrington Street
	South along Barrington Street to Stourbridge Street (including Sefton Place and Kinver Place)	South along Barrington Street to Stourbridge Street (including Sefton Place and Kinver Place)
	South west along Stourbridge Street to Lyttelton Street	South west along Stourbridge Street to Lyttelton Street
	South along Lyttelton Street to the junction with Frankleigh Street and Sparks Road.	South along Lyttelton Street to the junction with Frankleigh Street and Sparks Road.

Appendix 2:

Table 5.2 Illustration of Canonical Two-way Fixed Effect DID Model

<table>
<tr><th>Y</th><th>Control Group</th><th>Treatment Group</th><th>Across-group Difference</th><th>Difference-in-Differences</th></tr>
<tr><td>Pre-treatment Period(T=0)</td><td>E{Yit|D=0, T=0}</td><td>E{Yit|D=1, T=0}</td><td>E{Yit|D=1, T=0}-
E{Yit|D=0, T=0}</td><td rowspan="2">Across-group Difference of Post-Treatment Period
minus
Across-group Difference from the Pre-Treatment Period: (E{Yit|D=1, T=1}-
E{Yit|D=0, T=1})-
(E{Yit|D=1, T=0}-
E{Yit|D=0, T=0})</td></tr>
<tr><td>Post-Treatment Period(T=1)</td><td>E{Yit|D=0, T=1}</td><td>E{Yit|D=1, T=1}</td><td>E{Yit|D=1, T=1}-
E{Yit|D=0, T=1}</td></tr>
<tr><td>Across-time Difference</td><td>E{Yit|D=0, T=1} -
E{Yit|D=0, T=0}</td><td>E{Yit|D=1, T=1} -
E{Yit|D=1, T=0}</td><td colspan="2" rowspan="2">τ_{ATE}=
E{Yit|D=1, T=1}+
E{Yit|D=0, T=0}-
E{Yit|D=1, T=0}-
E{Yit|D=0, T=1})-</td></tr>
<tr><td>Difference-in-Differences</td><td colspan="2">Across-time Difference of Treatment Group minus
Across-time Difference of Control Group:
(E{Yit|D=1, T=1} - E{Yit|D=1, T=0})-
(E{Yit|D=0, T=1} - E{Yit|D=0, T=0})</td></tr>
</table>

Appendix 3: Explanation of Expanded Two-Treatment Four-Group Three-Period DID Model

$$Y_{it} = \alpha_i + \beta_1 GroupA + \beta_2 GroupB + \beta_3 GroupD + \gamma_1 t2 + \gamma_2 t3 + \sum_1^i \theta_i X_i + \sum_1^t \eta_t Year_t + \rho_1 GroupA * t2 + \rho_2 GroupA * t3 + \rho_3 GroupB * t2 + \rho_4 GroupB * t3 + \rho_5 GroupD * t2 + \rho_6 GroupD * t3 + \varepsilon_{it}. \quad (8)$$

Table 5.3 Interpretation of Coefficient Meanings in Model (8)

Terms	Coefficient	Calculation of differences	Interpretation
$GroupA * t2$	ρ_1	$=(\overline{y_{A,t2}} - \overline{y_{A,t1}}) - (\overline{y_{C,t2}} - \overline{y_{C,t1}})$	delta Y2- delta Y1 across group A and C, or delta YA- delta YC across Periods t2 t1
$GroupA * t3$	ρ_2	$= (\overline{y_{A,t3}} - \overline{y_{A,t1}}) - (\overline{y_{C,t3}} - \overline{y_{C,t1}})$	delta Y3- delta Y1 across group A and C, or delta YA- delta YC across Periods t3 t1
$GroupB * t2$	ρ_3	$= (\overline{y_{B,t2}} - \overline{y_{B,t1}}) - (\overline{y_{C,t2}} - \overline{y_{C,t1}})$	delta Y2- delta Y1across group B and C, or delta YB- delta YC across Periods t2 t1
$GroupB * t3$	ρ_4	$= (\overline{y_{B,t3}} - \overline{y_{B,t1}}) - (\overline{y_{C,t3}} - \overline{y_{C,t1}})$	delta Y3- delta Y1 across group B and C, or delta YB- delta YC across Periods t3 t1
$GroupD * t2$	ρ_5	$= (\overline{y_{D,t2}} - \overline{y_{D,t1}}) - (\overline{y_{C,t2}} - \overline{y_{C,t1}})$	delta Y2- delta Y1 across group D and C, or delta YD- delta YC across Periods t2 t1
$GroupD * t3$	ρ_6	$= (\overline{y_{D,t3}} - \overline{y_{D,t1}}) - (\overline{y_{C,t3}} - \overline{y_{C,t1}})$	delta Y3- delta Y1 across group D and C, or delta YD- delta YC across Periods t t31
constant	α_i		Group mean of C in t1
group A	β_1		Group difference between A and C in t1
group B	β_2		Group difference between B and C in t1
group D	β_3		Group difference between D and C in t1
t2	γ_1		Period difference between t2 and t1 of C
t3	γ_2		Period difference between t3 and t1 of C
X_i	θ_i		Fixed effect of unique house i
$Year_t$	η_t		Fixed effect of individual year t

Appendix 4

Table 6.5 Results of Parallel Trend Test Regression (Corresponding to Column (5) in Table 6.4)

	Model (8)
Dependent Variable:	LogHousePrices
GroupA*t2	-0.0859033
	(0.0544883)
GroupB*t2	-0.1039864
	(0.0669434)
GroupD*t2	0.0073428
	(0.0374655)
GroupA*t3	-0.1150069**
	(0.0541235)
GroupB*t3	-0.1254763*
	(0.0671756)
GroupD*t3	-0.0565454
	(0.0381693)
t2	0.3570941****
	(0.0257167)
t3	0.5309231****
	(0.0259770)
Year2013*A	0.0077857
	(0.0606935)
Year2014*A	0.0429306
	(0.0625481)
Year2015*A	-0.0067301
	(0.0564880)
Year2016*A	0.0111786
	(0.0605736)
Year2017*A	-0.0373027
	(0.0629335)
Year2018JanApr*A	-0.0543023
	(0.1289217)
Year2013*B	-0.1662774
	(0.0756965)
Year2014*B	-0.0538211
	(0.0792518)
Year2015*B	-0.0616197**
	(0.0663042)
Year2016*B	-0.1335479*
	(0.0745447)
Year2017*B	-0.2393695***
	(0.0787623)
Year2018JanApr*B	-0.0549855
	(0.1257970)
Year2013*D	0.0177507
	(0.0432878)
Year2014*D	0.0454727
	(0.0413589)
Year2015*D	-0.0195143
	(0.0402926)
Year2016*D	0.0433017
	(0.0417223)
Year2017*D	0.0373236
	(0.0435377)
Year2018JanApr*D	-0.0612111

	(0.0689873)
Fixed effects:	
Groups	N
Periods	N
Houseid	Y
Year	Y
N:	14738
R^2	0.28168
Adjusted R^2	-1.93
F-statistics	44.2743 on 32 and 3613 DF, p-value: < 2.22e-16

Notes: 1. Standard errors are reported in parentheses.
*2. Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 ' ' 1*