THE EFFECT OF SCHOOL CLOSURES ON HOUSING PRICES IN HAMILTON, ONTARIO

THE EFFECT OF SCHOOL CLOSURES ON HOUSING PRICES IN HAMILTON, ONTARIO

John Merrall, BA

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Arts

McMaster University

MASTER OF ARTS (2021) School of Earth, Environment and Society

TITLE: The Effect of School Closures on House Prices in Hamilton, Ontario

AUTHOR: John Merrall, B.A. (McMaster University, Economics)

SUPERVISOR: Dr. H. Antonio Páez

NUMBER OF PAGES: xi, 116

LayAbstract

Do homebuyers pay a premium to be located closer to a school, and do school closures affect house prices?

An analysis of Hamilton real estate transactions (2005-2017) finds evidence that houses closer to schools sell for more, and that a primary school closure has a negative impact on local house sale prices.

Abstract

Is school accessibility a valued good, and do school closures affect house prices?

This thesis applies two different methods of hedonic regression analysis, augmented by spatial regression methods, to a dataset of Hamilton real estate transactions (2005-2017) to investigate whether the closure of a school in an urban neighbourhood negatively affects house prices in that closed school catchment.

Evidence is found that school accessibility is a valued good, and that the closure of a primary school will negatively affect house prices from the period of closure announcement through several subsequent quarters. The use of spatial analysis corrects for bias in coefficient estimates.

Acknowledgements

This thesis was completed with the assistance of Dr. Antonio Paez (McMaster University) and Dr. Christopher Higgins (University of Toronto); the author also wishes to thank Wendy Stewart and staff at the Realtor's Association of Hamilton & Burlington for providing the dataset without which this work could not have been completed, as well as Dr. Saiedeh Razavi and staff at the McMaster Institute for Transportation and Logistics for facilitating this project.

The author also acknowledges the help given by staff at the Hamilton-Wentworth District School Board and the Hamilton-Wentworth District Catholic School Board in finding historical school catchment data, as well as Dr. Mark Burkey (North Carolina A&T) for his personal communications.

The research was funded by by a Social Sciences and Humanities Research Council Insight Development Grant (430-2017-00627) and a Joseph Armand Bombardier Canada Graduate Scholarship.

Contents

List of Figures

List of Tables

Declaration

This is to certify that this thesis represents my original work except where otherwise noted. Guidance was provided by Dr. Antonio Paez and Dr. Chris Higgins in determining appropriate methods of analysis, interpreting results, and review of literature.

The housing dataset provided by the Realtors' Association of Hamilton & Burlington was cleaned and geocoded by Dr. Chris Higgins, and use of that dataset for this thesis was based on his initial exploratory analysis.

1 Introduction

1.1 Background

Maclellan (2007) notes that a narrative arose in the 1990s in Canada which advocated for a reform of public education to improve accountability, budgeting, and the measure of student achievement. In May 1993, the NDP government in the Province of Ontario charged the Royal Commission on Learning with providing recommendations for the modernization of the Ontario public school system; their report, released in January 1995, included classroom, curriculum, accountability and governance recommendations (Begin, Caplan, & others, 1994). Following this, February 1995 saw the formation of the Ontario School Board Reduction Task Force, assigned with the task of rationalizing school board boundaries and reducing the number of school board trustees (Ontario Ministry of Finance, 1996); its report came in February 1996 after a provincial election that saw the Progressive Conservatives replace the NDP, with their election platform *The Common Sense Revolution* (Progressive Conservative Party of Ontario, 1994) offering a shift from spending money on "consultants, bureaucracy and administration" to "classroom-based budgeting" (p.8).

In 1997, the Conservatives changed public school funding in Ontario with the passing of the *Fewer School Boards Act*, which ended funding of local school boards from the local property tax base; all school funding was in future to be provided by the provincial government. This legislation also pursued the previously set goal of amalgamation of local school boards: on 1 January 1998 in Hamilton, this formed the Hamilton Wentworth District School Board (HWDSB) from the former boards for the city of Hamilton and the county of Wentworth. Standardized testing was also initiated at the grade 3, 6 and 9 levels. School board funding became grant-based, with the amount of funding determined on a per-student basis, and with school boards thereafter required to budget based on a per-pupil funding formula. School boards were reduced from 164 to 72; the number of school trustees in Ontario was reduced from 1900 to 700, with the number of trustees per board also determined by a student-based formula (MacLellan, 2007).

Given the acrimonious political environment in provincial politics in this period, resistance was to be expected. In 2002, the school boards in Toronto, Hamilton and Ottawa failed to submit balanced operating budgets in violation of the Education Act (MacLellan, 2007): HWDSB chair Judith Bishop asserted at the time that education budgeting had "become a political game with students as the pawns instead of the central concern" (Bishop, 2002b). As she noted elsewhere, the shift to the per-student funding formula had turned the HWDSB budgeting process into an exercise in cutting teaching and secretary positions (Bishop, 2002c).

Yet despite cuts to the operating budget, the HWDSB continued to operate in deficit from 1998 onward, funding their deficit through reserves and bank financing; By 2002, the HWDSB faced a $16 million budget deficit (Murray, 2002; Prokaska, 2002). Upon the board's continued refusal to submit a balanced budget, provincial Education Minister Elizabeth Witmer appointed auditor Charles Smedmor to investigate the HWDSB's spending. He found that "HWDSB's deficit is a direct result of the Trustees' unwillingness to close and consolidate schools" (Smedmor (2002b), p.2); while staff had recommended in 1998 the closure of 20 schools, the board had only closed 4 by 2002. Smedmor concluded that their yearly budgeting had been pursued without a long-range capital plan, failing to take advantage of new pupil grants that would finance the construction of new schools, reducing maintenance costs and increasing energy efficiency. This caused the HWDSB to spend money on capital maintenance and staff that should have instead gone toward classroom spending; rather than continued deficits, he found that a balanced budget was attainable in the 2002-3 school year (Smedmor, 2002b, 2002a).

Because Smedmor's report identified HWDSB staff and trustee resistance as an impediment to achieving a balanced budget, he recommended the Minister appoint a special supervisor to take over the Board; Witmer thereupon selected Jim Murray to oversee a rationalization (Murray, 2002; Prokaska, 2002). Murray set a tight timetable for community committees and the board executive council to develop proposals to quickly implement school closures and replacements; trustees were given a choice of proposals, but the final decision remained with Murray. Complaints were aired about the adversarial nature of this method; nevertheless, a number of immediate school closures followed, from 2002-2005

(Cox, 2003c). After Murray's exit at the end of 2003 a process of trustee-led accommodation review was instituted, by which future school closures and capital investments would be planned out.

The following years would see continued public board school closures and building replacements in Hamilton to reduce per-pupil education costs in the face of urban demographic change; board resistance to the consolidation process eventually relented, and this accommodation review process has since become institutionalized at the HWDSB. The Hamilton Catholic board (HWCDSB), facing similar overcapacity issues, followed suit.

A question was eventually posed, by local parents opposing the closure of Prince Philip primary school in West Hamilton in 2013, as to whether such closures can have an effect on housing prices. In February 2013, a local parent brought a suit against a school trustee, alleging that her decision to close Prince Philip School to free up \$8.5 million in budget for renovation and expansion of the neighbouring local school, G.R. Allan, constituted a conflict of interest which she should have declared: since the trustee owned a home in the G.R. Allan catchment, the application argued, she had a "pecuniary interest in the matters discussed and decided at the trustee meeting" where the decision to close Prince Philip was made (Pecoskie, 2013b).

The two positions in the lawsuit, with one side arguing that schools did not affect property values, highlights an important policy question: do schools contribute to the value of real estate property? Property taxes in Ontario are a key source of revenue for public services; therefore, if school proximity and quality is capitalized into housing values, schools could turn out to be generators of value to communities beyond the narrow criteria defined by the accommodation review process.

But how can we even determine whether school closures actually affect house prices? The hedonic regression method, developed from the work of Rosen (1974), is well suited for this analysis: a sample of houses with measurable attributes (describing structure, neighbourhood and environment) has its sale prices regressed on that set of attributes to find the implicit price for each attribute. Thus, we can discover the value that the consumer is willing to pay for each attribute of a house; if school closeness is a valued attribute of a house, its regression coefficient will be statistically significant.

Most school distance analysis in the literature, however, has been in a non-Canadian context - often parenthetical to the study of schools, usually just using distance to school as a spatial control variable - with only Ries & Somerville's (2010) Vancouver study and Des Rosiers et al's (2011) Quebec City study applicable to Canada and neither explicitly studying closures.

It should be added that it is known that conducting applied econometric work without accounting for the underlying spatial dependence and heterogeneity found in real-world spatial datasets can yield coefficients which are biased and inconsistent, and have incorrect standard errors; however, there have been few studies focusing on the value of school distance that have utilized advanced regression methods to determine the value of school accessibility, and only Rajapaksa et al. (2020) performed the analysis using a spatial regression method.

In response, the first part of this thesis will use a dataset of real estate transactions from the city of Hamilton to conduct a spatial regression, in an attempt to gain an unbiased estimate of the true impact of school accessibility on neighbourhood house prices in a specifically Canadian context; the final part of the thesis will instead use a natural experiment approach on the same Hamilton dataset in an attempt to factor out spatial confounders and look for a time-based impact of closures on house prices.

1.2 Literature Review

Using multivariate regression as a method for determining house prices has a long history: remarkably, Haas (1922) wrote a Master's thesis presenting a multiple regression for determining agricultural land prices nearly a hundred years ago. By the 1960s, as noted by Eisenlauer (1968), the development of computer technology allowed the appraisal profession to begin to experiment with computer-aided regression analysis (as a complement to comparables analysis) as a way of refining the process of mass appraisal, driven in part by the profession's need to reduce large discrepancies in valuation that cast their field in a bad light; Blettner (1969) detailed one multiple regression, using lot and house physical characteristics while controlling for neighbourhood variation, which yielded estimates within 15% of the house selling price 78% of the time.

During the development of multiple regression for mass appraisals, Lancaster (1966) presented a philosophical/economic basis for the development of hedonic modeling, which was later developed by Rosen (1974). As Lancaster noted, if it is assumed (as it was back then) that each consumer commodity in a market has its own separate utility, then there is no explanation as to why some commodities are intrinsically substitutes and others are intrinsically complements; even further, when a market sees the introduction of a new commodity to a market with n pre-existing commodities, previous theory would have no explanation for how to translate the old n-dimensional utility function to a new $n+1$-dimensioned one. His solution, to move the concept of utility from commodities themselves to the intrinsic *characteristics*, or goods, that make up a commodity, solved the "new commodity" problem by making a new commodity just a novel combination of existing intrinsic characteristics - and also solved the problem of substitutability by treating close-substitute commodities as closely-matched bundles of characteristics.

This then led to Rosen's (1974) theoretical model for commodities as combinations of intrinsic characteristics: while his paper presents a more complicated economic model than the simple additive functions used by Haas (1922) and Blettner (1969), it is still generally cited as the foundation for the method of hedonic price regression, rather than the pioneering empirical articles mentioned above.

With the development of hedonic regression for the prediction of housing prices, urban geographers were given an empirical tool to investigate the economic impact of neighbourhood characteristics. Per Rosen and Lancaster's theoretical model, as per Haas and Blettner's empirical model before them, a house can be considered a bundle of goods: each good that makes up a house transacts at a price, and the total price of a house is just a function of the prices of individual goods that make up a house. So, given a dataset containing a matrix of n house transactions, where each house has a recorded transaction price P and a vector of house characteristics X of length i, a regression of the form

$$P = X\beta + \epsilon$$

will give us a best-fit vector of i price coefficients, $\beta_0 ... \beta_i$, each of which represents an estimate of the price for each house characteristic $X_0 ... X_i$. (X_0 is a vector of ones to give us a β_0 intercept coefficient).

Importantly, there is no need to limit the inputs to just house characteristics: it was long ago recognized by appraisers that beyond the structural characteristics of houses, there are neighbourhood and accessibility characteristics that also contribute to the house's "bundle of goods." Trivially, this can be represented mathematically as

$$P = X\beta + N\gamma + A\delta + \epsilon,$$

where now each house has a vector X of house structural characteristics, but also a vector N of quantified neighbourhood characteristics and a vector A of quantified accessibility characteristics. Various house price effects for neighbourhood and accessibility characteristics have since been demonstrated: census data, accessibility to employment, distance to central business district, distance to shopping areas, air pollution, highway and airport noise, and so on. And of course such neighbourhood characteristics will autocorrelate, smoothly or not, over space.

Some such neighbourhood or accessibility characteristics can pertain to schools: Oates (1969), for example, demonstrated a relationship between local school board expenditures (which are a determinant of school quality in the United States) and house prices. Following Oates, there has been much interest in the impact of school "quality" on house prices (Nguyen-Hoang & Yinger, 2011; Ross & Yinger, 1999): multiple regression has been used to determine the impact on house prices from different measures of "quality" such as school proficiency scores (Downes & Zabel, 2002; Haurin & Brasington, 1996), a publicized state "report card" (Figlio & Lucas, 2004), or a more abstract measure of "value added" that attempts to quantify the amount by which a school can raise achievement of students given their socioeconomic background (e.g. Hayes & Taylor, 1996). Within this area of study, a natural experiment approach is sometimes pursued, where a house's change in school catchment between schools of differing quality is used to identify school quality's impact on house prices (Ding, Bollinger, Clark, & Hoyt, 2020; Machin & Salvanes, 2016; Ries & Somerville, 2010).

Where demographic shifts are driving changes in school boundaries, such as closure of underutilized urban schools, the above studies' findings would likely be important inputs for policy decisions: if regression can show that school quality is on average valued by homeowners, then we can empirically demonstrate that redistricting and closure-funded capital improvements may provide a net improvement to school quality for the remaining schools, and thus perhaps an improvement to house prices. But what about a distance effect? Do homebuyers value simple accessibility to local schools, and are residents losing something of value when their nearest school is closed?

Like many neighbourhood amenities, a school is a socialized good; that is, its provision is paid for by all taxpayers, but its benefits only accrue to those who make use of it. If school closeness has positive utility, and those homeowners who are closer are not otherwise paying a premium for that access otherwise (e.g. in their property tax), then this accessibility becomes a good that can be purchased by outbidding for the house that is closer. We should be able to see its value by adding school distance to a regression analysis of housing prices. Certainly McKibbin (1940) already noted long ago that appraisers in the United States identified distance to a local school as both a negative for house prices at close distance (where negative externalities such as noise and traffic congestion may depress value) and a positive for house prices (at distances far enough away that the benefit of accessibility would outweigh the school's negative impacts).

Many recent papers which include some form of distance-to-school function in a multiple regression analysis (Chin & Foong, 2006; Colwell & Guntermann, 1984; Des Rosiers, Lagana, & Theriault, 2011; Emerson, 1972; Kane, Staiger, Samms, Hill, & Weimer, 2003; Metz, 2015; Owusu-Edusei, Espey, & Lin, 2007; Sah, Conroy, & Narwold, 2016) have, varyingly, investigated McKibbin's qualitative finding that distance to school does matter to house prices. In some papers a maximum benefit distance is also found. An important consideration, as Chin

(2006) notes, is that accessibility of the local primary school may not just be valued simply for a student's shorter travel time; closeness also may influence parents' ability to exercise influence, for example by interaction with teachers and administration, or participation in school-based parent groups. If Chin is right that this greater parent influence also has utility, then its monetary value will also be found by regression.

Within the above-stated set of papers, however, there is tremendous variation in both results and in functional form used for the school distance variable. Emerson's (1972) early paper found a positive coefficient for log distance to school, at 90% significance; this indicated a negative (nuisance) value for school access, contrary to many subsequent papers' findings. However Colwell & Guntermann (1984), using a $1/d$ function alongside a second "negative externality" function, found a positive effect (even net of negative externality) for school access, to a high significance - although, importantly, the authors did warn that their "externality" function may simply have been compensating for the bad behaviour of a $1/d$ function as d becomes small. Des Rosiers, Lagana, & Theriault (2011) combined school distance with school size to find an "optimal" size of 350-400 pupils with an optimal distance of 300-500m; their published article unfortunately doesn't seem to investigate the robustness of the gamma transformation needed to achieve their result.

Kane, Staiger, Samms, Hill, & Weimer (2003), controlling for school quality with a school test score in standard deviation units, found a negative coefficient for linear distance, indicating a positive value for school access; these distance coefficients remained, though reduced in magnitude, after controlling for various neighbourhood fixed effects. The authors also reported that while they did test the quadratic of distance, it was discarded as being not a statistically significant addition to their model; interestingly, they also found that the school distance coefficient was indistinguishable from zero in high-income tracts, but significant in lower-income neighbourhoods - suggesting that school accessibility is in fact valued more in low-income neighbourhoods. Chin & Foong (2006) found a positive coefficient to distance from prestigious primary schools; their study, though, concentrating as it did on prestige schools in Singapore, may mean their finding is irrelevant in the Canadian public school context. Owusu-Edusei, Espey, & Lin (2007) studied the effect of school distance on house prices in Greenville SC differently, dispensing with a linear distance function entirely and instead dividing their house dataset into less-than-average, average and above-average closeness to schools; they found houses with above-average closeness to schools had higher prices.

Metz (2015) found a negative value for school access in a first basic regression, a statistically insignificant value in a second regression that incorporated census tract fixed effects, and a positive value in a third regression that incorporated school catchment fixed effects. As the author notes, unmeasured neighbourhood effects would render their first regression's results questionable; while when census tracts are smaller than school catchments, part of the value of school closeness may capitalize instead into the census tract, meaning their estimates in the second regression would be biased downward. The peculiarity with school catchment fixed effects, of course, is that catchments are collinear with school test results, meaning that school test scores cannot be included in the regression; they instead capitalize into the catchment, not a problem in itself as long as the effect of school scores is not important to the authors' thesis. A fourth regression by Metz, using dummy variables for distances similar to Owusu-Edusei, Espey, & Lin (2007), confirmed a slight disamenity to school closeness at under 500 feet distance, "ideal" price-maximizing distance of 500-1000 feet, and a drop off after that until mandatory busing kicks in whereupon prices rise again to 2500 feet and descend thereafter.

Of course, in a practical regression context, omitting significant variables from a regression will always result in biased results: if, for example, air quality correlates negatively with household income, omitting typical neighbourhood air quality will bias the marginal effect of household income away from zero. Fine, we can include more variables in a hedonic regression: but we can't eliminate the problem entirely, really, since there may always be some variable that is significantly valued by the homebuyer, but which is unknown or difficult to measure. Do we include neighbourhood tree cover, or number of dandelions on lawns? Road and sidewalk quality? The number of stores and restaurants in an area? And how do we quantify this data as it existed fifteen years ago, when the regression's home sales dataset was assembled?

Econometricians have developed alternative techniques for studying school effect on housing prices that can deal, to some extent, with this bias - for example, Black's (1999) boundary fixed effects method, which endeavoured to control for unobserved variables and which thus found a school quality effect approximately half that found in simple least squares regressions. Nevertheless, Bayer, Ferreira, & McMillan (2007) demonstrated that Black's assumption of neighbourhood uniformity along a school boundary would still be invalid if homeowners sort differentially into school boundaries; when controlling for neighbourhood variation on either side of the boundary, they found that the effect of school quality on house prices was cut in half yet again, though still significant.

1.3 The Spatial Method

In any case, natural experiment papers, while a fair alternate method for controlling for unobserved variables, still will leave out the effect of autocorrelation in the dependent variable, if it exists. Certainly failure to include social, accessibility or neighbourhood variables that are determinants of house price can lead to omitted variable bias (though these can be differenced out); but where such unobserved variables are spatially autocorrelated with each other but uncorrelated with observed variables in the characteristics vector X, this violation of OLS assumptions (i.e. the assumption that $E(\epsilon_i|\epsilon_{j\neq i}) = 0)$) will result in underestimation of the standard errors of the coefficients' estimates, still leading to overestimation of coefficient significance and possible unwarranted rejection of the null hypothesis. Where the unobserved variables are correlated with X, this will cause OLS to also generate bias in the estimates. But importantly for spatial analysis, for a data generating process where dependent variable y_i is partially determined by neighbouring dependent variables $y_{j\neq i}$ (e.g. if a house's sale price can be expected to be influenced by recent sales of other houses in the neighbourhood, in practice as the result of a process of price determination by comparison to other recently sold houses (Small & Steimetz, 2012)), then failure to include this spatial dependency in the regression will also lead to biased and inconsistent estimates (Lesage & Pace, 2009).

The use of spatial econometric techniques for real estate can address these problems. Given a vector of house prices P and a matrix of house, accessibility and neighbourhood characteristics X, a spatial lag model (SLM):

$$P = \rho W P + X\beta + u,$$

will take into account each house price P_i's dependence on the price of neighbouring houses $P_{j \neq i}$, where neighbour relationships are included using a weights matrix W (which can be either binary or with weights dependent on a function of the distance separating i and j, with the diagonal set to 0). On the other hand, a spatial error model (SEM):

$$P = X\beta + \epsilon, \epsilon = \lambda W \epsilon + u,$$

will address the problem of unobserved variables that are spatially autocorrelated.

Where a Moran's I test on a non-spatial regression's residuals indicates spatial autocorrelation, Lagrange Multiplier tests presented in Anselin (1988) are a method to determine whether a spatial lag model or spatial error model should be tried. If both of these spatial models still exhibit residual spatial autocorrelation, a "bottom up" approach (Florax, Folmer, & Rey, 2003) extends the spatial model one step higher to a spatial Durbin model (SDM) (Osland, 2010), where

$$P = \rho W P + X\beta - \rho W X \gamma + \epsilon.$$

Despite the look of the structural form above, this spatial Durbin setup does allow both for P_i to be determined by $P_{j \neq i}$, and for prices to also be determined by spatially autocorrelated unobserved variables in ϵ. With $\gamma = -\rho\beta$, this model will simplify to a spatial error model; this "common factor constraint" can also be tested (Anselin, 2003).

Other more complicated spatial regression models have been developed, and there is no single agreed-upon method of specification search; one can use this "bottom-up" method of estimating a nonspatial regression, testing for residual autocorrelation, running LM tests, trying spatial lag and spatial error models, and adding complexity to the model until residual autocorrelation disappears. Alternately, one can instead start with a model that sets a lag on P, X *and* ϵ, and then use Likelihood Ratio tests to reduce the model down to a more parsimonious one. Lesage & Pace (2009) even suggest starting with a Spatial Durbin model, as it has been found to produce coefficient estimates that are accurate even when the underlying spatial process is not necessarily Spatial Durbin in character.

This thesis will follow the "bottom-up" model, for which a real estate application has been detailed elegantly in Osland (2010). The next two chapters will look at a static model, and attempt to see if distance to school is a significant variable in a nonspatial, and then in a spatial, hedonic regression. The final chapter will take a different approach, using a natural experiment to instead look for a temporal effect of school closure on house prices; comparison of control and treatment groups over time allows us both to concentrate on the "educational good" value of the school itself (as opposed to the effect of a school and its surrounding appurtenances, which generally remain after closure and may have their own independent utility) and to factor out bias from possible unaccounted-for geographic effects that change slowly over time.

2 Data Exploration

2.1 Real estate dataset

A real estate sales transaction data set was generously provided by the Realtors' Association of Hamilton and Burlington (RAHB), covering 299,092 transactions from January 1990 through July 2017. Initial investigation found that RAHB agents only began coding houses into age categories by 2005; since age category is an obviously important input into any real estate regression, and since there is also no complete HWDSB school boundary data available from before 2005, all pre-2005 house transactions were dropped from the study's dataset. This first filtering left 159,435 transactions from January 2005 through July 2017.

Dr. Chris Higgins had previously geocoded all transaction locations in the RAHB dataset; approximately 4%-7% of all houses in any given year from the selected timeframe failed geocoding, as can be seen in Figure 2.1. A Wilcoxon test for difference of means indicated that any given year's population of geocode fails is different in mean price from the population of geocode successes, at a significance level of $p < 0.01$, in 6 of the 13 selected years; these test results, along with counts for geocode failures and successes by year and the mean price of each, are summarized in Table 2.1.

It would of course be possible to use a brute-force method to reduce the number of geocode fails, for example by going through the approximately 8800 failed geocode records and using the address field and nearest cross-street field to manually geolocate. However, it was decided to instead remove the geocode fails from the dataset.

After filtering the 2005-2017 dataset to remove geocode fails, non-residential and non-freehold transactions were also stripped, as were non-sale transactions and sale transactions at a price below \$20,000. Homes classified as "APARTMENTUNIT," "MODULAR," "MOBILE," "NA" and so on were also deleted. This left the dataset with 105,411 records for standard ground-level home sales. Next, correction and recoding of individual columns in the dataset was performed.

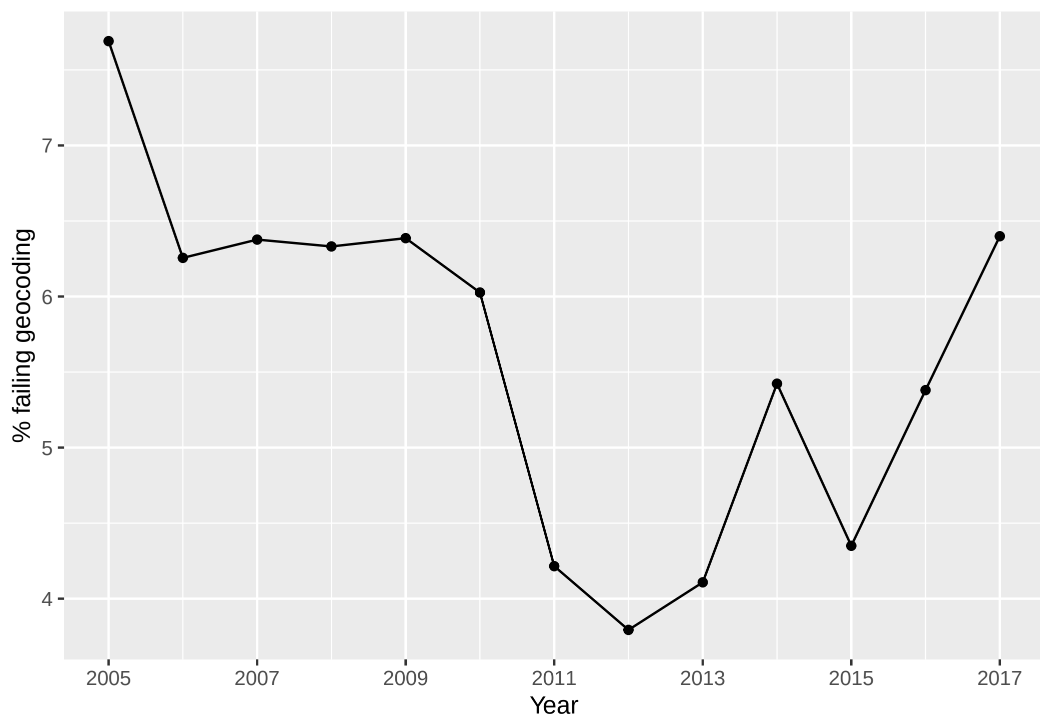

Figure 2.1: Geocode fails by year

Table 2.1: Likelihood of same mean prices for geocode fails, by year

Year	n Fails	Mean Price of Fails	n Succeeds	Mean Price of Succeeds	p-value
2005	943	261001	11319	224362	0.0000
2006	755	256828	11314	239234	0.5158
2007	811	287387	11907	258352	0.0255
2008	739	290569	10933	269245	0.7746
2009	732	295213	10730	267920	0.6918
2010	725	312277	11305	290979	0.0520
2011	504	330022	11453	310139	0.0125
2012	463	378247	11743	338186	0.0000
2013	509	423240	11881	361246	0.0001
2014	719	440108	12539	374088	0.0000
2015	641	479166	14095	410816	0.0000
2016	771	506831	13559	454279	0.0000
2017	534	534088	7811	524738	0.1895

2.1.1 Approximate Age

Age coding in the RAHB dataset uses a categorical variable, with the oldest category corresponding to 51-99 years old. However, approximately 34% of houses in the urban dataset have an "NA" value for house age. Some houses in our dataset might be over 100 years old; there may also be sale transactions where age information was just not available from the seller. However, since the NA proportion for house age ranges from over 40% in 2005 (and over 90% before 2005) down to under 25% by 2017 (see Figure 2.2), it may instead be that some number of the NAs in the dataset are simply due to poor coding.

In any case, it is found that the mean house price for Approximate Age=NA houses is significantly different in every year from the mean price for houses with a recorded age category, confirmed by a Wilcoxon test, as can be seen in Table 2.2; it is apparent, however, that the percentage price difference remains fairly constant after 2005. Because of the different price profile for Age=NA houses and a desire to not skew our results, it was decided to retain the Age=NA houses in the dataset but keep them as a separate category, with the assumption that (at least for 2006 onward) they simply reflect either very old houses or houses for which age information was not available from the seller.

2.1.2 Air Conditioning

The dataset's original air conditioning variable was coded with two levels: NA and CENTRAL. The number of houses coded NA for air conditioning dropped from over 25% in 2005 to around a constant 19% after 2010 (see figure 2.3); if the NA codings only reflected no central air conditioning installed and not incorrect coding, then this would suggest prevalence of air conditioning in sold Hamilton homes increased by 5% in 5 years, which may or may not be implausible. However, since a simple econometric regression will later show this variable to be a significant determinant of house price, it was decided to not omit the air conditioning variable and accept the possible error in the earlier years of the dataset.

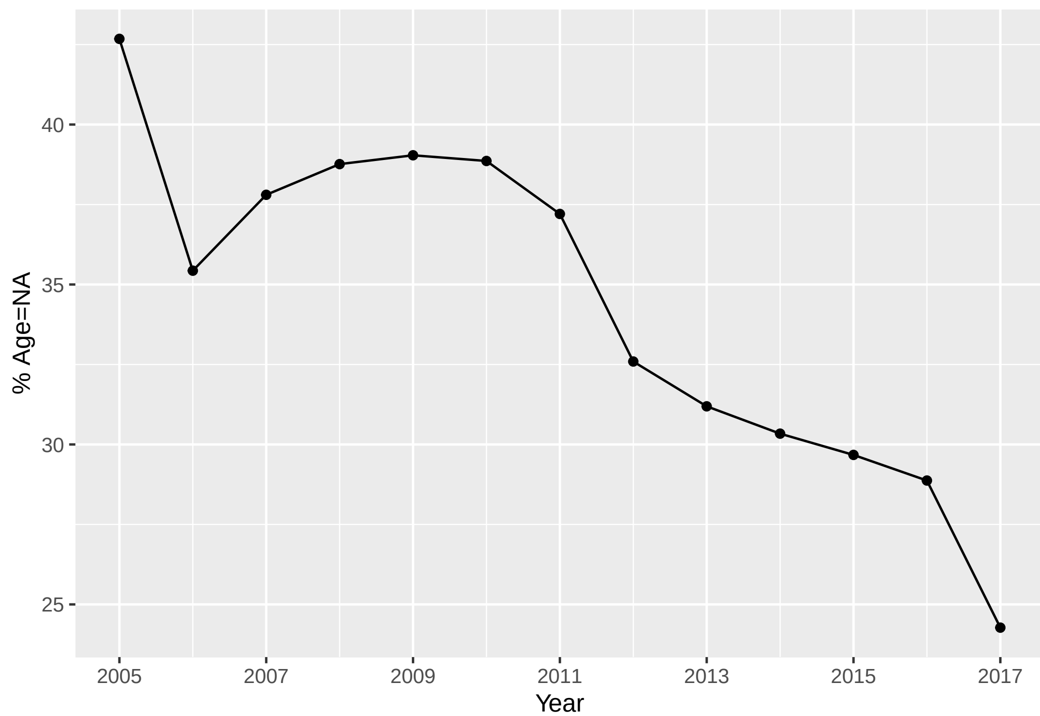

Figure 2.2: Percent of houses coded NA for age

Table 2.2: Likelihood of same mean prices for uncoded house age, by year

Year	n uncoded	Mean Price uncoded	n coded	Mean Price coded	% price diff.	p-value
2005	3509	216544	4713	267073	0.1892	0
2006	2904	198312	5292	296741	0.3317	0
2007	3250	215257	5347	325553	0.3388	0
2008	3003	232773	4744	346816	0.3288	0
2009	2933	236115	4580	348830	0.3231	0
2010	3072	268800	4833	375140	0.2835	0
2011	2956	280409	4989	408347	0.3133	0
2012	2683	285588	5549	443675	0.3563	0
2013	2583	303074	5698	474740	0.3616	0
2014	2579	323424	5922	500714	0.3541	0
2015	2886	355445	6840	542653	0.3450	0
2016	2688	398872	6622	604262	0.3399	0
2017	1271	477049	3965	689778	0.3084	0

2.1.3 Basement

The RAHB dataset contains two categorical variables for basement: Basement.YN (i.e. does a basement exist) and Basement.Finish (i.e., how finished is the basement). Basement.YN is coded with approximately 1% NA values exclusively before 2008, despite being a yes/no variable; this is obviously problematic. Meanwhile, Basement.Finish is meant to have only three levels ("FINISHED," "PARTIAL" and "UNFINISHED"), but also has 6173 records coded "NA"; and these NA records actually increase in number from 3% in 2005 to over 10% by 2017, indicating a possible increase in incorrect coding over time. (See Figure 2.3). Further, while there are 241 records with Basement.YN="NA," 119 of those 241 records have a non-NA value for Basement.Finish, indicating that at least some Basement.YN="NA" records actually have a basement.

A perfunctory ($n = 10$) inspection was performed via Google Maps Street View for a sample of houses that were coded NA for both the above variables. In one case no basement was apparent, and in several cases it was not possible to discern whether the house had a basement; but in another two cases, it was obvious the house was built with a basement (because of visible basement windows, for example). It was decided to follow the following methodology for coding basements:

- if Basement.Finish is not NA, we assume it is properly coded, there is a basement (i.e., no matter what Basement.YN says), and we know how finished it is;
- If Basement.Finish is NA and Basement.YN is 1, we assume a basement exists; and because we are not told any more than that, we can only assume it is unfinished (4.8% of the dataset);
- If Basement.Finish is NA and Basement.YN is 0, we assume correct coding of the latter: thus there is no basement (1% of the dataset);
- If Basement.Finish is NA and Basement.YN is NA, we have no proof that there is a basement, and so assume no basement exists (0.1% of the dataset).

This set of assumptions will induce error into the estimation of coefficients for the value of the level of finish of a basement, but it was decided that it was better to retain the approximately 5% of problematic houses in the dataset due to the relative insignificance of the value of a basement to our study.

2.1.4 Bathrooms

The RAHB dataset contains a numerical variable, Baths.Full, counting the number of full baths in a house; there is a second variable, Baths.Half, for half baths. Baths.Full is obviously miscoded with zeroes for 1%-4% of transactions before 2011, and under 1% of transactions starting in 2011. (See figure 2.3.) It was decided to remove the 1252 transactions recorded with 0 full baths.

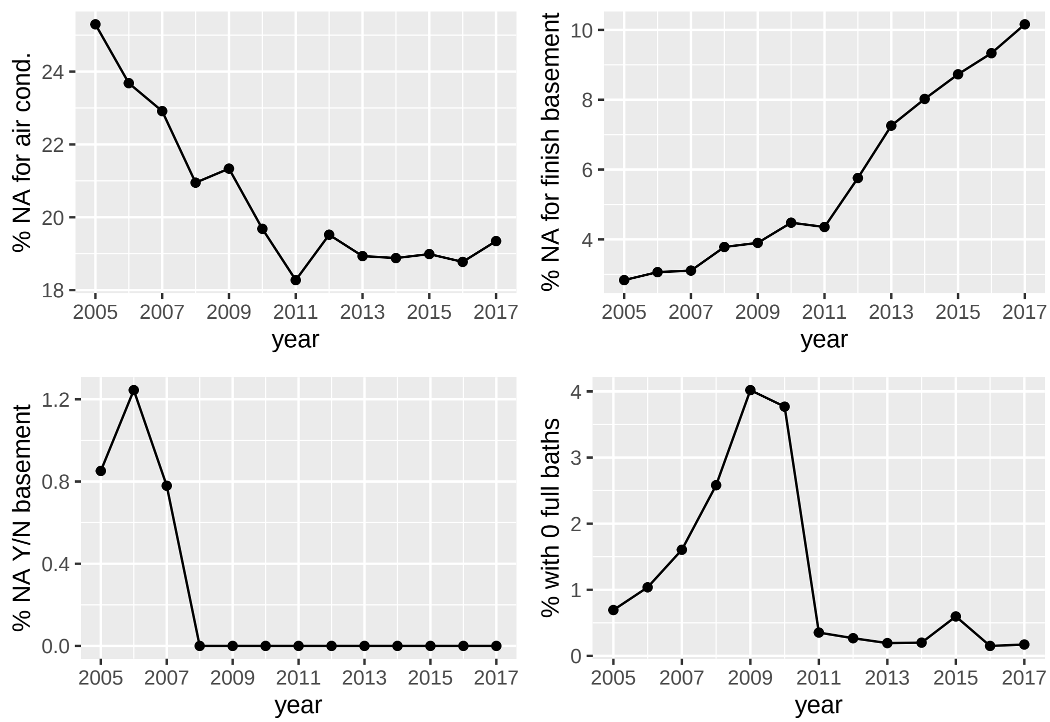

Figure 2.3: Counts for selected variables in RAHB dataset

2.1.5 Garage and Parking

Form factor of car parking for houses in Hamilton varies roughly by age of house and neighbourhood. Typically, on-street or alleyway parking predominates in very old neighbourhoods (pre-1950s); houses from the immediate postwar era tend to have either an unsheltered driveway or a carport, and garages only became standard by the 1980s or later. Unfortunately, the RAHB dataset's coding for its type of car shelter variable has a large number of NA codings, making it a problematic variable to use. As such, it was decided to leave the form factor of car storage to be absorbed by the age of house variable, and instead include number of parking spaces and number of garage spaces (which would better correlate to the size of the house, a variable which we cannot directly observe in our dataset).

For the two variables related to parking, it was assumed that any NA coding for Garage.Spaces.Number represented a 0 (since there were no records where Garage.Spaces.Number was explicitly coded as 0); for Parking.Spaces.Total, there are 22998 records coded as 0, and one single record was found to be recorded as NA. By inspection, the one NA record was determined to be a row house with on-street parking; it was decided to set this as 0 parking spaces.

2.1.6 Heat Type

The RAHB dataset includes a multiple-categorical variable for house heating type, which can include several heating types: a record for one particular house can, for example, indicate both forced-air heating and solar heating.

This category was simplified into a new category that simply states whether forced air heating is present in the house (as opposed to baseboard, radiant, heat pump, solar or "other").

2.1.7 Room Count

The RAHB dataset has two separate variables, Number.Rooms and Room.Count, that disagree for each record; by perfunctory inspection, it seems that Number.Rooms refers to rooms not otherwise counted in the records (i.e. rooms that aren't bathroom or bedroom, either including or not including kitchens or garages or basements). It was decided to use Room.Count, a variable which seems to count the total number of rooms in a house.

The Room.Count variable has a small number of 0s or NAs recorded each year (less than $n = 30$ per year) before 2008. It was decided to delete these records from the dataset, since an NA record here strongly suggests an incorrectly coded record.

2.1.8 Sewer and Water

The RAHB dataset includes categorical variables for presence and type of water supply and sewer type; these two variables obviously correlate very highly with whether a house is located in the contiguous urban area of Hamilton or in its rural surroundings.

As explained later in this chapter, rural houses were ultimately excluded from this study; therefore, these RAHB dataset variables were ignored.

2.1.9 Style and Type of house

Categorical variables exist in the RAHB dataset to identify house style (e.g. number of storeys, sidesplit or backsplit) and house type (detached, townhouse etc). These variables were converted to factors and recoded, as it was found they were statistically significant additions to regressions later on; four NA codings for style, which were found in the 2005 dataset only, were removed as lazy codings.

2.1.10 Time of sale

Variables in the RAHB dataset identify the month, year, and day of sale. It was decided to retain only the year and month variables; a quarter of sale variable was also generated from the month variable, to give another option for dividing periods by time; the quarter of sale variable ultimately was used.

2.1.11 City

This study concentrates on changes to the Hamilton-Wentworth District School Board's school catchments. Therefore, RAHB transactions in the cities of Burlington and Oakville were removed from our dataset. This reduced our dataset to 73507 observations for the city of Hamilton from January 2005 to July 2017.

2.1.12 Lot sizes

Lot sizes for each property were obtained by first converting the house transaction dataset to a simple features GIS dataset, and then intersecting it with a City of Hamilton assessment datafile. Three properties with NA for lot area, and two properties with lot area below 20 sq.m. which were determined by inspection to have been miscoded, were then removed from the dataset.

2.1.13 Log transformations of house variables

In real estate regressions, a log transformation of closing price is typically generated for each transaction in the dataset; this takes into account that house prices are typically roughly log distributed, and allows estimation of the percent change in price caused by a marginal increase in an independent variable. A log transformation of prices was added to our dataset to follow the standard.

Note that a dataset of house prices is only *approximately* log distributed. This real estate dataset's log transformation of closing price still fails a Shapiro-Wilk normality test; interestingly, non-normality of the price distribution increases as the time window for price selection gets wider, suggesting that part of the problem is the steady drift in prices over time which could be expected to make a normal distribution less normal.

Lot size, in our dataset, is also not normally distributed; however, even a log transformation of lot area will still yield a distribution with high excess kurtosis and a long right tail, and this holds to an only slightly lesser extent when the analysis excludes rural areas outside the old City of Hamilton. Nevertheless, a log transformation of lot area was generated for the dataset.

2.1.14 RAHB Dataset Exploration

For each year from 2005-2017, the median sale price of houses was calculated; this generated a simplistic median price index for Metro Hamilton houses. As can be seen in Figure 2.4, year-over-year price accumulation was steady and moderate (between 5% and 7.5%) from 2005-2013, except for 2008-2009 where the effect of the US mortgage-backed securities crash and financial panic was to drive yearly price increases down to below 3% for just two years. It can also be seen that house prices began to experience aggressive year-over-year increases starting in 2015; this price accumulation has continued to the present (2021) year.

Interestingly, when breaking up the price index into a stratified index for former cities of Hamilton (Figure 2.5), the weakness in house prices presaging the US MBS crisis can be seen already by 2007 in the peripheral areas of Flamborough and Glanbrook; the weakness doesn't hit the urban areas of Hamilton and Stoney Creek until 2009. Over the recent period of aggressive price increases, Hamilton and Stoney Creek house prices have appreciated by somewhat more than prices for outlying areas. Both the shorter period of price weakness 2007-2009 and the faster house appreciation after 2015 for urban Hamilton and Stoney Creek may indicate that buyers have a lasting preference for urban houses over exurban houses.

Note that a simple median house price index confounds change in price over time with change in quality (Hill, 2013); a hedonic price index, which compensates for change in quality over time, will be calculated later.

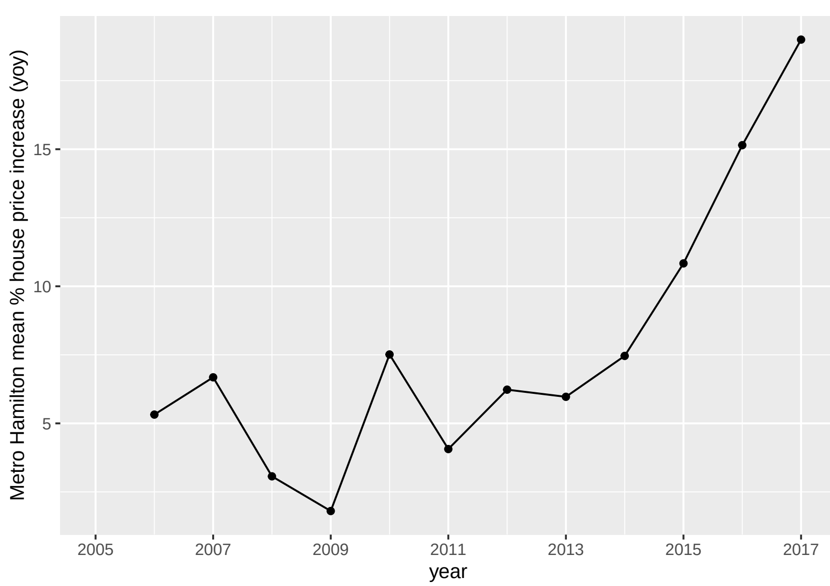

Figure 2.4: Simple median price index for metro Hamilton, 2005-2017

Spatially, house prices in the RAHB dataset generally reflect the trend of lowest prices in the downtown and next-lowest prices in the postwar-to-1980 area of

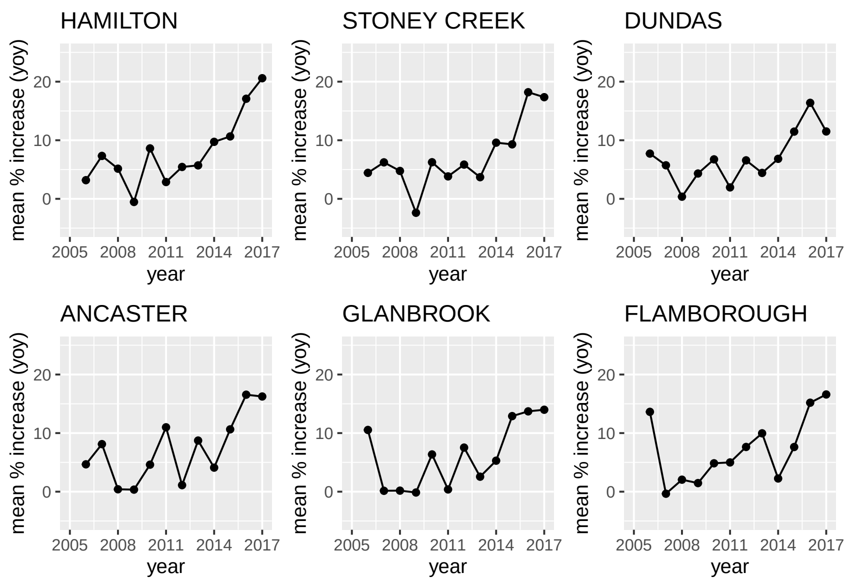

Figure 2.5: Simple median price index for Hamilton, 2005-2017, stratified by former municipality

the north Mountain, where houses are generally older and smaller; higher prices are found further out in the more remote or recently-built neighbourhoods, with the notable exception of the expensive houses in Westdale, along Aberdeen, and along the top of the Mountain Brow, as can be seen in Figures 2.6 (2006) and 2.7 (2016).

2.1.15 Initial regressions and corrected price trend graph

The results for an initial multivariate regression (on house characteristics only), performed on the complete Jan 2005-Jul 2017 dataset, are presented below in Table 2.3; a by-year time dummy is omitted from the summary. Generally, coefficients make sense: houses are penalized for age, there is a significant benefit to having central air conditioning, basements are valued for their level of finish, extra bathrooms and bedrooms (and rooms in general) are positively-valued goods, detached houses are valued more, parking spaces and garage spaces are normal goods, and lot area is also positively valued. Strangely, forced air heating seems to be penalized with a 10% reduction in house price compared to radiator, baseboard and other forms of heating.

An adjusted R^2 of 0.76 is rather impressive for a house price regression which

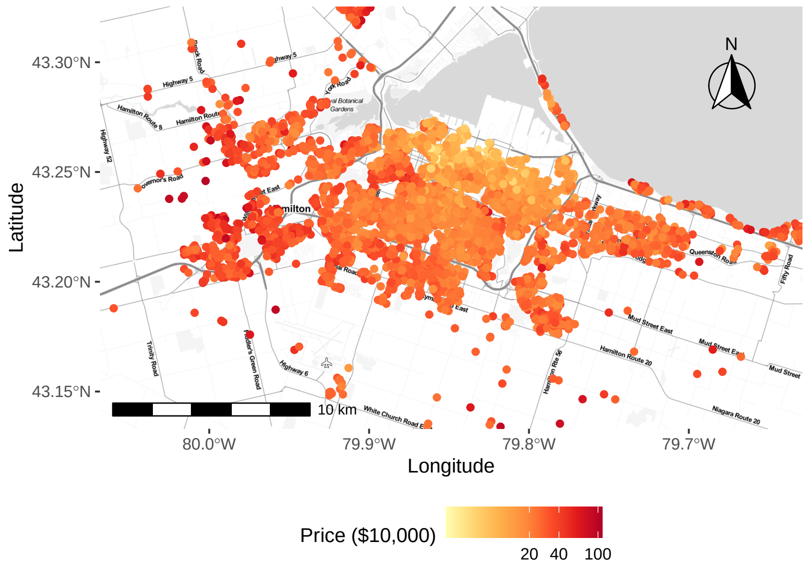

Figure 2.6: Spatial distribution of house prices in Hamilton, 2006

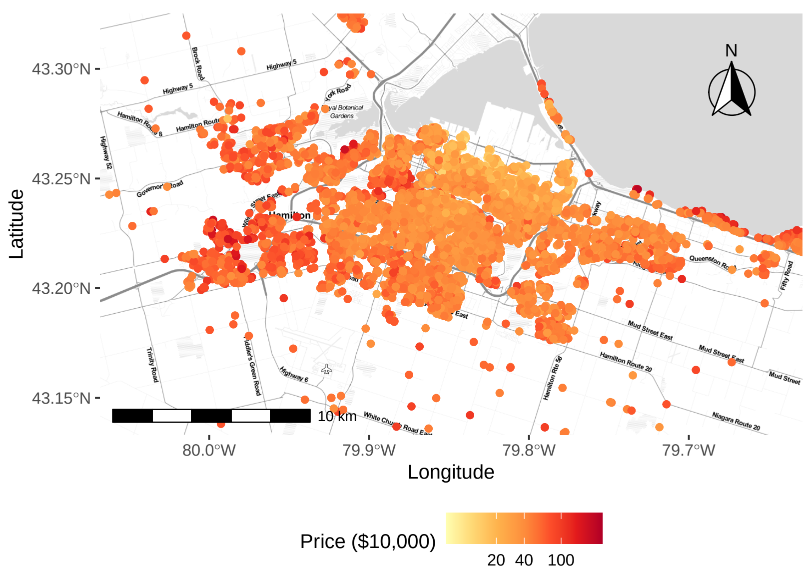

Figure 2.7: Spatial distribution of house prices in Hamilton, 2016

forces homogeneity of coefficients over 13 years, and which doesn't even include house floor area or any sort of neighbourhood characteristics.

This initial multivariate regression can be used along with a time dummy to construct a hedonic price index; this provides a correction for any possible change in the distribution of house characteristics over time that would affect apparent house price appreciation. (Hill, 2013) A corrected price trend graph for Metro Hamilton, using time dummy coefficient values from the regression, is shown in figure 2.8: as can be seen, when we take into account change in characteristics over time, the weakness in house prices due to the 2009 MBS crisis actually sent the metro Hamilton hedonic index negative for 2009. Meanwhile, the price trend post-2015 looks slightly less aggressive, after accounting for changes in house characteristics over time.

A predicted-observed plot (see Figure 2.9) shows the regression is fairly predictive of prices, at least through the middle of the range of observed prices; The deflected lower left tail, however, indicates that the predictive power of the regression degrades as we examine houses in the lower range of observed prices.

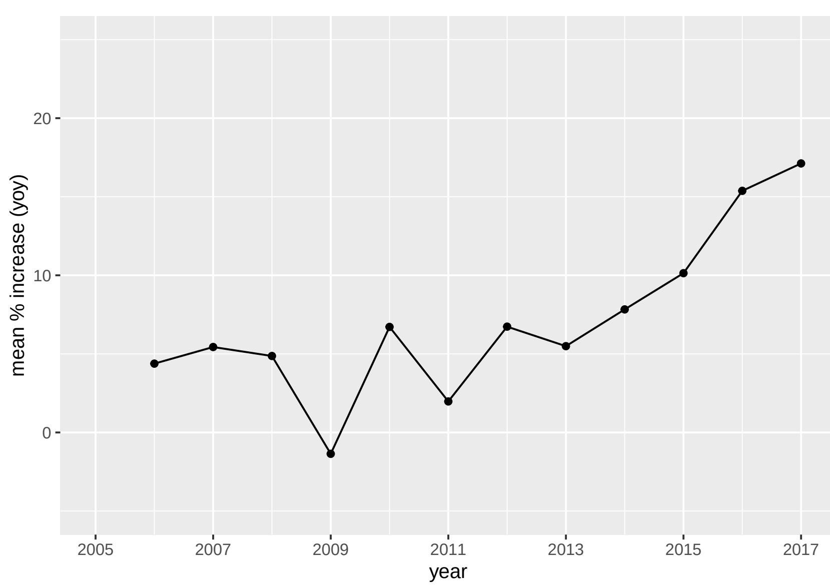

Figure 2.8: Hedonic price index for Hamilton, 2005-2017

Of course, the above combined-years regression forces constants to remain steady over years. The assumption of temporal homogeneity can be relaxed

Table 2.3: Initial non-spatial regression results, 2005-2017 dataset

Variable	results			
term	β	s.e.	statistic	p
(Intercept)	10.706	0.017	630.371	***
Approximate.Age0TO5	-0.098	0.008	-12.000	***
Approximate.Age6TO15	-0.220	0.008	-28.092	***
Approximate.Age16TO30	-0.327	0.008	-39.052	***
Approximate.Age31TO50	-0.355	0.008	-42.489	***
Approximate.Age51TO99	-0.335	0.008	-41.645	***
Approximate.AgeUnknown	-0.428	0.008	-54.696	***
AC.YN-NoCentralAir	-0.164	0.002	-66.322	***
Basement-Unfinished	0.058	0.010	5.568	***
Basement-Partial	0.084	0.011	7.910	***
Basement-Finished	0.101	0.011	9.515	***
Baths.Full	0.101	0.002	53.719	***
Baths.Half	0.116	0.002	50.846	***
Beds.Total	0.029	0.001	23.787	***
Garage.Spaces.Number	0.049	0.002	25.238	***
Heat.FA-ForcedAir	-0.097	0.005	-20.629	***
HouseType-Semi	-0.069	0.004	-16.555	***
HouseType-Town	-0.094	0.004	-21.262	***
lnlotareasqm	0.212	0.001	143.895	***
Parking.Spaces.Total	0.019	0.001	17.896	***
as.factor(YEAR)2006	0.044	0.005	9.502	***
Room.Count	0.017	0.001	31.950	***
Storeys-RaisedBungalow	0.035	0.005	6.324	***
Storeys-Sidesplit	0.074	0.007	11.300	***
Storeys-Backsplit	0.056	0.004	12.957	***
Storeys-1.5	-0.018	0.003	-5.726	***
Storeys-2	0.015	0.003	4.713	***
Storeys-2.5	0.011	0.004	2.597	**
Storeys-3	0.090	0.013	6.892	***
Storeys-Other	0.084	0.021	3.966	***

Note:

R-squared (1) = 0.762

*** $p < 0.001$

** $0.001 < p < 0.01$

* $0.01 < p < 0.05$

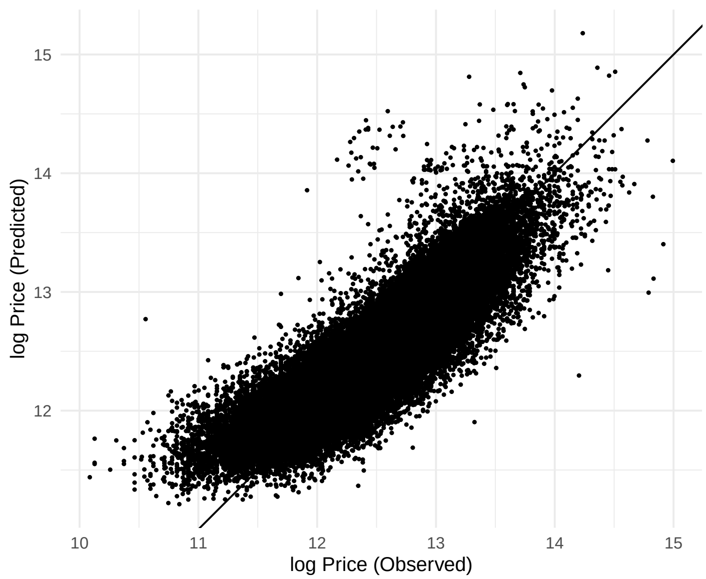

Figure 2.9: Predicted-Observed plot for multi-year real estate regression, house data only

by doing year-by-year regression, and examining the stability of estimates; Figure 2.10 compares the coefficient estimates (plotted along with a graphical representation of their standard error) for yearly regressions from 2005 to 2009: it can be seen that, at least for this time interval, yearly coefficients for house characteristics were fairly stable over time, with their estimate distributions overlapping with the distributions of other years' estimates.

2.2 Census variables

Census data at the Dissemination Area (DA) level for the city of Hamilton was downloaded from Statistics Canada for 2006, 2011, and 2016; it was then intersected with the home sales data sets for those three years. Census variables were chosen on the basis of their possible impact on house prices; these variables also correlate with school scores, which we will later interact with school distance (Johnson, 2005). Using DAs for the level of aggregation provides an average of 6.2 (in 2011) to 7.6 (in 2016) house transactions per dissemination area.

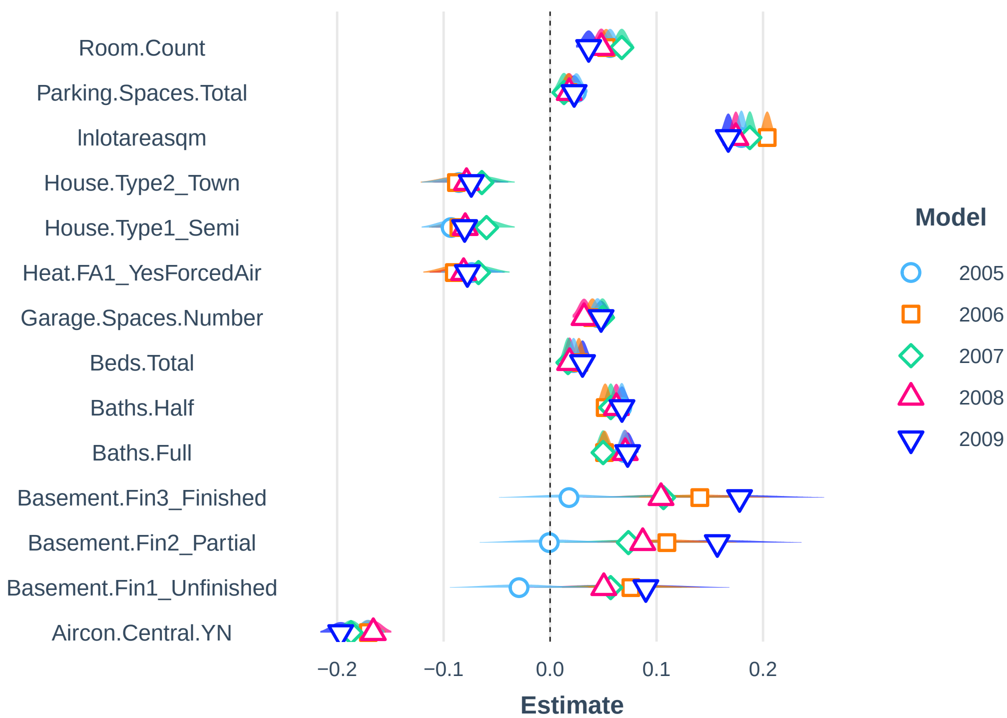

Figure 2.10: Comparison of regression coefficients, 2005-2009

2.2.1 Corrections for missing data

In the 2006 DA-level census dataset, there are seven DAs with NA for some variables: of these, three are DAs with no private housing (St. Joseph's Villa in Dundas, the St. Joseph's Healthcare facility at West Fifth and Fennell, and a strip of land by Centennial Parkway and Queenston Road that has no dwellings). These three DAs were left uncorrected, since they will have no housing transactions. For the remaining four DAs in 2006 with missing data, missing variables were replaced with those variables which were available at the DA level from the 2011 dataset. An eighth problematic DA was found at the St. Elizabeth Village gated neighbourhood at Garth and Rymal, where raw values of 0 yield 0/0 errors for "percent 25-64 with no highschool" and "percent journey to work by active transit." Given the demographic nature of this DA, a 0 value was set for each of these variables.

Six more DAs in the 2006 dataset were found to have 0 set for average after-tax household income; they had their incomes approximated by taking the average income of their neighbouring DAs.

In the 2011 dataset, there are 19 records with NA for at least one value. Three of these are again the non-residential DAs that were found in 2006 (a further problem DA is the St. Elizabeth Village seniors gated residential area, which is again missing a number for active transit and for percent 25-64 without highschool). For the remaining 16 DAs where NAs are found for Average After-Tax Household Income, those which had an AVGATHHINC recorded in 2006 and 2016 had their 2011 value calculated by the log median, $\sqrt{x_{2006}x_{2016}}$. Two DAs with no 2006 and/or 2016 income data instead had their incomes approximated by taking the average income of all their neighbouring DAs. Finally, in the 2016 dataset, twelve records again were missing values for average after-tax household income; these were approximated by again taking the average income of neighbouring DAs.

Total DAs requiring generation of at least one synthetic variable make up less than 2.5% of each year's DA dataset.

2.2.2 Spatial distribution of census variables

Census variables, like house prices, are also found to have a definite spatial distribution. Figure 2.11 and Figure 2.12, for example, show 2016 average after-tax household income is generally lower, and 2016 unemployment rate is generally higher, in the downtown core; incomes are also generally higher in the west part of the urban metropolitan area.

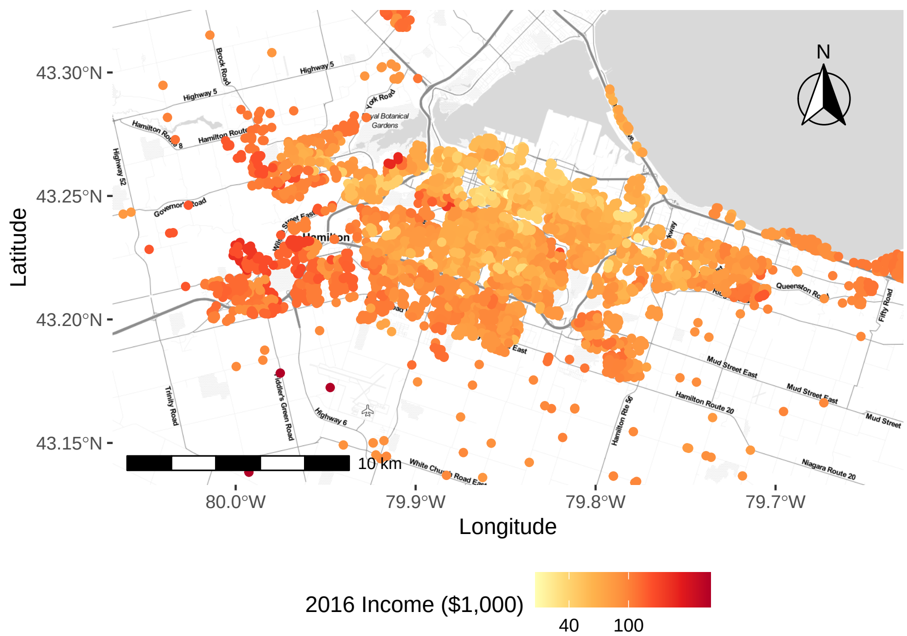

Figure 2.11: Spatial distribution of average after-tax household incomes in Hamilton, 2016

These spatial distributions are similar for other census variables denoting economic stress, such as percent of homes needing major repair (Figure 2.13) or percent of persons aged 25-64 who have not completed highschool (Figure 2.14).

As seen in the spatial distribution plots, there is high cross-correlation of census variables; this is confirmed with Pearson correlation plots. In particular, it is most obvious in Figures 2.15, 2.16 and 2.17 that average after-tax household income is most negatively correlated with percent journey to work using active transit, percent living in apartments, and percent age 25 to 64 who have not completed highschool; most likely, people with very low incomes can't afford

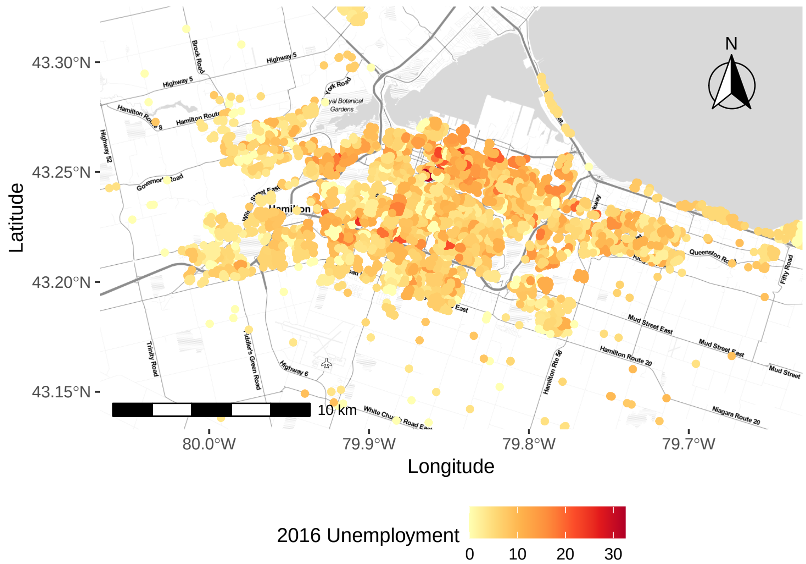

Figure 2.12: Spatial distribution of unemployment rate in Hamilton, 2016

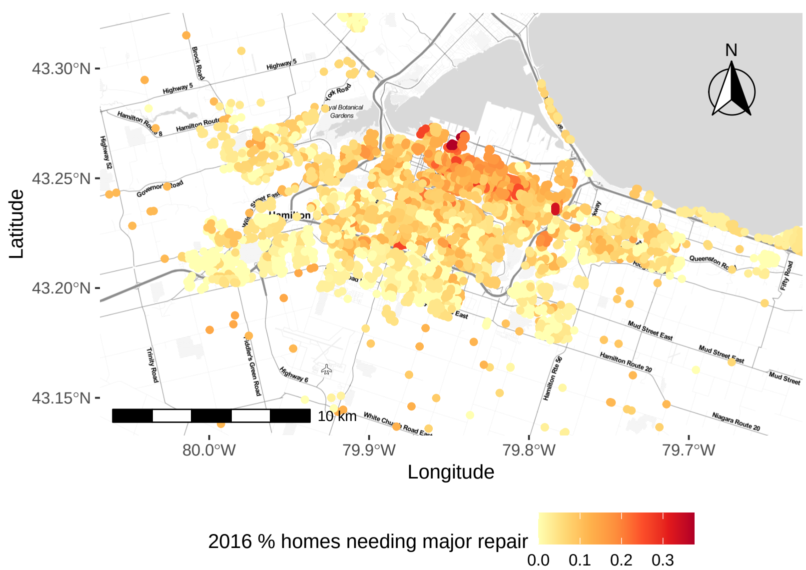

Figure 2.13: Spatial distribution of percent homes needing major repair in Hamilton, 2016

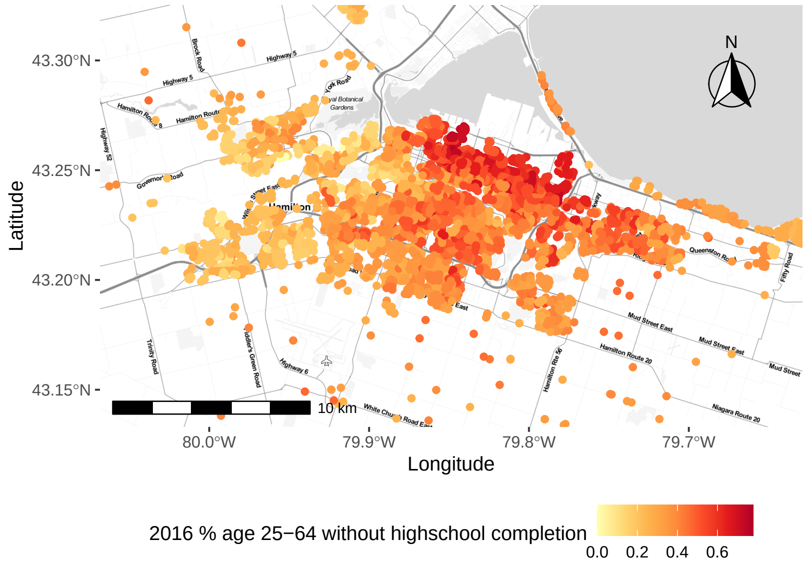

Figure 2.14: Spatial distribution of persons 25-64 not completed highschool in Hamilton, 2016

personal transit, can't save enough money to buy or rent a house, and are constrained by lack of education from gaining higher-paying jobs. Household income is also negatively correlated with percent of residents who have moved in the past year; this variable was selected to be a measure of economic stress, and it seems to do this job well in these datasets.

Percent ages 65 and up, interestingly, has very little correlation with the other census variables, other than a strong negative correlation with percent aged 0 to 14 - perhaps expected if the typical neighbourhood in Hamilton is very homogeneous in resident age and family makeup. Percent of DA aged 0 to 14 is also rather negatively correlated with percent of neighbourhood living in apartments.

2.2.3 Regressions using census data

The corrected census data was intersected with the RAHB dataset, producing three smaller datasets for the years 2006, 2011 and 2016 with each house's DA-level census statistics included in its record. New hedonic regressions were performed for each year, one with both house and census data and one with house data only; the results are presented in Table 2.4.

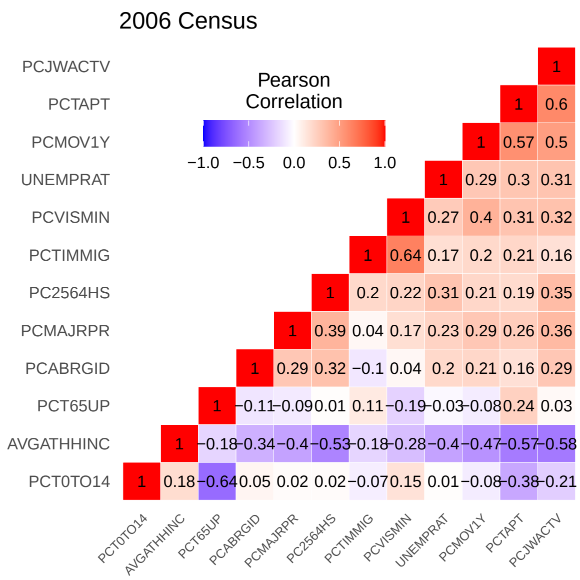

Figure 2.15: Pearson correlation matrix for 2006 census

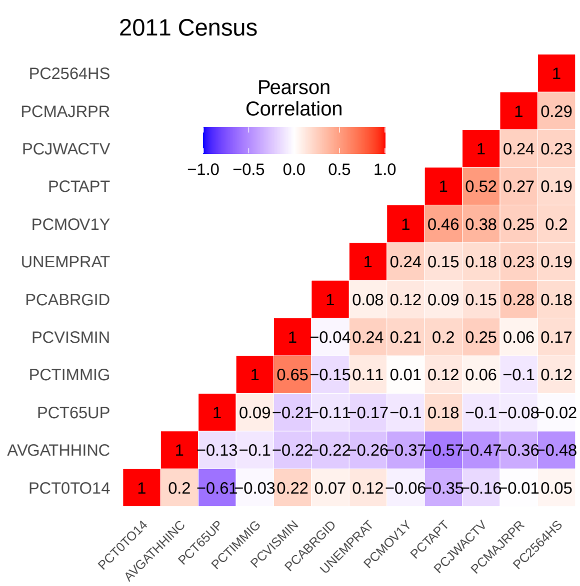

Figure 2.16: Pearson correlation matrix for 2011 census

Table 2.4: Non-spatial regression results with census

Variable	2006			2006 with census			2011			2011 with census			2016			2016 with census		
term	β	s.e.	p	β	s.e.	p	β	s.e.	p	β	s.e.	p	β	s.e.	p	β	s.e.	p
(Intercept)	10.4478	0.0571	***	10.8358	0.0529	***	10.8853	0.0751	***	11.0116	0.0674	***	11.4786	0.0477	***	11.9482	0.0462	***
Approximate.Age0TO5	-0.1698	0.0316	***	-0.1867	0.0258	***	-0.1839	0.0267	***	-0.1573	0.0222	***	-0.0200	0.0235		-0.0527	0.0191	**
Approximate.Age6TO15	-0.3039	0.0303	***	-0.2993	0.0248	***	-0.2992	0.0249	***	-0.2645	0.0207	***	-0.1463	0.0223	***	-0.1750	0.0181	***
Approximate.Age16TO30	-0.4489	0.0329	***	-0.4258	0.0269	***	-0.3913	0.0273	***	-0.3580	0.0229	***	-0.2467	0.0236	***	-0.2559	0.0193	***
Approximate.Age31TO50	-0.4663	0.0319	***	-0.4397	0.0261	***	-0.4264	0.0281	***	-0.3503	0.0237	***	-0.2800	0.0237	***	-0.2544	0.0195	***
Approximate.Age51TO99	-0.4629	0.0310	***	-0.4327	0.0255	***	-0.3976	0.0261	***	-0.3445	0.0222	***	-0.2755	0.0228	***	-0.2532	0.0189	***
Approximate.AgeUnknown	-0.5558	0.0304	***	-0.4859	0.0249	***	-0.4986	0.0249	***	-0.3967	0.0211	***	-0.3451	0.0225	***	-0.2834	0.0186	***
AC.YN-NoCentralAir	-0.1687	0.0083	***	-0.1283	0.0068	***	-0.2131	0.0100	***	-0.1469	0.0085	***	-0.1146	0.0074	***	-0.0786	0.0061	***
Basement-Unfinished	0.0826	0.0330	*	0.0874	0.0269	**	0.1183	0.0576	*	0.1321	0.0478	**	0.0388	0.0277		0.0614	0.0225	**
Basement-Partial	0.1156	0.0332	***	0.0992	0.0270	***	0.1420	0.0579	*	0.1510	0.0481	**	0.0424	0.0283		0.0653	0.0229	**
Basement-Finished	0.1464	0.0333	***	0.1342	0.0271	***	0.1563	0.0579	**	0.1737	0.0481	***	0.0608	0.0282	*	0.0777	0.0229	***
Baths.Full	0.0754	0.0068	***	0.0606	0.0056	***	0.1051	0.0074	***	0.0848	0.0062	***	0.1014	0.0054	***	0.0790	0.0044	***
Baths.Half	0.0967	0.0081	***	0.0603	0.0066	***	0.1214	0.0087	***	0.0789	0.0072	***	0.0930	0.0067	***	0.0644	0.0054	***
Beds.Total	0.0273	0.0042	***	0.0134	0.0035	***	0.0409	0.0048	***	0.0278	0.0040	***	0.0289	0.0034	***	0.0200	0.0028	***
Garage.Spaces.Number	0.0506	0.0068	***	0.0370	0.0056	***	0.0453	0.0076	***	0.0375	0.0064	***	0.0434	0.0056	***	0.0344	0.0045	***
Heat.FA-ForcedAir	-0.0890	0.0148	***	-0.0577	0.0121	***	-0.1269	0.0189	***	-0.0672	0.0157	***	-0.0747	0.0153	***	-0.0454	0.0124	***
HouseType-Semi	-0.0849	0.0143	***	-0.1012	0.0117	***	-0.0734	0.0167	***	-0.0829	0.0139	***	-0.0961	0.0123	***	-0.1100	0.0100	***
HouseType-Town	-0.0899	0.0171	***	-0.1269	0.0140	***	-0.0527	0.0170	**	-0.1103	0.0142	***	-0.1204	0.0124	***	-0.1749	0.0101	***
lnlotareasqm	0.2765	0.0054	***	0.2067	0.0048	***	0.2128	0.0057	***	0.1574	0.0050	***	0.1686	0.0043	***	0.1167	0.0038	***
Parking.Spaces.Total	0.0109	0.0036	**	0.0097	0.0029	***	0.0269	0.0041	***	0.0165	0.0034	***	0.0183	0.0030	***	0.0095	0.0025	***
Quarter2	0.0280	0.0094	**	0.0290	0.0077	***	0.0500	0.0110	***	0.0414	0.0091	***	0.0913	0.0086	***	0.0829	0.0070	***
Quarter3	0.0592	0.0094	***	0.0595	0.0076	***	0.0667	0.0109	***	0.0610	0.0091	***	0.1364	0.0082	***	0.1252	0.0067	***
Quarter4	0.0380	0.0100	***	0.0491	0.0081	***	0.0464	0.0114	***	0.0385	0.0095	***	0.1667	0.0087	***	0.1641	0.0071	***
Room.Count	0.0182	0.0020	***	0.0169	0.0016	***	0.0153	0.0021	***	0.0129	0.0017	***	0.0165	0.0014	***	0.0130	0.0012	***
Storeys-RaisedBungalow	0.0296	0.0181		0.0273	0.0148		0.0116	0.0215		-0.0040	0.0179		0.0378	0.0165	*	0.0298	0.0134	*
Storeys-Sidesplit	0.0309	0.0229		0.0025	0.0186		0.0598	0.0249	*	0.0115	0.0207		0.0765	0.0192	***	0.0320	0.0156	*
Storeys-Backsplit	0.0631	0.0145	***	0.0637	0.0120	***	0.0530	0.0168	**	0.0562	0.0141	***	0.0428	0.0133	**	0.0406	0.0110	***
Storeys-1.5	0.0002	0.0107		0.0105	0.0088		-0.0086	0.0126		0.0006	0.0106		-0.0289	0.0095	**	-0.0041	0.0077	
Storeys-2	0.0154	0.0107		0.0247	0.0088	**	0.0008	0.0123		0.0158	0.0103		0.0244	0.0089	**	0.0360	0.0073	***
Storeys-2.5	0.0164	0.0145		0.0539	0.0122	***	0.0032	0.0170		0.0504	0.0146	***	0.0530	0.0126	***	0.1009	0.0106	***
Storeys-3	0.2320	0.0780	**	0.1611	0.0635	*	0.0645	0.0755		0.0777	0.0627		0.1046	0.0293	***	0.1264	0.0237	***
Storeys-Other	0.1841	0.0632	**	0.1350	0.0514	**	0.0272	0.1088		0.0753	0.0903		0.0068	0.0623		0.0037	0.0504	
PCJWACTV				0.2381	0.0315	***				0.1248	0.0303	***				0.0717	0.0307	*
PCMOV1Y				0.0378	0.0413					0.2133	0.0388	***				0.0805	0.0407	*
UNEMPRAT				-0.0038	0.0007	***				-0.0021	0.0004	***				-0.0013	0.0006	*
PCTIMMIG				0.1084	0.0356	**				-0.0299	0.0378					0.0933	0.0386	*
PCT0TO14				-0.2572	0.0734	***				-0.4208	0.0888	***				-0.2827	0.0668	***
PCVISMIN				-0.1560	0.0369	***				0.0905	0.0363	*				-0.0861	0.0325	**
PCT65UP				0.1569	0.0482	**				0.0946	0.0548					0.0490	0.0443	
PCTAPT				0.1315	0.0177	***				0.1950	0.0200	***				0.2169	0.0149	***
PCABRGID				-0.7174	0.0955	***				-0.3169	0.0797	***				-0.8033	0.0916	***
PCMAJRPR				-0.4102	0.0485	***				-0.3491	0.0395	***				-0.3007	0.0508	***
AVGATHHINC				0.0037	0.0002	***				0.0047	0.0002	***				0.0027	0.0002	***
PC2564HS				-0.8432	0.0306	***				-0.7171	0.0305	***				-0.6535	0.0234	***

Note:
R-squared (1) = 0.738
R-squared (2) = 0.827
R-squared (3) = 0.682
R-squared (4) = 0.782
R-squared (5) = 0.691
R-squared (6) = 0.798
*** $p < 0.001$
** $0.001 < p < 0.01$
* $0.01 < p < 0.05$

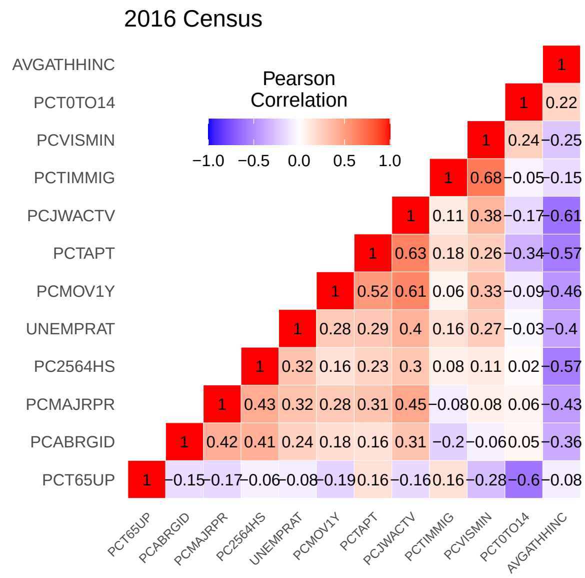

Figure 2.17: Pearson correlation matrix for 2016 census

As can be seen, the coefficients for the baskets of house-related "goods" generally drop slightly when census variables are included in a regression. To some extent, this may be due to correlations between social variables and house type: smaller houses, with less features, will be found in neighbourhoods with lower average incomes.

While house variable coefficients were found to drift only slightly in the year-by-year regression set from 2005 to 2009 (see Figure 2.10), over the inter-census intervals these variables do show a much greater capacity for drift; coefficient estimates for 'Percent Visible Minority' and 'Percent Immigrants' are found to be especially unstable, even dropping out of significance in 2011.

A log transformation for average after-tax household income was investigated; it would make sense that percent marginal effect on house price would be determined by percent change in income, and not by absolute change in income. Tukey tests on a random sample of 5000 rows from each of the 2006, 2011 and 2016 datasets indicated the appropriateness of a lambda near zero for income in each of the three years' datasets (i.e. most likely a log transformation); a log transformation of income was found to produce a distribution with near-zero skewness and slightly negative excess kurtosis (see Table 2.5), and using the log

Table 2.5: Effect of log transform on income

	2006	2011	2016
skewness of log income	0.134	-0.116	0.063
kurtosis of log income	2.816	2.851	2.932
skewness of untransformed regression's residuals	-0.311	-0.332	-0.126
skewness of transformed regression's residuals	-0.315	-0.321	-0.057
kurtosis of untransformed regression's residuals	5.235	5.404	6.955
kurtosis of transformed regression's residuals	5.286	5.396	7.028
R-squared for untransformed regression	0.827	0.782	0.798
R-squared for transformed regression	0.832	0.785	0.801
F test for untransformed regression	653.046	452.283	618.226
F test for transformed regression	674.103	461.016	628.545

transformation in each regression slightly improves the R^2 and F-test without negatively impacting the skewness or kurtosis of the residuals. This suggests that a log transformation of income will be appropriate.

Re-running the regressions from Table 2.4 with log transformations of income produces the results in Table 2.6. (The house structure type variable has been removed from this table for simplicity.) Note that after incorporating log transformation for after-tax household income, some variables are found to be insignificant in some regressions: PCTIMMIG is insignificant in 2006 and 2011, PCTMOV1Y in 2006, and PCT65UP and UNEMPRAT in 2016. However, since these insignificance results are not constant across all years, it was decided to keep all variables that are significant in at least one year, to ensure homogeneity of regression equations across years.

An example predicted-observed plot for the 2006 regression, including census and log-transformed household income, is shown in Figure 2.18. Visual inspection of this Figure suggests that this regression is best at predicting prices for houses which sold in the range of approx. 100,000-440,000 dollars in 2006; this range is approximately 88.6% of the dataset. Underprediction of prices at the low end of the price range seems lesser than it was in the multi-year non-census regression in Figure 2.9.

Table 2.6: Non-spatial regression results with census, log transformed income

Variable	2006			2011			2016		
term	β	s.e.	p	β	s.e.	p	β	s.e.	p
(Intercept)	7.3808	0.1722	***	6.8914	0.1799	***	8.6047	0.1832	***
Approximate.Age0TO5	-0.1860	0.0255	***	-0.1522	0.0221	***	-0.0516	0.0190	**
Approximate.Age6TO15	-0.2940	0.0244	***	-0.2528	0.0206	***	-0.1733	0.0180	***
Approximate.Age16TO30	-0.4188	0.0266	***	-0.3403	0.0228	***	-0.2544	0.0191	***
Approximate.Age31TO50	-0.4326	0.0258	***	-0.3348	0.0235	***	-0.2521	0.0193	***
Approximate.Age51TO99	-0.4260	0.0251	***	-0.3294	0.0220	***	-0.2503	0.0188	***
Approximate.AgeUnknown	-0.4776	0.0246	***	-0.3794	0.0210	***	-0.2805	0.0185	***
AC.YN-NoCentralAir	-0.1259	0.0067	***	-0.1429	0.0084	***	-0.0757	0.0060	***
Basement-Unfinished	0.0807	0.0265	**	0.1378	0.0475	**	0.0645	0.0223	**
Basement-Partial	0.0912	0.0267	***	0.1534	0.0477	**	0.0659	0.0228	**
Basement-Finished	0.1274	0.0267	***	0.1751	0.0477	***	0.0789	0.0227	***
Baths.Full	0.0601	0.0055	***	0.0853	0.0061	***	0.0788	0.0044	***
Baths.Half	0.0574	0.0065	***	0.0786	0.0072	***	0.0637	0.0054	***
Beds.Total	0.0129	0.0034	***	0.0281	0.0040	***	0.0203	0.0028	***
Garage.Spaces.Number	0.0368	0.0055	***	0.0371	0.0063	***	0.0341	0.0045	***
Heat.FA-ForcedAir	-0.0552	0.0119	***	-0.0649	0.0156	***	-0.0398	0.0124	**
HouseType-Semi	-0.0965	0.0116	***	-0.0844	0.0138	***	-0.1097	0.0100	***
HouseType-Town	-0.1286	0.0138	***	-0.1131	0.0141	***	-0.1757	0.0101	***
lnlotareasqm	0.2025	0.0047	***	0.1549	0.0050	***	0.1149	0.0038	***
Parking.Spaces.Total	0.0097	0.0029	***	0.0169	0.0034	***	0.0094	0.0024	***
Quarter2	0.0286	0.0076	***	0.0402	0.0091	***	0.0829	0.0069	***
Quarter3	0.0588	0.0075	***	0.0599	0.0090	***	0.1260	0.0066	***
Quarter4	0.0485	0.0080	***	0.0384	0.0094	***	0.1638	0.0070	***
Room.Count	0.0170	0.0016	***	0.0131	0.0017	***	0.0128	0.0012	***
PCT0TO14	-0.2489	0.0724	***	-0.4414	0.0882	***	-0.2837	0.0664	***
PCT65UP	0.2123	0.0477	***	0.1108	0.0544	*	0.0711	0.0440	
PCTAPT	0.1972	0.0181	***	0.2715	0.0208	***	0.2834	0.0158	***
PCTIMMIG	0.0688	0.0352		-0.0677	0.0375		0.0829	0.0383	*
PCABRGID	-0.5670	0.0950	***	-0.2771	0.0792	***	-0.7015	0.0914	***
PCVISMIN	-0.1104	0.0366	**	0.1233	0.0361	***	-0.0818	0.0323	*
PCMAJRPR	-0.3645	0.0480	***	-0.3074	0.0393	***	-0.2524	0.0506	***
PC2564HS	-0.7564	0.0310	***	-0.6462	0.0311	***	-0.5723	0.0246	***
PCJWACTV	0.2857	0.0313	***	0.1363	0.0301	***	0.1446	0.0310	***
PCMOV1Y	0.0577	0.0408		0.2358	0.0386	***	0.1009	0.0405	*
log(AVGATHHINC * 1000)	0.3336	0.0150	***	0.3979	0.0152	***	0.3114	0.0155	***
UNEMPRAT	-0.0028	0.0007	***	-0.0016	0.0004	***	-0.0007	0.0006	

Note:
R-squared (1) = 0.832
R-squared (3) = 0.785
R-squared (5) = 0.801
*** $p < 0.001$
** $0.001 < p < 0.01$
* $0.01 < p < 0.05$

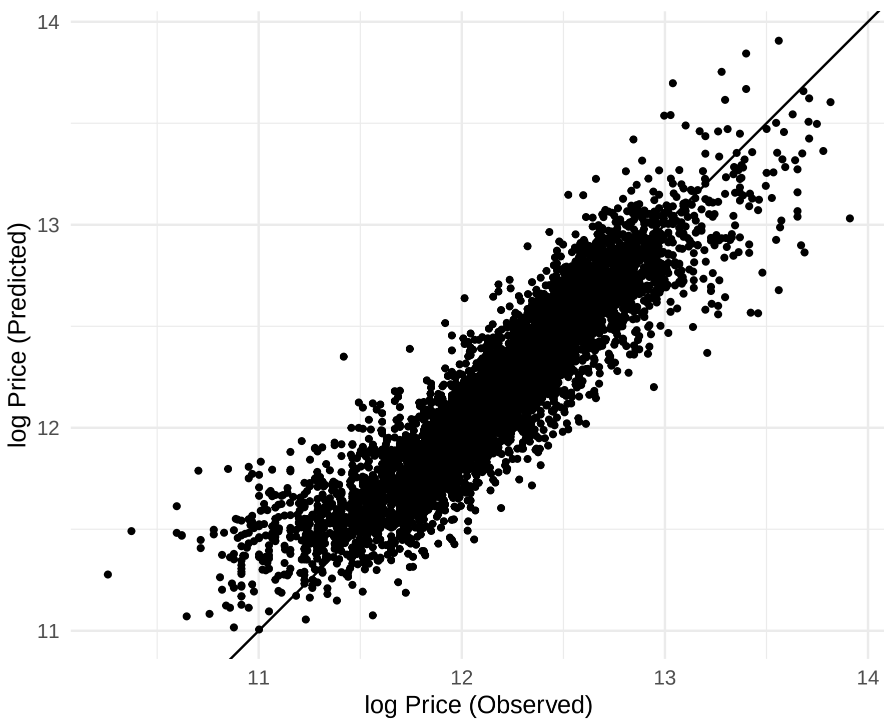

Figure 2.18: Predicted-Observed plot for multi-year real estate regression, 2006, including census data

2.2.4 Effect of DA-level census variables on house prices

The importance of the DA-level census variables to determination of house prices can be seen when the selling price of the "median house" of the Hamilton contiguous urban area dataset is predicted for census dissemination areas ordered by income. Using the 2006 dataset, one dissemination area was selected to represent each percentile of average after-tax household income; the predicted selling price for a 2006 "median house" (Age 51-to-99, detached 1.5 storey, with central air conditioning and forced-air heating, partially-finished basement, 2 full baths, 0 half baths, 3 bedrooms, 12 total rooms, 1 garage space and 1 total parking space, 386 m^2 lot area) was then calculated for each of these representative dissemination areas, using the actual census variables for that DA. The resulting predicted prices, along with a Loess trend line, is shown in figure 2.19.

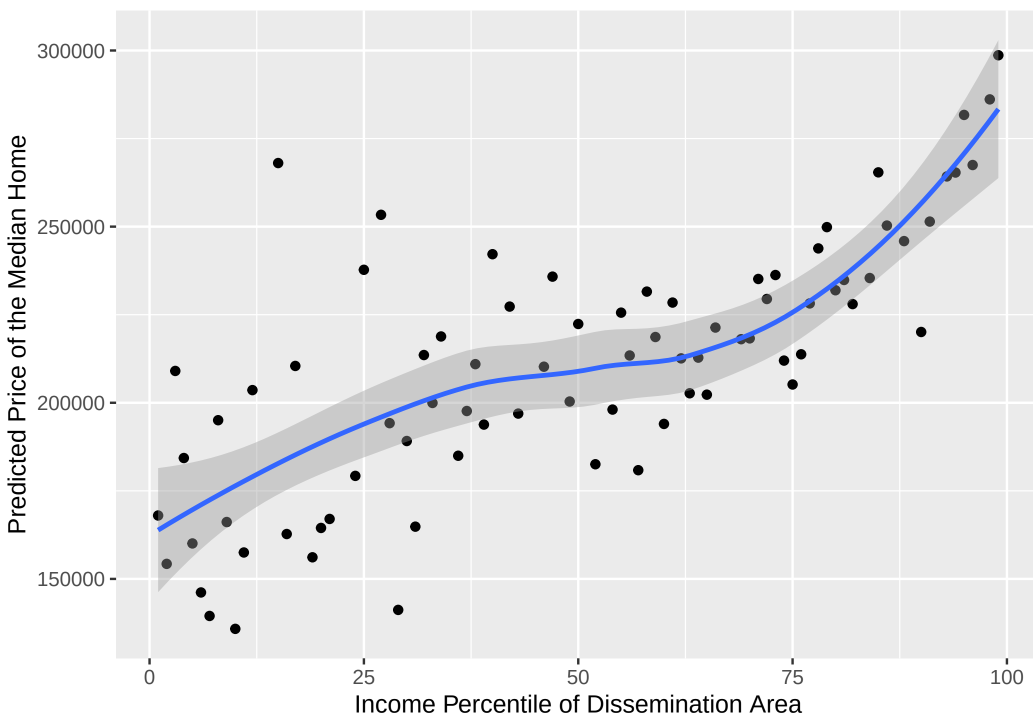

Figure 2.19: Predicted price for a median house, by census DA sorted by average after-tax household income, 2006

It can be seen, at least for 2006, that house prices are relatively unaffected by the social characteristics of neighbourhoods whose incomes are within the 30th-70th percentiles of income; but for houses in the DAs with the bottom ten percent of incomes, the average impact of their social variables is often a decrease in the selling price of a median home by fifty thousand dollars or more (but with large variation from DA to DA); while houses in DAs with incomes in the top ten percent see a consistent selling bonus of thirty thousand dollars or more over typical price. The incorporation of census variables into a housing price regression seems, then, to more significantly correct house prices at the extremes of the selling price range.

2.2.5 Trimming to contiguous urban residential area

A decision was made to trim the real estate dataset down to the contiguous urban residential area of Hamilton. Theoretical reasons can be advanced for doing this:

- the form and character of rural areas is significantly different than urban areas;
- partly due to the above, it can be expected that fully-rural house buyers (and perhaps buyers in rural towns like Freelton and Carlisle) have markedly different preferences, making these houses a significantly different submarket that if included in a regression may confound estimates for urban coefficient values;
- the topics under investigation (school distance, school contribution to neighbourhood character, walkability and so on) are exclusively urban geographical topics and have not been asserted to have any relevance to rural geography;
- primary school distance in particular will have a very different effect on rural house prices, since HWDSB policy of busing primary students who face a >1.6km trip to school will leave most rural houses out of the "can walk to school" group;
- most importantly, when we later move on to spatial analysis, the assumption of spatial interaction and spatial interconnection seems untenable both for rural houses (whose nearest neighbour distances are much greater) and for suburban housing areas that are unconnected to the

larger Hamilton area (such as Waterdown, Binbrook, Mount Hope, or even the lakefront properties in north Stoney Creek).

The last point above can be tested by examining the residual autocorrelation of a rural-only dataset. For this example, the 2016 sales dataset was trimmed to properties outside the contiguous urban area of Hamilton; exurban properties in Waterdown, and along the north shore of Lake Ontario in Stoney Creek, were also removed due to their peri-urban situation. The remaining entirely rural dataset of properties exhibit no systematic spatial autocorrelation of residuals, as can be seen in Figure 2.20.

Further, the p-values calculated for neighbour correlations are also insignificant after the first neighbour, as can be seen in Figure 2.21. This indicates that the expectation of house price autocorrelation is not typically met in rural areas; perhaps this is simply because of the larger distances between homes. Alternately, perhaps this indicates the lower significance of social variables to rural house prices: after all, one cares a lot less about the social characteristics of one's neighbours when they are much further away and there is less interpersonal interaction. If the latter is the case, then it would definitely be wise to exclude rural properties from a house price hedonic regression that includes social variables.

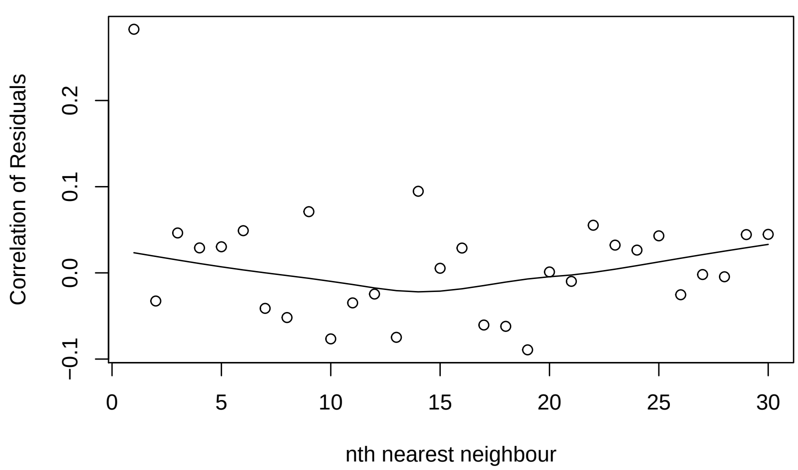

Figure 2.20: Residual autocorrelation for rural houses, 2016

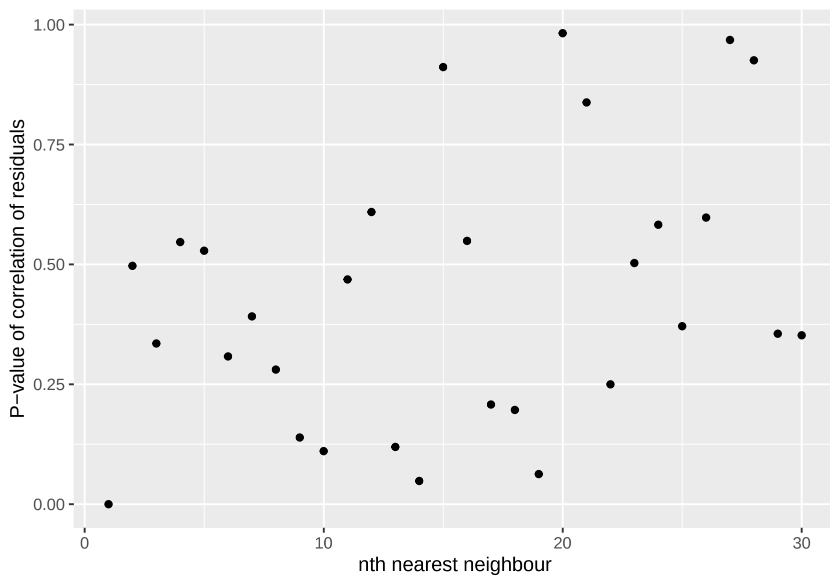

Figure 2.21: Residual autocorrelation p-values, rural houses only, 2016

2.3 School Data

Hamilton public school catchment boundaries and school locations were provided by the Hamilton Wentworth District School Board (HWDSB) for school years from 2005-6 through 2018-19; to ensure their accuracy, all HWDSB catchment files received were checked against previous and subsequent catchment files as well as against local news stories from the period detailing the ongoing school reorganizations in Hamilton.

Note that school catchments in Hamilton are fairly permeable; the HWDSB has a fairly open enrollment policy, allowing students to attend out-of-catchment schools as long as there are spaces available in those schools. The HWDSB also operates dedicated French immersion primary schools with their own separate, larger, catchments. Since 1990, the board has also operated a self-directed learning program at Westmount Secondary School, and more recently added an Advanced Placement program there; Westmount has no “catchment” as such, and takes in students from across the city.

Note that, in the HWDSB system, regression analysis of the importance of school distances is complicated by the Board’s primary grade regime; while some primary school catchments in the city are serviced by a single JK-to-grade-8 school, other primary catchments have separate JK-to-5 and 6-to-8 schools (or in the case of Ancaster and Dundas, JK-to-6 and 7-to-8), and these "middle schools" generally have catchments that span several primary "feeder" catchments. The HWDSB has been slowly converting all its primary schools to JK-8; however, even now in 2021 some JK-5 and 6-8 schools do remain in the HWDSB inventory. Thus, in some locations, the calculated distance to "primary" school is the same as the calculated distance to "middle" school; in other locations, the calculated distances will be different.

For the 2015-16 and 2016-17 school years, no public board school closures or catchment adjustments happened. For the 2010-11 and 2011-12 school years, the only apparent changes were:

- the reassignment of the city blocks bounded by John, James, King and Barton from Hess Street school to Dr. J. Edgar Davey school, for primary grades only;
- the assignment of a middle school catchment in a part of Waterdown to Guy Brown school, which does not affect the dataset because it is outside the contiguous urban area of Hamilton.

Several changes occurred in HWDSB primary and middle-school catchments during the 2005-6 and 2006-7 school years, concentrated mainly in the downtown core (Faulkner, 2006c):

- **Queen Victoria** and **Stinson**: The old Queen Victoria school (JK-5, built 1965) was closed in June 2006, so that it could be demolished and a new school built on its site. JK-3 students in the Queen Victoria and Stinson catchments were temporarily housed at Stinson school (also formerly JK-5); its catchment's grades 4-8 students, as well as the grades 4-8 students from the Stinson catchment, had formerly been bound for grades 6-8 at Sanford school, but their official middle-school catchment changed in summer 2006 to Queen Victoria. These grade 4-8 students were nevertheless temporarily housed at Sanford school, which itself had closed and sent its own students to Cathy Wever for September 2006. The new Queen Victoria school (JK-8) did not open until March 2009. (Faulkner, 2006c, 2006f; Kruchak, 2006)

- **Sanford**, **Gibson**, **Dr. Davey** and **Cathy Wever**: Cathy Wever school was built as a new JK-8 school for the city centre, on the same property as Sanford School; it opened on 5 September 2006 (Faulkner, 2006a; Kruchak, 2006). All prior Sanford and Gibson students transferred to the new Cathy Wever school, but Sanford and Gibson continued to be utilized as temporary holding school locations for other schools. Dr. Davey school, which had formerly been a JK-5 school feeding Sanford, became a JK-8 school and gained its own middle school catchment.

- **Prince of Wales**: This JK-8 school (built 1922) was closed in June 2006, and its students went to Gibson (JK-3) and Sanford (4-8) as holding schools until the new Prince of Wales building was completed; however, the Prince of Wales student body also included a number of students who lived in the over-capacity catchment for Dr. J. Edgar Davey school, and it is known that some of its catchment's students transferred instead to the new Cathy Wever school. For a month in September 2006, the Prince of Wales building was used as a temporary holding school for students of Hillcrest and Hillsdale whose own replacement building's opening was delayed. A new Prince of Wales school finally opened to students on 25 March 2009. (Faulkner, 2006b, 2007c; Kruchak, 2006)

- **Lawfield**: this Grade 6 to 8 school on the East Mountain was damaged by an F1 tornado on 9 Nov 2005. While there was no change to the catchment boundary, Lawfield students had to be bussed to Seneca Adult Learning Centre (a former primary school that had been closed in 2002) on the West Mountain until 19 November 2007, when a newly-constructed Lawfield school opened beside the site of the old school; the new Lawfield also absorbed the JK-5 cohort from neighbouring Vern Ames school, which had continued operating as a JK-5 school until the new Lawfield opened. (Faulkner, 2006e, 2007b; Fragomeni, 2007)

- **Hillcrest** and **Hillsdale**: A new Hillcrest school (JK-8) was built in the same East End park as Hillcrest (6-8, built in 1920) and Hillsdale (JK-5, built in 1971); meant to open for September 2006, poor weather during winter construction meant its opening was delayed until 10 October. Since the old Hillcrest and Hillsdale had already been demolished, however, all students had to bussed for a month to Prince of Wales school, which itself had been closed earlier that summer. Upon opening, the new Hillcrest absorbed the JK-5 cohort of Hillsdale and became a JK-8 school. (Faulkner, 2007b; Kruchak, 2006)

- **Central Park**, **Dundana**, **Dundas District** and **Sir William Osler**: Central Park, a JK-6 school, and Dundas District, a grades 7-8 school, were originally meant to be closed by the start of the 2006-7 school year and replaced with a new JK-8 school, which was eventually named Sir William Osler; this school was also to take over the north catchment of Dundana school (the area surrounding Veterans Park). Ultimately, Central Park and Dundas District had to remain open until 12 November 2007, as construction of the new Osler school was delayed by a failed property purchase and a labour dispute during construction. (Faulkner, 2007d, 2007a; Pona, 2007)

As can be seen, addressing the above complicated school catchment adjustments in our dataset cannot be done both correctly and simply: when calculating distance to school for each house, should a distance be calculated to a temporary holding school, or to the final school location? Do home buyers demand a "discount" for having neighbourhood students bussed to a temporary holding school, especially when the timeline of future school openings is perceived by them as uncertain? Even more, do prospective buyers pay a premium to be located near a brand-new school like Cathy Wever, or the new Prince of Wales or Queen Victoria, which has been outfitted with new equipment like smartboards, weight rooms and brand-new musical instruments for its music department?

It was decided to address the 2006 school catchment changes in a simplistic manner:

- in the case of Sanford, Gibson and Cathy Wever, primary school distance for 2006 will be calculated to Sanford (which in any case is right next door to the new Cathy Wever school);
- in the case of Dr. Davey, its primary school catchment's middle school distance will be calculated to Dr. Davey school, since it gains its own

middle school catchment by September 2006;

- in the case of Queen Victoria and Stinson, primary school and middle school distance will be calculated to Queen Victoria (which will be the primary and middle school when it opens);
- Prince of Wales and Hillcrest distances will be calculated per 2005-6 school locations, since the school locations don't change significantly;
- in the case of Central Park/Dundana/Dundas District/Sir William Osler, primary school distances for Central Park and north Dundana catchments will be calculated to the future William Osler site, and middle school distances for Central Park's three catchments will be calculated to the future William Osler site;
- holding schools will be ignored in simple cases where administrative boundaries don't change (Hillcrest, Hillsdale, Lawfield), under the assumption that if distance to school is important to homebuyers it is the distance to the future permanent school that would matter.

Generally, in the case of long-term disruption of the school network (such as the multi-year changes in the downtown core), it was considered better to use distances to final destination schools, even if the final destination school will not be open by the 2006-7 school year.

2.3.1 Note on the Catholic school board

Note that a Catholic school system exists in Hamilton in parallel with the public system; in the 2016-2017 school year, the the Hamilton-Wentworth Catholic District School Board (HWCDSB) made up approximately 35% of primary school students in Hamilton participating in provincial EQAO tests at the grade 3 and grade 6 level and approximately 44% of students participating in EQAO tests at the grade 9 level. These enrollment percentages correspond approximately to the 34% share of Hamilton's population who identified their religion as Catholic in the 2011 National Household Survey; however, the HWCDSB does not limit enrollment to students with a Catholic baptismal certificate. Nevertheless, while a nontrivial number of Protestant, Muslim and other (i.e. non-Roman) Catholic students in Hamilton do attend Roman Catholic primary and secondary schools, it is assumed for the purpose of this study that Catholic schools are not a simple substitute for public schools for the general homebuyer, since the general homebuyer would not be expected to either change their religion or change their child's schooling from Public to Catholic based on neighbourhood school accessibility.

2.3.2 EQAO data

"Quality of school" is a characteristic that may affect a homebuyer's willingness to purchase a house at a particular price; it may also affect a homebuyer's desire to be close to a school. It was decided to operationalize "quality of school" by using the yearly results of the tests conducted by the Education Quality and Accountability Office (EQAO); while this information is not generally available to the public, related publicly-available measures like the Fraser Institute school rankings will be correlated with school scores, and EQAO testing data has the added benefit of providing a score for the Hamilton JK-5 schools which are not rated by the Fraser Institute.

Detailed yearly (2004 to 2018) results for Ontario's province-wide standardized tests (for reading, writing and math in grade 3 and grade 6, for academic and applied math in grade 9, and for the grade 10 highschool literacy test) were provided by the Province of Ontario's Education Quality and Accountability Office. For each test score, a "school quality" indicator was generated from its standardized test scores. Firstly, the EQAO score for "percent of students meeting or exceeding expectations" score was found for each test at each school; the mean "meet or exceed expectations" score for each year's test was then calculated for the city, and then each test score for each test at each school grade and year was converted into a value representing that school's test results' standard deviation above or below this citywide mean.

This was done for each school's EQAO test grade, both for the census year (2006, 2011 and 2016) and for the following year (2007, 2012 and 2017). Finally, a both-years average of all mean-standardized scores for each grade level was calculated. This two-year averaging method is similar to Black's (1999) averaging of test results over three years; the intent of averaging is to eliminate year-to-year variation in test results to get a more invariant indicator of school quality.

Figures 2.22 and 2.23 plot the average of grade 3, grade 6 and grade 9 normalized 2-year-average school scores across the urban contiguous residential area of Hamilton; as can be seen, the average school score is by far the lowest in the downtown core, average throughout the rest of Hamilton, and by far the highest in the peripheral neighbourhoods of Dundas, Westdale, Ancaster and Stoney Creek. Interestingly, the differentiation between high scores on the periphery and low scores in the city and Mountain seems to have increased over those ten years.

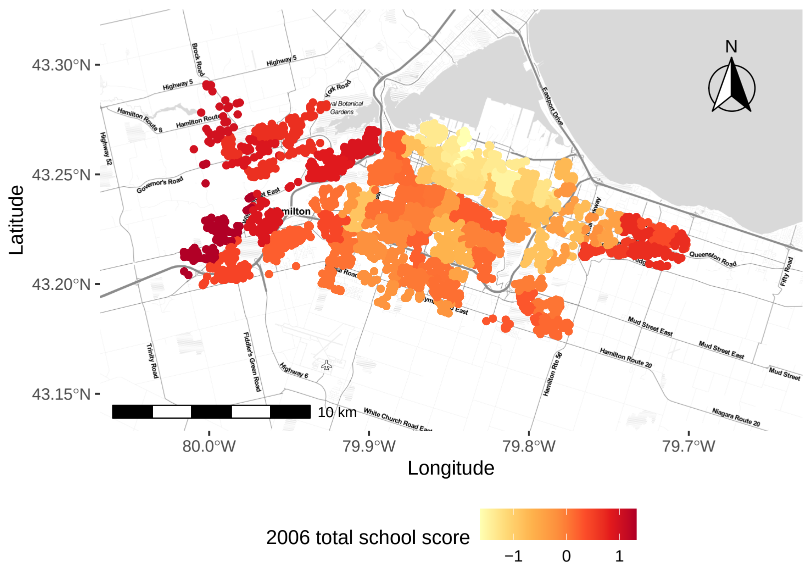

Figure 2.22: Spatial distribution of total school score, 2006

Analysis of Pearson correlation matrices (Figures 2.24, 2.25 and 2.26), calculated at the level of the house transaction, do confirm that grade 3, 6 and 9 school scores do correlate highly with each other in all three census years. It is interesting, though, that intra-grade score correlation is approximately as high as the correlation between school scores and both average household income and closing price of house, with lot area close behind; schools in areas with higher incomes, higher house prices and larger lot areas tend to have higher scores. A strong negative correlation is also found between school scores and percent of neighbourhood aged 25-64 with less than highschool diploma: it would seem that aggregate student ability depends to a fair extent on the educational level of their neighbourhood.

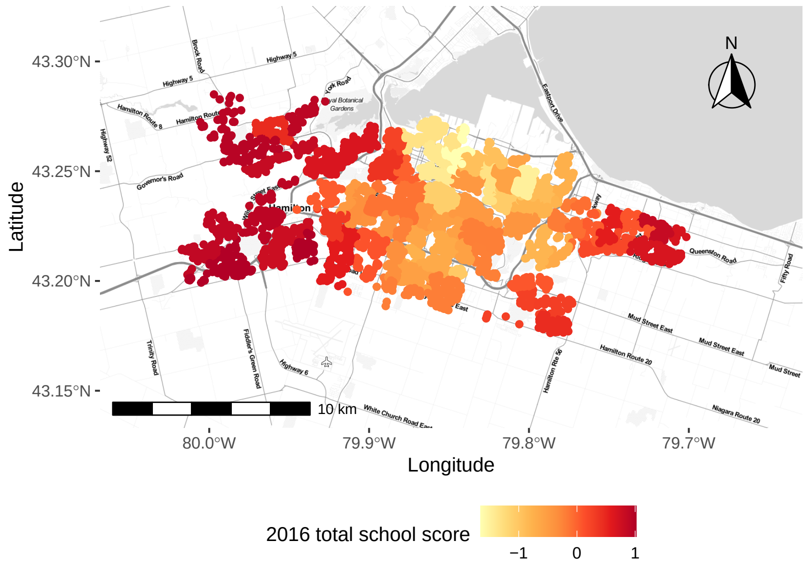

Figure 2.23: Spatial distribution of total school score, 2016

Regression of school scores on the contents of these Pearson correlation matrices (see Table 2.7) shows that the marginal effects of these variables are unambiguous in sign and significance across all three grade scores for house price (positive), lot area (positive), percent ages 25-64 with no highschool diploma (negative), and average household income (positive).

2.3.3 Distance to school calculations

Each house in the sales data sets for 2006, 2011 and 2016 was tagged with its primary, middle school and highschool catchments so that each house's distance to its catchment school (which is not necessarily its "nearest school") could be calculated. Distances to schools were calculated as Euclidean only.

For the 2016 dataset, network distances were also calculated, and were found to have a minimum correlation with Euclidean distances of 0.939 for JK-5 schools and a maximum of 0.978 for secondary schools (see Figure 2.27). Imperfections were found in the walking network which introduced sufficient enough error in calculated network distances to make it an obviously inferior option to using Euclidean distance: for example, there was an Ancaster house in the 2016 dataset in the catchment for Fessenden school whose network distance was

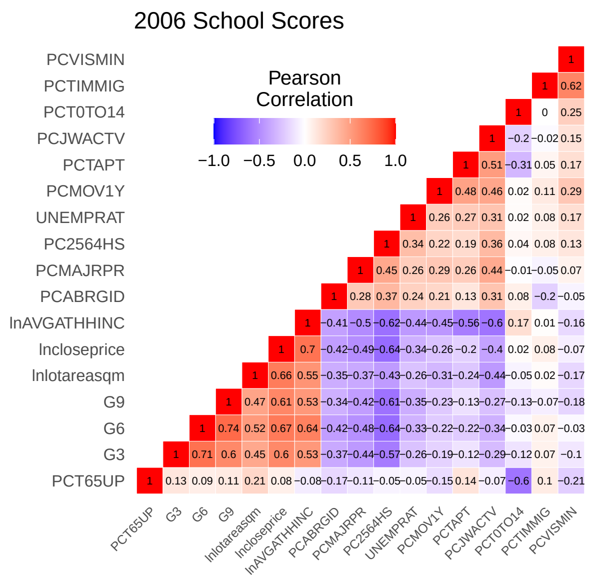

Figure 2.24: Pearson correlation matrix, school scores, 2006

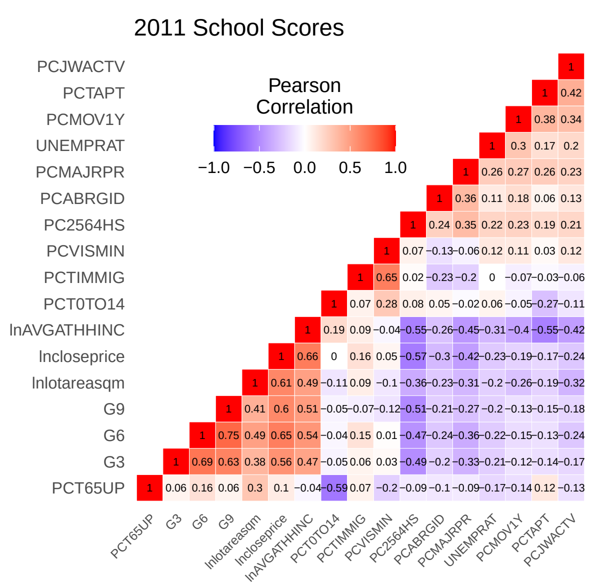

Figure 2.25: Pearson correlation matrix, school scores, 2011

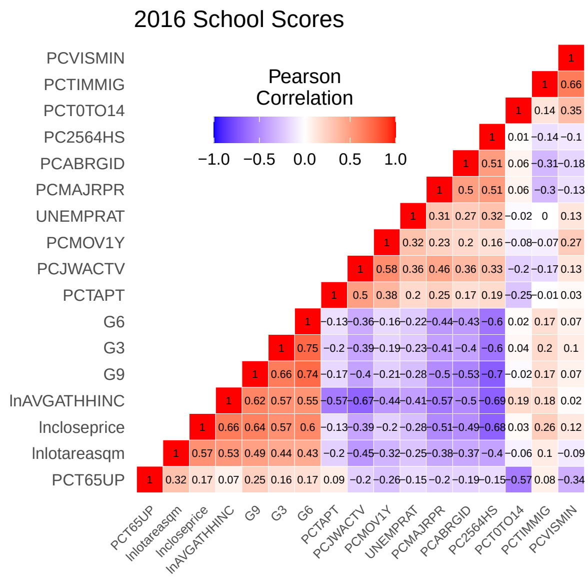

Figure 2.26: Pearson correlation matrix, school scores, 2016

Table 2.7: Regression of school scores on census variables, 2016

Variable	Grade 3 (1)			Grade 6 (2)			Grade 9 (3)		
term	β	s.e.	p	β	s.e.	p	β	s.e.	p
(Intercept)	-8.6014	0.6809	***	-11.1036	0.6853	***	-7.0886	0.5940	***
lncloseprice	0.3060	0.0352	***	0.4655	0.0355	***	0.3359	0.0307	***
lnlotareasqm	0.1606	0.0183	***	0.1183	0.0184	***	0.1693	0.0160	***
PCT0TO14	-0.1656	0.2243		0.3114	0.2257		-0.4818	0.1957	*
PCTAPT	0.0766	0.0574		0.4222	0.0578	***	0.2404	0.0501	***
PCTIMMIG	0.3161	0.1268	*	0.0022	0.1276		-0.2631	0.1106	*
PCABRGID	-0.2435	0.3267		-1.7511	0.3289	***	-3.3119	0.2850	***
PCVISMIN	0.2415	0.1116	*	-0.0642	0.1123		0.1739	0.0973	
PCMAJRPR	0.0337	0.1810		-0.7203	0.1822	***	-0.6981	0.1579	***
PC2564HS	-1.8627	0.0913	***	-1.5009	0.0919	***	-2.1581	0.0797	***
PCJWACTV	-0.7591	0.1077	***	-0.5873	0.1084	***	-0.5920	0.0940	***
PCMOV1Y	0.3734	0.1499	*	0.5061	0.1509	***	0.0739	0.1308	
lnAVGATHHINC	0.3723	0.0593	***	0.4227	0.0597	***	0.2369	0.0517	***
UNEMPRAT	0.0049	0.0020	*	0.0079	0.0020	***	0.0031	0.0018	

Note:
R-squared (1) = 0.45
R-squared (2) = 0.46
R-squared (3) = 0.586
*** $p < 0.001$
** $0.001 < p < 0.01$
* $0.01 < p < 0.05$

calculated circuitously through the road network of the neighbourhood, despite the house being located directly next to a catwalk that emptied out directly into Fessenden school's park. Another house in the 2016 dataset, in Hamilton, had a circuitous path to Glendale highschool calculated despite the house being located literally across the road from Glendale.

The very high correlation of Euclidean and network distances, and the presence of an unknown number of imperfections in the walking network, suggested that it would be better to use the Euclidean distance in further calculations.

Do any social or house variables correlate with school distance? Regressing primary school distance on the full set of house characteristics and DA-level census values (at the level of the house), the following consistent, unambiguous and statistically significant marginal effects were found for all three dataset years:

- for house age 15 or older, the house is more likely to be closer to primary schools;
- townhouses have slightly higher distances to primary schools;
- higher lot area correlates with higher distances to schools;
- a higher percent of the population taking active transit to work correlates with shorter distance to schools;
- a higher percent of the DA population who moved in the past year (a sign of economic stress) correlates with higher distances to schools;
- a higher DA unemployment rate (a sign of economic stress) correlates with higher distances to schools.

2.4 Summary of datasets

The categorical variables used in all regressions for this paper are described in Table 2.8; descriptive statistics for continuous variables are summarized (for 2016) in Table 2.9.

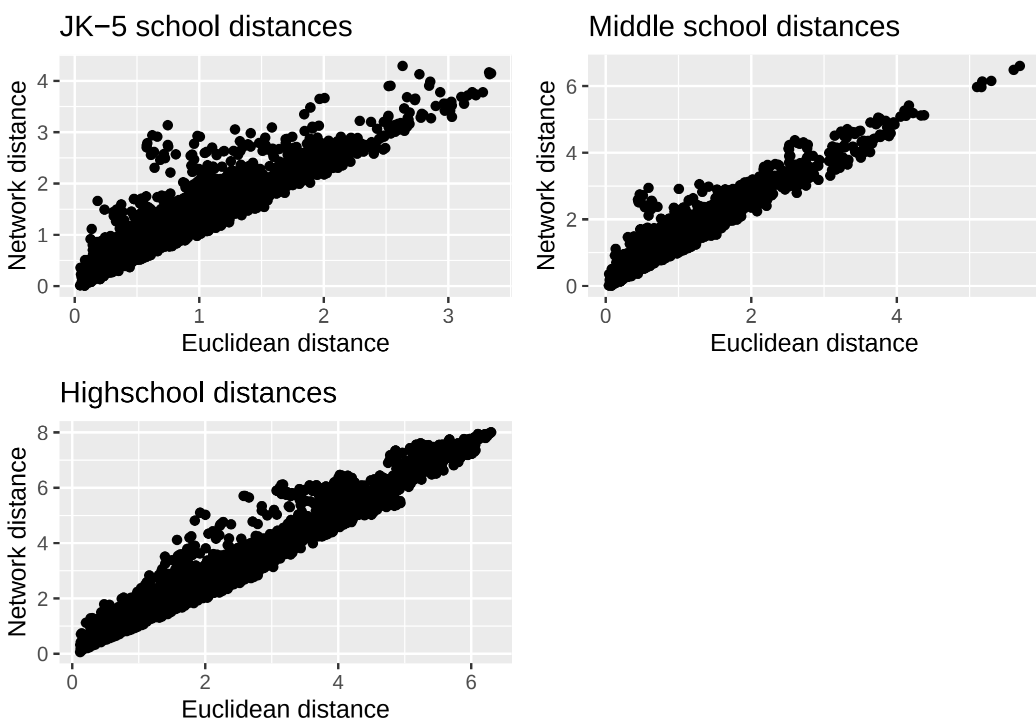

Figure 2.27: Euclidean vs network distance, kilometres, 2016

Table 2.8: Descriptive statistics for categorical variables

Statistic	Definitions
Approximate.Age	categorical for approx. age of house (reference=new)
Aircon.Central.YN	house has central air conditioning (reference=yes)
Basement.Fin	categorical for basement (reference = none)
Heat.FA	categorical for forced air heating (reference = none)
House.Type	categorical for house type (reference=detached)
Quarter	categorical for quarter of sale (reference = Q1)
Storeys	categorical for number of storeys (reference = bungalow)

Table 2.9: Descriptive statistics for continuous variables, 2016

Statistic	Definitions	Min	Q1	Median	Mean	Q2	Max
lncloseprice	natural log of closing price	10.968	12.614	12.848	12.846	13.081	14.832
Baths.Full	number of full bathrooms	1.000	1.000	2.000	1.777	2.000	10.000
Baths.Half	number of half bathrooms	0.000	0.000	0.000	0.438	1.000	3.000
Beds.Total	number of bedrooms	0.000	3.000	3.000	3.483	4.000	13.000
Garage.Spaces.Number	number of garage spaces	0.000	0.000	1.000	0.735	1.000	4.000
lnlotareasqm	natural log of lot area, sq. m.	3.879	5.617	5.952	5.968	6.236	10.206
Parking.Spaces.Total	total number of parking spaces	0.000	0.000	1.000	1.645	3.000	7.000
Room.Count	total number of rooms	1.000	10.000	12.000	11.864	14.000	20.000
PCT0TO14	% DA pop. ages 0 to 14	0.005	0.128	0.151	0.157	0.184	0.360
PCT65UP	% DA pop. ages 65 and up	0.026	0.110	0.146	0.162	0.200	0.776
PCTAPT	% DA pop. living in apartment	0.000	0.000	0.025	0.126	0.170	1.027
PCTIMMIG	% DA pop. immigrants	0.000	0.157	0.211	0.230	0.303	0.644
PCABRGID	% DA pop. of aboriginal identity	0.000	0.000	0.022	0.029	0.045	0.195
PCVISMIN	% DA pop. of visible minority status	0.000	0.078	0.147	0.175	0.238	0.803
PCMAJRPR	% of houses in DA requiring major repair	0.000	0.032	0.067	0.077	0.114	0.375
PC2564HS	% DA pop. ages 25-64 not completed postsecondary	0.000	0.301	0.406	0.405	0.517	0.778
PCJWACTV	% daily work trips using active transport	0.000	0.092	0.162	0.189	0.275	0.700
PCMOV1Y	% population in DA moved in past year	0.000	0.071	0.107	0.122	0.162	0.450
UNEMPRAT	Unemployment rate (0-100 scale)	0.000	4.800	6.900	7.754	10.400	32.500
lnAVGATHHINC	natural log of avg. DA after-tax household income	10.081	10.914	11.155	11.159	11.379	12.244
G3	2-year avg. of EQAO Gr. 3 test scores (s.d.)	-1.804	-0.946	-0.358	-0.278	0.471	1.645
G6	2-year avg. of EQAO Gr. 6 test scores (s.d.)	-1.816	-1.061	-0.265	-0.313	0.256	1.255
G9	2-year avg. of EQAO highschool test scores (s.d.)	-1.804	-0.945	-0.198	-0.300	0.618	0.928
PPdist	dist. to catchment public primary school (km)	0.043	0.374	0.586	0.726	0.963	3.343
PMdist	dist. to catchment public middle school (km)	0.043	0.429	0.715	0.862	1.139	5.690
PHdist	dist. to catchment public highschool (km)	0.118	1.042	1.685	1.976	2.525	6.299

3 Census year school distance regressions

3.1 The Base model

The starting point for analysis is the estimation of a final non-spatial hedonic model, with the natural log of closing sale price regressed on the variables in the 2006, 2011 and 2016 datasets. Results for these regressions (house structure type coefficients not included) are summarized in Table 3.1.

As can be seen, in all three years a mild premium is paid for houses in high-scoring school catchments, even after controlling for several socioeconomic variables obtained from the census datasets: a one-standard-deviation difference in grade 6 EQAO scores is associated with an unambiguous 2.8%-2.9% change in house price for all three census years, while a one-standard-deviation difference in highschool EQAO tests yields a much larger premium in all three years. The results for grade 3 EQAO tests, however, are ambiguous across the three census years: no premium is found in 2016.

There is also a mild premium found in all three census years for being located closer to a school: a 1km difference in primary (JK-5) school distance affects house prices by more than 1% in each year studied, though admittedly the coefficient for 2006 is not significant at a 5% level. A slightly lesser penalty is found in all three years for increasing distance to highschool.

Note, again, that in Hamilton some primary schools are JK-8, while others are JK-5 and feed into a different school for grades 6 to 8; this makes interpretation of the value of distance to primary or middle school more complicated. Where the primary school is JK-8, the total distance effect is $PPdist + PMdist$; where the JK-5 school is different, there is a separate calculated benefit for middle school distance, which in the table of results is not statistically different from zero in 2011 or 2016. However, in a case where the "primary school" is a JK-8 school, the combination $PPdist + PMdist$ *is* statistically significant at a 5% level in 2006 and at close to a 5% level in 2011.

As an aside, another interesting result is the unambiguous positive effect on house prices for neighbourhoods with a high percentage of employed residents taking active transit (bus, walking or cycling) to work: a 10% difference in active travel to work is associated with an approximately 2.5% increase in house prices. It may be hypothesized that people who take active transit to work are mostly found in areas with high transit (e.g. city bus system) availability: highest active-transit-to-work areas are in the downtown core (Wentworth to Locke) and the area around McMaster University, where buses run every few minutes and an express bus is available. If this hypothesis is true, then homebuyers may indeed pay a premium to be located in areas with high transit availability; then again, they may simply be paying a premium to be located near the extra commercial amenities that are typically found along the main streets in areas supplied with higher public transit access.

It is also interesting that the age penalty for older houses decreases from 2006 to 2016. As can be seen in the Quarter dummy variables, and as demonstrated previously with the construction of a hedonic house price index, 2016 was the year where aggressive (>15% yoy) house price increases began in Hamilton; it may be, then, that decreased penalties for house age may be predictive of a housing bubble. Further to this, it may also be related that the premium for lot size decreased in 2016, and that the penalty for lack of central air conditioning also decreased: small lot size and older HVAC systems are typically found with the houses in the urban core.

3.2 A first spatial model

In order to move to a spatial model, a conjecture about the scale and form of spatial interaction must first be hypothesized. In the case of form, it is possible either that the econometric model is fully specified in variables but that house prices in the model are influenced by neighbouring house prices (i.e. that $E(P_i|P_{j\neq i}) \neq E(P)$), or that the model is underspecified and that there are relevant spatially-autocorrelated omitted variables that are contributing to a spatially autocorrelated error term (which would mean that $E(\epsilon_i|\epsilon_{j\neq i}) \neq 0$).

One can then determine which of these two types of spatial interaction are occurring using Lagrange multiplier tests (Anselin, 1988), assuming a theory about the scale of spatial interaction has been settled on. Here, a scale of interaction is determined by plotting and inspecting a variogram of Table 3.1's regression's residuals to determine approximately to what distance limit spatial autocorrelation exists within the data set.

A first sill can be seen in all three residual variograms shown in Figure 3.1, indicating that spatial autocorrelation of residuals is reduced after some distance. Interestingly, while in 2006 and 2016 the sill seems to be at approximately 1000m, the sill in 2011 seems to instead be at either 600m or 1200m.

As an aside, it would be interesting to investigate in future whether the location of a housing market's residual variogram sill is indicative of some characteristic of housing market strength: in 2011, Hamilton's hedonic house price index posted its second-weakest year-over-year returns of the 2006-2017 period, and (as far as conventional asset markets are concerned, at least) it is known that cross-market price correlation does increase when markets are under stress (Bertero & Mayer, 1990; King & Wadhwani, 1990). This could also be an empirical truth for scale of spatial autocorrelation of house prices, though the question will not be addressed here.

Given the sill locations for the 2006 and 2016 variograms, a 1000m cutoff for spatial interaction was chosen for spatial modelling. Interestingly, this is

Table 3.1: Non-spatial regression results

Variable	2006 (1)			2011 (2)			2016 (3)		
term	β	s.e.	p	β	s.e.	p	β	s.e.	p
(Intercept)	8.332	0.179	***	8.429	0.183	***	8.943	0.198	***
Approximate.Age0TO5	-0.178	0.027	***	-0.135	0.027	***	-0.091	0.023	***
Approximate.Age6TO15	-0.326	0.026	***	-0.291	0.023	***	-0.213	0.020	***
Approximate.Age16TO30	-0.447	0.027	***	-0.391	0.024	***	-0.300	0.021	***
Approximate.Age31TO50	-0.455	0.027	***	-0.409	0.025	***	-0.299	0.021	***
Approximate.Age51TO99	-0.455	0.026	***	-0.391	0.023	***	-0.290	0.020	***
Approximate.AgeUnknown	-0.501	0.026	***	-0.439	0.023	***	-0.327	0.020	***
AC.YN-NoCentralAir	-0.120	0.007	***	-0.142	0.008	***	-0.077	0.006	***
Basement-Unfinished	0.058	0.028	*	0.236	0.049	***	0.039	0.025	
Basement-Partial	0.075	0.028	**	0.257	0.049	***	0.045	0.026	
Basement-Finished	0.113	0.028	***	0.283	0.049	***	0.062	0.026	*
Baths.Full	0.050	0.006	***	0.071	0.006	***	0.067	0.005	***
Baths.Half	0.050	0.007	***	0.069	0.007	***	0.056	0.005	***
Beds.Total	0.007	0.003	*	0.018	0.004	***	0.017	0.003	***
Garage.Spaces.Number	0.037	0.005	***	0.025	0.006	***	0.027	0.005	***
Heat.FA-ForcedAir	-0.066	0.012	***	-0.062	0.015	***	-0.037	0.012	**
Parking.Spaces.Total	0.010	0.003	***	0.013	0.003	***	0.008	0.003	**
Quarter2	0.034	0.007	***	0.043	0.009	***	0.088	0.007	***
Quarter3	0.064	0.007	***	0.065	0.009	***	0.132	0.007	***
Quarter4	0.056	0.008	***	0.046	0.009	***	0.165	0.007	***
Room.Count	0.020	0.002	***	0.015	0.002	***	0.014	0.001	***
lnlotareasqm	0.203	0.007	***	0.183	0.008	***	0.120	0.006	***
PCT0TO14	-0.142	0.076		-0.516	0.088	***	-0.244	0.074	***
PCT65UP	0.250	0.047	***	0.116	0.052	*	0.067	0.045	
PCTAPT	0.121	0.018	***	0.166	0.020	***	0.242	0.016	***
PCTIMMIG	0.054	0.034		0.012	0.036		0.144	0.038	***
PCABRGID	-0.339	0.090	***	-0.142	0.073		-0.414	0.091	***
PCVISMIN	-0.031	0.036		0.141	0.034	***	-0.064	0.033	
PCMAJRPR	-0.198	0.047	***	-0.239	0.037	***	-0.210	0.050	***
PC2564HS	-0.435	0.032	***	-0.378	0.031	***	-0.401	0.027	***
PCJWACTV	0.246	0.031	***	0.133	0.028	***	0.211	0.031	***
PCMOV1Y	0.026	0.041		0.139	0.038	***	0.073	0.042	
UNEMPRAT	-0.001	0.001		-0.001	0.000		-0.001	0.001	
lnAVGATHHINC	0.249	0.016	***	0.245	0.016	***	0.280	0.017	***
G3	0.014	0.005	**	0.044	0.005	***	-0.008	0.005	
G6	0.029	0.005	***	0.028	0.006	***	0.029	0.005	***
G9	0.080	0.005	***	0.079	0.006	***	0.053	0.005	***
PPdist	-0.012	0.007		-0.034	0.009	***	-0.016	0.006	*
PMdist	-0.024	0.006	***	0.004	0.007		0.003	0.005	
PHdist	-0.012	0.003	***	-0.010	0.003	**	-0.010	0.002	***

Note:
R-squared (1) = 0.84 ; F statistic (1) = 578.82
R-squared (2) = 0.82 ; F statistic (2) = 429.67
R-squared (3) = 0.81 ; F statistic (3) = 508.06
*** $p < 0.001$
** $0.001 < p < 0.01$
* $0.01 < p < 0.05$

approximately the size of the typical city block in Hamilton (main streets are mostly laid out in a $\frac{1}{2}$ by $\frac{6}{10}$ mile grid, to yield a concession net of city easements of 160 acres); this suggests that residual autocorrelation, whether caused by price correlation due to comparables analysis by realtors, or by some spatially autocorrelated unobserved variable that affects house prices, is approximately a city-block-scale phenomenon.

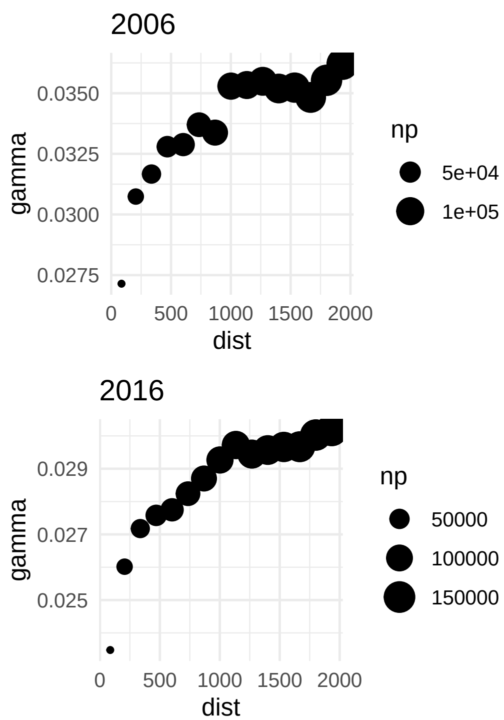

Figure 3.1: Variograms for Hamilton urban contiguous area, 2006, 2011 and 2016

Instead of investigating correlation at the distance scale, correlation can also be investigated at the scale of number of neighbours, k, as is shown in Figures 3.2, 3.3 and 3.4. We find that correlation between neighbours becomes not significantly different from 0 after approximately the 110th-nearest neighbour in 2016; the 200th nearest neighbour in 2011; and the 120th nearest neighbour in 2006. Parenthetically, the average distance between a house in the 2016 data set and its 100th nearest neighbour is approximately 925m; for the 200th nearest neighbour in the 2011 dataset, mean distance is 1622m; and for the 120th nearest neighbour in the 2006 dataset, mean distance is 1077m.

Table 3.2: Moran's I test results for 2006, 2011 and 2016

	Moran's I statistic standard deviate	Observed Moran I	Expectation
2006	68.2560	0.1128	-0.0017
2011	81.1290	0.1425	-0.0018
2016	51.3719	0.0814	-0.0016

Again, the 2011 correlation plot is different than the 2006 and 2016 plots: the estimated price correlation between a house and its 100th nearest neighbour seems to be significantly higher in the 2011 dataset, again suggesting a higher autocorrelation of house prices during a stressed market.

It is therefore decided to use 100 nearest neighbours in our analysis; it should remembered for later that the results of the spatial regressions might be misleading for the 2011 year, if there truly is temporal heterogeneity in the mechanism of spatial autocorrelation. Using a value of 100 for k, Moran's I tests on the residuals of the base models for 2006, 2011 and 2016 (see Table 3.2) yield highly significant (p«0.001) Moran's I p-values, as can be inferred by the standard deviates reported in the table. This indicates that there is definite spatial autocorrelation in the residuals of the non-spatial, econometric models: thus one should have little confidence in the estimates, standard errors, or p-values for the non-spatial regression results in Table 3.1, and it would be advised to move to a spatial regression model.

But which spatial model? Following the examples laid out in Florax, Folmer, & Rey (2003) and Osland (2010), we first perform Anselin's robust Lagrange multiplier tests: the LM-lag and LM-error test, respectively, of a null hypothesis of no spatial autocorrelation in the Y variable and in the error term. (The robust forms of these tests each correct for the possible presence of local lags in the other test's variable.) As can be seen in Table 3.3, in all three census-year housing datasets the robust LM-lag and robust LM-error tests indicate that lag-Y and lag-error are both picking up spatial autocorrelation: in this case, spatial regression should begin with the error model, whose Lagrange test provides the higher value.

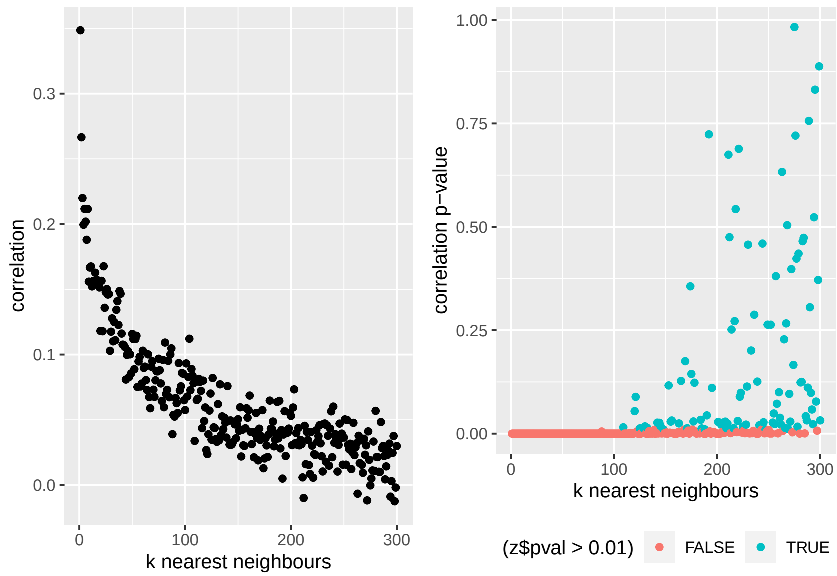

Figure 3.2: Neighbour correlation and p-value, 2006

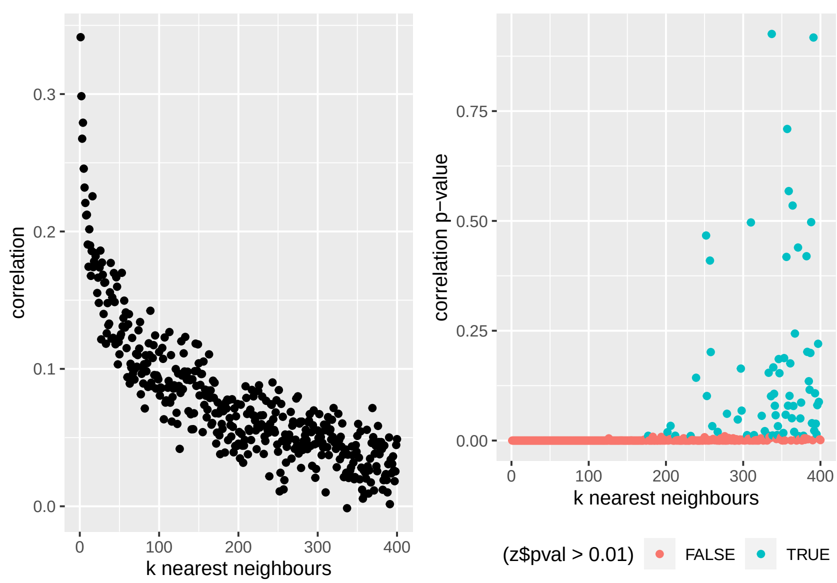

Figure 3.3: Neighbour correlation and p-value, 2011

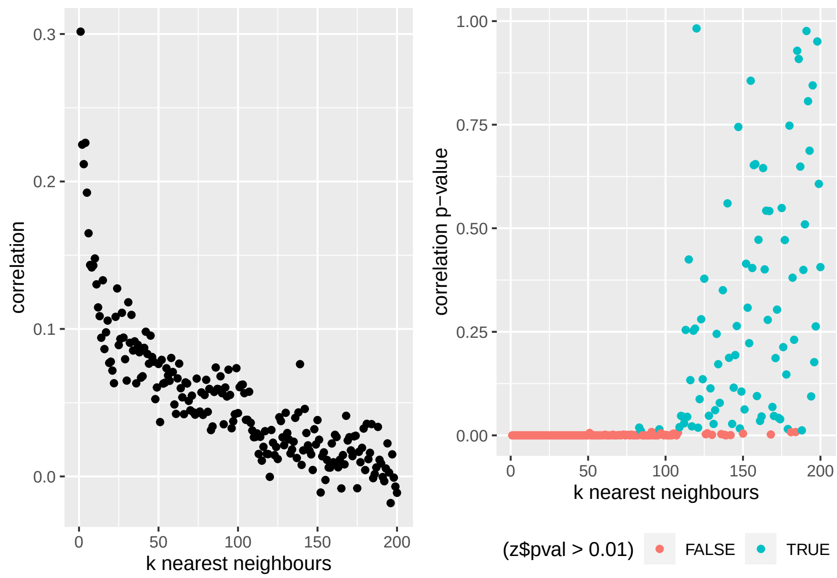

Figure 3.4: Neighbour correlation and p-value, 2016

Table 3.3: Robust Lagrange Multiplier test results for 2006, 2011 and 2016

	RLM error	RLM lag
2006	2885.1	369.5
2011	4010.9	498.7
2016	1530.7	241.4

3.2.1 A note on an alternative method for eliminating spatial autocorrelation

One alternative method that might successfully remove longer-distance spatial autocorrelation from a non-spatial regression would be to add a dummy variable for each Dissemination Area: doing this does eliminate the sills from the variograms. Adding DA dummies was found to produce results that had no residual autocorrelation according to Moran's I tests.

However, this is not a productive road to go down: the small number of house sales in each DA (on average, around 7 per year) means the individual DA dummy estimates cannot be estimated to any significance due to low n. But more significantly for this study, each DA's dummy estimate will also incorporate the value of that DA's average house distance to school, making a separate per-house estimation for the distance to school coefficients impossible. This is therefore discarded as a non-productive solution for this chapter.

3.3 Beginning the spatial model

Conducting a curve-fitting exercise on the kth-neighbour correlation values in Panel 1 of Figure 3.4 suggests that we may use an inverse square root distance weighting, i.e. that each weight w_{ij} between neighbours in the matrix should be calculated as:

$$w_{ij} = \frac{1}{\sqrt{d}_{ij}}.$$

Before calculating the inverse square-root distance weighting for all neighbour pairs in the weights matrix, all identical points (i.e. within-year repeat sales) are removed; only the first sale of a property in each year is used. After calculating inverse distance weighting, the weights are then row-standardized.

3.3.1 Spatial error models

Table 3.3 shows higher Lagrange test values for error than for Y-lag, for all three years. This indicates that we should first attempt a spatial error model, and then test to see if residual autocorrelation has been removed.

Table 3.4: Lag Y and lag error regression results

	2006		2011		2016	
term	lag error	lag Y	lag error	lag Y	lag error	lag Y
log likelihood	2000.5300	1902.4329	1503.4453	1396.9837	2401.2237	2294.9804
log likelihood of lm	1533.5429	1533.5429	921.2905	921.2905	2021.8365	2021.8365
Moran's I p-value	0.0296	0.0001	0.0205	0.0001	0.0879	0.0001
rho		0.5040		0.5872		0.5042
lambda	0.9684		0.9704		0.9592	
Spatial Hausman test p-value	0.0373		0.2543		0.0000	

A spatial error model was thus estimated for each census year; for completeness, a spatial lag model was also estimated. In all spatial regressions, a detrended log price was used; it was calculated by subtracting the house's quarter-of-sale Quarter coefficient (taken from the above final nonspatial regression) from its log house price, to avoid the problem of lagging a time-based variable in a spatial regression.

Next, pseudo *p*-values were calculated for the residuals of each regression using Monte Carlo simulation for Moran's I; they are reported in Table 3.4, along with log likelihood comparisons for each model, a Spatial Hausman test result for each lag error regression (Pace & LeSage, 2008), and rho or lambda statistics for the respective models. It can be seen that the lag error models do a fair job at removing residual spatial autocorrelation: the null hypothesis that residual autocorrelation has been eliminated cannot be rejected at the 99% level for any of the three years. However, each error model also yields a seemingly unreasonable value for λ: given that the error model is

$$P = X\beta + \epsilon, \epsilon = \lambda W \epsilon + u,$$

a $\lambda > 0.9$ would suggest almost 50% of log house price variation comes from unobserved autocorrelated spatial variables, and not from observed house characteristics. In addition, a Spatial Hausman test conducted on each lag error regression suggests that while lag-error might be a correct specification for 2006 and 2011, for some reason it fails in 2016.

While the lag error models might eliminate residual spatial autocorrelation according to the residual Moran's I values reported, but with unreasonable λ values, it is obvious that the lag-Y models all fail to eliminate residual spatial autocorrelation, given the pseudo p-values reported in the table. Given this, the next step is to move on to a Spatial Durbin model.

3.4 Spatial Durbin model

As noted previously, the Spatial Durbin model is of the form

$$P = \rho WP + X\beta - \rho WX\gamma + \epsilon,$$

where the object $\rho WX\gamma$ represents a spatial lag of the X vector, whose coefficient vector γ is allowed to have some elements set to 0 in order to subset out elements of the X vector that do not need to be lagged. If, in addition, $\gamma = -\rho\beta$, this model would then simplify to the spatial error model

$$P = X\beta + \epsilon, \epsilon = \lambda W\epsilon + u.$$

As such, while the spatial Durbin model estimates lags on both the Y and X variables in the regression equation, this is equivalent to estimating lags on Y and error. The inclusion of an X lag, though, causes particular trouble for interpretation, as each coefficient is estimated not just for the house characteristic but also for its lag neighbours. In the case of the Quarter variable, which is meant to be included only to allow for drift in the house price index over time, this variable must be removed from all the Spatial Durbin regressions; all prices in each census year were, as before, de-trended to first-quarter prices by subtracting the Quarter variable found in the previous econometric regressions.

The spatial Durbin models that are finally estimated are found to have no residual spatial autocorrelation and a superior AIC to either the lag-Y or lag-error models; a Likelihood Ratio test of the breaking of the common factor hypothesis (i.e. that the lag error model is not nested within the Spatial Durbin model) finds the breaking to be valid for all three years, with an exceedingly low p-value. Breusch-Pagan tests do still indicate heteroskedasticity in the residuals; this warns us that the estimates of standard error and p-values in our regressions are approximate only. Figures 3.7, 3.6 and 3.5 show that the model does a reasonably good job of predicting prices in aall three years - even at the extremes of the distribution, which was a notable problem in hedonic regression models.

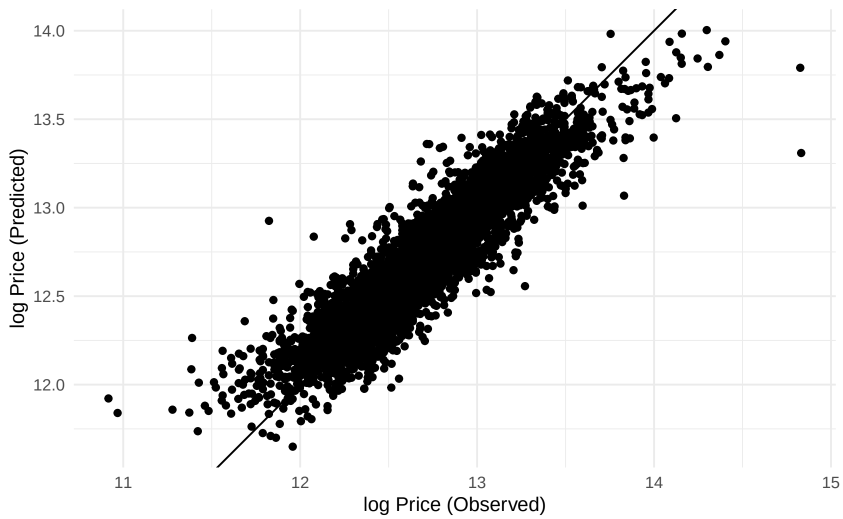

Figure 3.5: Predicted-Observed plot for Spatial Durbin regression, 2016

Unlike in a standard regression, an estimated coefficient β_k in a spatial Durbin model does *not* represent variable x_k's marginal effect $\frac{\delta y}{\delta x_k}$. Rather, a change in house i's variable $x_{i,k}$ is going to affect the price P_i for house i, but also indirectly affect the prices P_j of all houses $j \neq i$ that are defined as neighbours in matrix W through both the x-lags and the y-lags; these effects on prices $P_{\forall j \neq i \in W}$ will then also further affect the price P_i of house i through the y-lag. These feedback effects need to be calculated before anything can be said about the marginal effects of areal characteristics (such as census variables) in any spatial lag model (Lesage & Pace, 2009).

The marginal effects in a spatial Durbin model therefore needed to be calculated separately (see Piras (2014) for an explanation); a comparison between the base and spatial Durbin model's estimated approximate marginal effects can be seen in Tables 3.5, 3.6 and 3.7.

The move to a spatial Durbin model does complicate interpretation of the

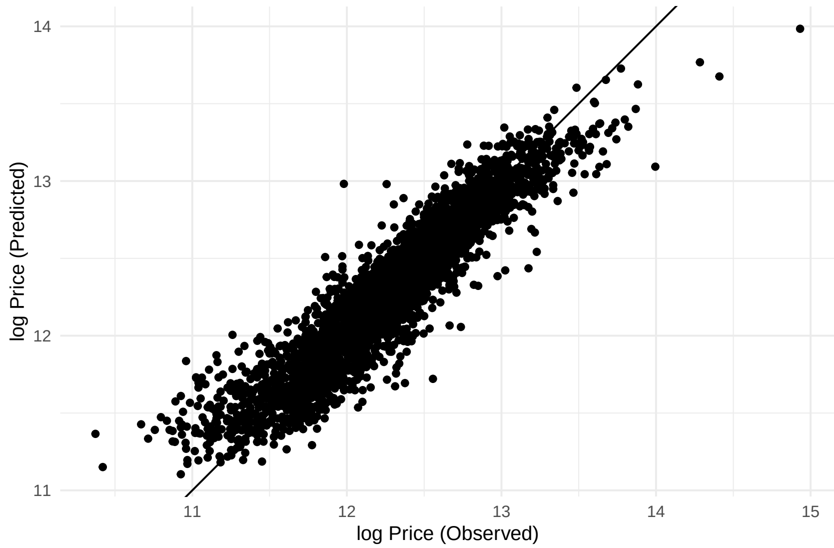

Figure 3.6: Predicted-Observed plot for Spatial Durbin regression, 2011

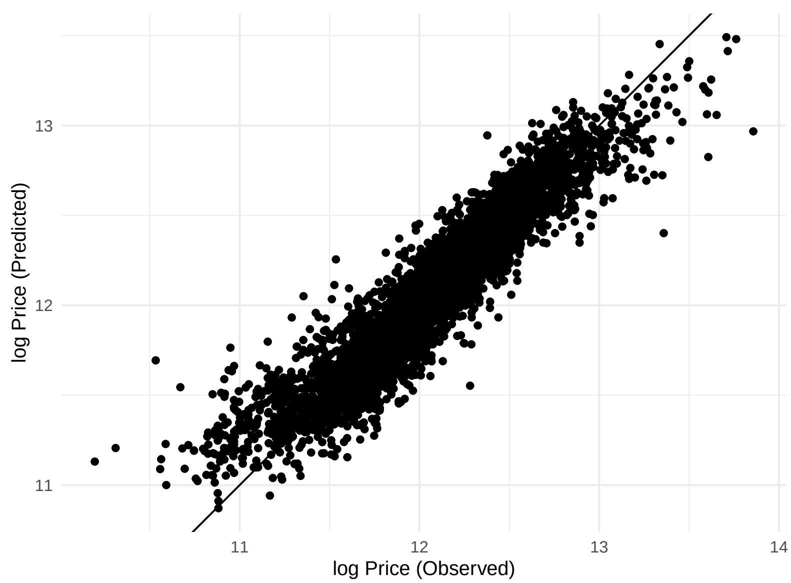

Figure 3.7: Predicted-Observed plot for Spatial Durbin regression, 2006

regression results: not only does the marginal effect of a change in an x_k have to be calculated separately instead of interpreted from β, but the interpretation of the empirical mechanisms of direct, indirect, and total effects due to a change in x_k becomes difficult. See Small & Steimetz (2012) for a discussion of the economic point of view in interpreting these results; in our case, we will limit ourselves to interpretation of the effect of school distance on house prices, and leave interpretation of control variables alone.

3.5 Results

As can be seen in Tables 3.5, 3.6 and 3.7, the marginal effects of independent variables have *far* greater standard errors in a spatial regression than in a non-spatial, econometric regression; it can be seen that estimating a regression without taking into account autocorrelation in y or ϵ does lead to systematic overconfidence in estimated effects.

In the case of distance to public schools, we do find that

- estimates of the coefficients for total effect (i.e. $\frac{\delta y}{\delta x_k}$) of school distance are almost all more negative in the SDM (the one exception being highschool distance in the 2006 regression) than in the hedonic non-spatial regression;
- all estimates of total effect have a negative sign in all three years;
- but unfortunately, the standard error of the estimate is also much higher in all three years, meaning we are only able to assert an effect with 90% confidence for five of the nine school distance coefficient estimates, and only four at 95% confidence;
- in addition, from 2006 to 2016 there is significant variation in the estimate for value of middle school distance and highschool distance.

In any case, the higher value found for school distance in a spatial model may be due to both the direct benefit to the homeowner from living closer to a school, and the indirect benefit on the homeowner's price due to realtors' comparables analysis indirectly affecting their house's price by raising neighbours' prices - and possibly also an increase in welfare, due to some indirect effect of school proximity on neighbours' behaviour (Small & Steimetz, 2012).

Table 3.5: Spatial and non-spatial regression results, 2006

Variable	base regression			spatial coefficients			spatial lag coefficients			total spatial effects		
term	β	s.e.	p	β	s.e.	p	β	s.e.	p	$\frac{\delta y}{\delta x_k}$	s.e.	p
(Intercept)	8.325	0.181	***	2.984	0.862	***						
Approximate.Age0TO5	-0.180	0.027	***	-0.174	0.024	***	-0.262	0.224		-0.902	0.469	
Approximate.Age6TO15	-0.327	0.026	***	-0.274	0.023	***	-0.132	0.215		-0.839	0.423	*
Approximate.Age16TO30	-0.450	0.027	***	-0.382	0.025	***	-0.241	0.218		-1.288	0.436	**
Approximate.Age31TO50	-0.457	0.027	***	-0.397	0.024	***	-0.296	0.218		-1.434	0.425	***
Approximate.Age51TO99	-0.456	0.026	***	-0.406	0.024	***	-0.231	0.221		-1.318	0.441	**
Approximate.AgeUnknown	-0.503	0.026	***	-0.439	0.023	***	-0.187	0.212		-1.295	0.409	**
AC.YN-NoCentralAir	-0.122	0.007	***	-0.107	0.006	***	-0.027	0.062		-0.278	0.133	*
Basement-Unfinished	0.052	0.029		0.052	0.026	*	0.552	0.303		1.249	0.640	
Basement-Partial	0.069	0.029	*	0.061	0.026	*	0.522	0.300		1.205	0.638	
Basement-Finished	0.107	0.029	***	0.092	0.026	***	0.662	0.302	*	1.560	0.647	*
Baths.Full	0.050	0.006	***	0.046	0.005	***	-0.027	0.061		0.039	0.126	
Baths.Half	0.049	0.007	***	0.044	0.006	***	-0.103	0.059		-0.123	0.126	
Beds.Total	0.006	0.003		0.008	0.003	*	-0.062	0.025	*	-0.112	0.057	*
Garage.Spaces.Number	0.036	0.006	***	0.044	0.005	***	-0.058	0.054		-0.029	0.116	
Heat.FA-ForcedAir	-0.068	0.012	***	-0.045	0.011	***	-0.401	0.108	***	-0.922	0.246	***
HouseType-Semi	-0.099	0.012	***	-0.111	0.011	***	0.059	0.083		-0.108	0.180	
HouseType-Town	-0.148	0.015	***	-0.161	0.014	***	0.183	0.116		0.045	0.257	
Parking.Spaces.Total	0.010	0.003	***	0.008	0.003	**	-0.022	0.028		-0.030	0.060	
Room.Count	0.020	0.002	***	0.021	0.001	***	-0.012	0.013		0.019	0.027	
Storeys-RaisedBungalow	0.051	0.015	***	0.039	0.013	**	0.053	0.116		0.191	0.241	
Storeys-Sidesplit	0.033	0.018		0.027	0.016		0.102	0.146		0.267	0.301	
Storeys-Backsplit	0.075	0.012	***	0.054	0.011	***	-0.099	0.075		-0.093	0.162	
Storeys-1.5	0.015	0.008		0.007	0.008		0.034	0.059		0.086	0.126	
Storeys-2	0.044	0.009	***	0.050	0.008	***	-0.059	0.057		-0.019	0.124	
Storeys-2.5	0.111	0.012	***	0.103	0.012	***	0.049	0.068		0.314	0.140	*
Storeys-3	0.171	0.058	**	0.192	0.052	***	-0.301	0.669		-0.225	1.492	
Storeys-Other	0.134	0.049	**	0.096	0.044	*	2.006	0.667	**	4.346	1.289	***
lnlotareasqm	0.203	0.007	***	0.176	0.007	***	0.081	0.046		0.531	0.093	***
PCT0TO14	-0.137	0.077		0.062	0.079		-0.330	0.316		-0.554	0.649	
PCT65UP	0.253	0.047	***	0.191	0.049	***	-0.551	0.224	*	-0.745	0.463	
PCTAPT	0.119	0.018	***	0.039	0.018	*	0.019	0.081		0.119	0.180	
PCTIMMIG	0.057	0.035		0.021	0.037		0.402	0.122	**	0.875	0.256	***
PCABRGID	-0.342	0.092	***	0.091	0.094		-0.325	0.293		-0.485	0.591	
PCVISMIN	-0.032	0.036		0.025	0.036		-0.274	0.143		-0.516	0.300	
PCMAJRPR	-0.209	0.047	***	-0.069	0.045		-0.503	0.220	*	-1.183	0.481	*
PC2564HS	-0.442	0.033	***	-0.004	0.035		-0.579	0.130	***	-1.206	0.264	***
PCJWACTV	0.253	0.031	***	0.135	0.033	***	0.512	0.141	***	1.337	0.273	***
PCMOV1Y	0.023	0.041		0.031	0.039		-0.068	0.202		-0.076	0.426	
UNEMPRAT	-0.001	0.001		0.000	0.001		-0.001	0.003		-0.003	0.006	
lnAVGATHHINC	0.250	0.016	***	0.155	0.016	***	0.024	0.070		0.369	0.134	**
G3	0.014	0.005	**	0.009	0.007		-0.014	0.013		-0.011	0.024	
G6	0.030	0.005	***	-0.002	0.009		-0.004	0.017		-0.012	0.025	
G9	0.080	0.005	***	0.076	0.013	***	-0.084	0.018	***	-0.016	0.026	
PPdist	-0.012	0.007		0.016	0.010		-0.054	0.023	*	-0.077	0.042	
PMdist	-0.023	0.006	***	-0.016	0.009		-0.015	0.019		-0.062	0.029	*
PHdist	-0.012	0.003	***	-0.009	0.006		0.006	0.009		-0.006	0.012	
Quarter2	0.033	0.007	***									
Quarter3	0.064	0.007	***									
Quarter4	0.054	0.008	***									
rho				0.517	0.053	***						

Table 3.6: Spatial and non-spatial regression results, 2011

Variable	base regression			spatial coefficients			spatial lag coefficients			total spatial effects		
term	β	s.e.	p	β	s.e.	p	β	s.e.	p	$\frac{\delta y}{\delta x_k}$	s.e.	p
(Intercept)	8.379	0.185	***	1.787	0.978							
Approximate.Age0TO5	-0.129	0.027	***	-0.111	0.024	***	-0.119	0.184		-0.569	0.445	
Approximate.Age6TO15	-0.291	0.023	***	-0.273	0.020	***	-0.042	0.127		-0.779	0.306	*
Approximate.Age16TO30	-0.389	0.024	***	-0.367	0.022	***	0.032	0.142		-0.831	0.336	*
Approximate.Age31TO50	-0.408	0.025	***	-0.391	0.023	***	-0.041	0.149		-1.070	0.369	**
Approximate.Age51TO99	-0.389	0.024	***	-0.389	0.022	***	0.125	0.135		-0.652	0.335	
Approximate.AgeUnknown	-0.439	0.023	***	-0.411	0.021	***	0.005	0.129		-1.007	0.324	**
AC.YN-NoCentralAir	-0.144	0.008	***	-0.122	0.007	***	-0.117	0.087		-0.594	0.223	**
Basement-Unfinished	0.243	0.051	***	0.210	0.043	***	0.211	0.500		1.041	1.189	
Basement-Partial	0.265	0.051	***	0.228	0.044	***	0.300	0.505		1.307	1.195	
Basement-Finished	0.291	0.051	***	0.245	0.044	***	0.292	0.497		1.331	1.179	
Baths.Full	0.071	0.006	***	0.073	0.005	***	0.013	0.069		0.213	0.170	
Baths.Half	0.069	0.007	***	0.060	0.006	***	-0.047	0.059		0.032	0.150	
Beds.Total	0.017	0.004	***	0.009	0.003	**	-0.043	0.028		-0.083	0.074	
Garage.Spaces.Number	0.023	0.006	***	0.031	0.006	***	-0.171	0.063	**	-0.345	0.171	*
Heat.FA-ForcedAir	-0.066	0.015	***	-0.036	0.013	**	-0.470	0.113	***	-1.254	0.310	***
HouseType-Semi	-0.081	0.014	***	-0.116	0.012	***	-0.045	0.115		-0.400	0.270	
HouseType-Town	-0.125	0.017	***	-0.150	0.015	***	0.191	0.123		0.102	0.304	
Parking.Spaces.Total	0.013	0.003	***	0.013	0.003	***	0.026	0.033		0.096	0.089	
Room.Count	0.015	0.002	***	0.013	0.002	***	0.000	0.018		0.034	0.044	
Storeys-RaisedBungalow	0.013	0.018		0.023	0.015		0.106	0.165		0.318	0.437	
Storeys-Sidesplit	0.016	0.020		0.034	0.017		0.107	0.172		0.350	0.452	
Storeys-Backsplit	0.066	0.014	***	0.056	0.012	***	-0.030	0.102		0.065	0.236	
Storeys-1.5	0.008	0.010		0.012	0.009		0.070	0.068		0.204	0.176	
Storeys-2	0.042	0.010	***	0.056	0.009	***	0.048	0.068		0.258	0.178	
Storeys-2.5	0.135	0.014	***	0.112	0.014	***	0.016	0.080		0.317	0.199	
Storeys-3	0.197	0.072	**	0.161	0.063	*	1.103	0.921		3.130	2.333	
Storeys-Other	0.163	0.089		0.099	0.077		1.269	1.184		3.388	2.961	
lnlotareasqm	0.184	0.008	***	0.171	0.007	***	0.006	0.047		0.438	0.117	***
PCT0TO14	-0.501	0.089	***	-0.053	0.089		-0.501	0.394		-1.371	0.964	
PCT65UP	0.126	0.053	*	0.129	0.053	*	-0.323	0.204		-0.479	0.470	
PCTAPT	0.162	0.020	***	0.049	0.020	*	0.069	0.078		0.293	0.175	
PCTIMMIG	0.025	0.037		0.026	0.035		-0.013	0.132		0.032	0.314	
PCABRGID	-0.170	0.076	*	-0.075	0.067		0.772	0.317	*	1.727	0.787	*
PCVISMIN	0.140	0.035	***	-0.005	0.033		0.356	0.147	*	0.870	0.375	*
PCMAJRPR	-0.243	0.038	***	-0.073	0.034	*	-0.366	0.163	*	-1.086	0.399	**
PC2564HS	-0.379	0.031	***	-0.028	0.031		0.141	0.121		0.281	0.311	
PCJWACTV	0.132	0.029	***	0.037	0.026		-0.032	0.152		0.012	0.378	
PCMOV1Y	0.143	0.038	***	0.010	0.035		0.363	0.169	*	0.923	0.439	*
UNEMPRAT	-0.001	0.000	*	0.001	0.000		-0.001	0.002		0.000	0.005	
lnAVGATHHINC	0.248	0.016	***	0.120	0.015	***	0.093	0.075		0.528	0.170	**
G3	0.044	0.005	***	-0.020	0.007	**	0.060	0.016	***	0.099	0.034	**
G6	0.027	0.006	***	0.009	0.010		-0.032	0.019		-0.056	0.035	
G9	0.079	0.006	***	0.056	0.014	***	-0.055	0.020	**	0.003	0.035	
PPdist	-0.034	0.009	***	-0.010	0.011		-0.014	0.030		-0.058	0.065	
PMdist	0.004	0.007		-0.009	0.010		0.007	0.022		-0.005	0.041	
PHdist	-0.009	0.003	**	-0.003	0.007		-0.012	0.011		-0.036	0.017	*
Quarter2	0.042	0.009	***									
Quarter3	0.062	0.009	***									
Quarter4	0.042	0.009	***									
rho				0.596	0.051	***						

Table 3.7: Spatial and non-spatial regression results, 2016

Variable	base regression			spatial coefficients			spatial lag coefficients			total spatial effects		
term	β	s.e.	p	β	s.e.	p	β	s.e.	p	$\frac{\delta y}{\delta x_k}$	s.e.	p
(Intercept)	8.949	0.200	***	0.963	0.953							
Approximate.Age0TO5	-0.086	0.023	***	-0.072	0.022	***	0.382	0.137	**	0.639	0.279	*
Approximate.Age6TO15	-0.211	0.021	***	-0.176	0.019	***	0.124	0.140		-0.108	0.289	
Approximate.Age16TO30	-0.300	0.022	***	-0.264	0.020	***	0.075	0.131		-0.388	0.271	
Approximate.Age31TO50	-0.300	0.021	***	-0.279	0.020	***	0.206	0.141		-0.151	0.291	
Approximate.Age51TO99	-0.292	0.021	***	-0.288	0.020	***	0.146	0.136		-0.294	0.279	
Approximate.AgeUnknown	-0.330	0.021	***	-0.312	0.019	***	0.187	0.139		-0.257	0.283	
AC.YN-NoCentralAir	-0.076	0.006	***	-0.066	0.006	***	-0.088	0.065		-0.316	0.126	**
Basement-Unfinished	0.028	0.026		0.023	0.024		-0.572	0.246	*	-1.130	0.508	*
Basement-Partial	0.033	0.027		0.021	0.025		-0.556	0.240	*	-1.102	0.500	*
Basement-Finished	0.053	0.027	*	0.044	0.025		-0.601	0.247	*	-1.146	0.513	*
Baths.Full	0.064	0.005	***	0.062	0.004	***	-0.006	0.040		0.116	0.084	
Baths.Half	0.054	0.006	***	0.060	0.005	***	-0.062	0.055		-0.004	0.116	
Beds.Total	0.016	0.003	***	0.011	0.003	***	0.038	0.024		0.100	0.051	
Garage.Spaces.Number	0.026	0.005	***	0.031	0.004	***	-0.074	0.038		-0.088	0.081	
Heat.FA-ForcedAir	-0.039	0.012	**	-0.010	0.011		-0.383	0.103	***	-0.811	0.212	***
HouseType-Semi	-0.114	0.011	***	-0.132	0.010	***	0.064	0.067		-0.141	0.142	
HouseType-Town	-0.173	0.013	***	-0.190	0.012	***	0.214	0.088	*	0.049	0.190	
Parking.Spaces.Total	0.008	0.003	**	0.010	0.002	***	-0.081	0.023	***	-0.147	0.047	**
Room.Count	0.014	0.001	***	0.014	0.001	***	0.012	0.012		0.053	0.024	*
Storeys-RaisedBungalow	0.031	0.013	*	0.029	0.012	*	0.039	0.114		0.142	0.252	
Storeys-Sidesplit	0.037	0.016	*	0.039	0.014	**	-0.418	0.136	**	-0.780	0.295	**
Storeys-Backsplit	0.040	0.011	***	0.047	0.010	***	-0.062	0.066		-0.030	0.141	
Storeys-1.5	0.007	0.008		0.007	0.007		0.053	0.043		0.125	0.090	
Storeys-2	0.052	0.007	***	0.053	0.007	***	0.028	0.051		0.168	0.101	
Storeys-2.5	0.135	0.011	***	0.129	0.011	***	-0.028	0.059		0.207	0.126	
Storeys-3	0.101	0.031	***	0.104	0.028	***	-0.564	0.284	*	-0.946	0.614	
Storeys-Other	-0.011	0.054		-0.041	0.049		-0.520	0.629		-1.154	1.305	
lnlotareasqm	0.123	0.006	***	0.115	0.005	***	0.084	0.034	*	0.409	0.075	***
PCT0TO14	-0.246	0.074	***	0.058	0.078		-1.143	0.279	***	-2.233	0.578	***
PCT65UP	0.069	0.046		0.112	0.046	*	-0.291	0.215		-0.370	0.466	
PCTAPT	0.243	0.016	***	0.089	0.017	***	0.158	0.067	*	0.509	0.126	***
PCTIMMIG	0.134	0.039	***	0.088	0.043	*	0.157	0.130		0.504	0.244	*
PCABRGID	-0.429	0.092	***	0.013	0.092		0.232	0.398		0.505	0.784	
PCVISMIN	-0.053	0.033		-0.089	0.035	*	0.218	0.136		0.266	0.271	
PCMAJRPR	-0.219	0.051	***	-0.103	0.051	*	0.162	0.245		0.120	0.513	
PC2564HS	-0.404	0.027	***	-0.040	0.030		0.092	0.099		0.108	0.208	
PCJWACTV	0.210	0.032	***	0.032	0.035		0.283	0.124	*	0.649	0.250	**
PCMOV1Y	0.072	0.043		0.071	0.041		-0.149	0.156		-0.159	0.323	
UNEMPRAT	-0.001	0.001		0.000	0.001		-0.007	0.002	**	-0.016	0.005	**
lnAVGATHHINC	0.280	0.017	***	0.170	0.018	***	0.258	0.081	**	0.882	0.156	***
G3	-0.008	0.005		-0.007	0.007		-0.035	0.013	**	-0.086	0.021	***
G6	0.029	0.005	***	0.006	0.008		0.013	0.014		0.039	0.020	*
G9	0.053	0.005	***	0.041	0.014	**	-0.024	0.018		0.036	0.020	
PPdist	-0.016	0.006	*	0.008	0.008		-0.056	0.019	**	-0.098	0.036	**
PMdist	0.003	0.005		0.000	0.008		-0.010	0.016		-0.019	0.024	
PHdist	-0.011	0.002	***	-0.005	0.006		-0.009	0.008		-0.028	0.009	**
Quarter2	0.086	0.007	***									
Quarter3	0.129	0.007	***									
Quarter4	0.159	0.007	***									
rho				0.514	0.053	***						

3.6 Discussion

It is necessary at this point to note some caveats. Schools do come with extra amenities that do not necessarily disappear if the school closes; in other words, it is possible be that the effect that we found is not just the capitalization for the school itself, but also the capitalization of the public appurtenances that go along with a school - its surrounding park and playground, for example, or the historical and architectural value of the school building. Again, as noted in the introduction, if a park and playground is a valued good, and if it is socially provided, then we should expect that we might find homebuyers paying a premium simply to be located near a park and playground. A simple distance-to-school analysis is completely unsuited to the task of teasing out the differential effects of school access and park access: it may be that part of what these regressions are picking up is accessibility of school *and* related amenities.

Nevertheless, given that a school closure will typically result in most houses being reassigned to catchments for other, further-away schools, the fact remains that if our regression yields a statistically significant negative coefficient for the school distance variable then we are justified in suspecting that school closure *possibly* has *some* negative effect on house prices by affecting accessibility.

As well, our positive finding in a time-static model says nothing about market reaction over time: it may be that house prices adjust to a school closure over several years, as the closure of a school initiates a slow process of re-sorting resulting in a lower-than-expected price reduction as new residents who are ambivalent to school distance take advantage of the neighbourhood's new school inaccessibility discount.

For these reasons, it would be valuable to analyze property values using a longitudinal data set with observations before and after school closures; this will be the task of the next chapter.

4 Treatment effect of school closures

4.1 Method

What if we perform another multivariate regression of house prices, but change to a time series format similar to what we used to construct a hedonic price index in Chapter 2, including census data for neighbourhood variables, and we also add a categorical variable that flags whether a house's local school is being closed? The regression would then be of the form

$$P = X_i\beta + N_i\gamma + t_i + C_i + \epsilon_i,$$

where P_i is the price of house i, X_i is the vector of house i's structural characteristics, N_i is a vector of house i's local census variables, t_i is a time index for the year and quarter in which house i transacts, and C_i is a second time-index categorical variable which specifies the relationship between the time that house i is sold and the time when its local school is closed, as illustrated in Table 4.1.

Using this categorical variable C in a dataset of all houses sold over several years, we may have sufficient n to be able to determine a generalized school closure price effect on its catchment's houses over time; using $C =$"none" as the reference, the estimated value for each C category will be the estimated price premium or discount for buying a house in a catchment whose school is closing, is going to close, or which has closed some quarters before or after the transaction has occurred. This categorical variable will then allow us to see the effect of a school closure on house prices over time - though the precision of the estimate will critically depend on the value of n for each category.

4.2 Primary school closures in urban Hamilton from 2005 to 2017

In order to determine the sample area for analysis, we should first consider which schools have closed in Hamilton during the available data period. The following information is available from local newspaper reports via Nexis Uni, GIS files of HWDSB catchment data, images available on Google Street View,

Table 4.1: Meaning of C_i categorical variable

C	Definition
none	house's school is not closed
Qminus3	house's school is closed 3 quarters after sale
Qminus2	house's school is closed 2 quarters after sale
Qminus1	house's school is closed 1 quarter after sale
Q0	house's school is closed during the quarter of sale
Qplus1	house's school is closed 1 quarter before sale
Qplus2	house's school is closed 2 quarters before sale
Qplus3	house's school is closed 3 quarters before sale
etc.	etc. (extended for as many quarters as possible)

the HWDSB archives at Hill Park, and old HWDSB website pages preserved at the Internet Archive. Closed schools outside the urban contiguous residential area of metro Hamilton have been omitted.

Maple Lane and **Grange (2005)**: These two Ancaster primary schools (Maple Lane was JK-2, and Grange was grades 3-6) were both recommended in 2002 for closure and replacement with a new JK-8 school servicing their catchment, eventually to be named Ancaster Meadow. Memorial School in Ancaster, the grades 7-8 school for this area, was also identified for closure under this plan. The Grange school building still stood as of June 2009, but was demolished in 2010; eventually the City of Hamilton purchased the property from the board, preserving 2.4 acres as passive parkland and selling a 1.93 acre section to Schuit Homes, who began building high-end homes on their part of the site (Werner, 2011, 2013). Maple Lane had formerly been the Glenwood school for students with special needs; its first classes as a general public school had been in September 1999. After closure, Maple Lane continued to be used by the HWDSB as a board resource and training centre; in 2013, the land was earmarked to be sold to a developer, and in 2018 a draft plan of subdivision and zoning amendment was approved for the developer to build single-family homes. (Leitner, 2013; Werner, 2018; Wheeler, 1999)

Gibson (2006): initially recommended for closure in 2002 (Bishop, 2002a), Gibson absorbed the student body of Robert Land school after that school's closure in 2004 (Cox, 2004). Gibson itself was eventually eliminated as a separate administrative entity in September 2006; all its students were transferred to the new Cathy Wever school that had been built beside Sanford school. However, Gibson school remained in operation until 2009 as a temporary holding school for the JK-3 students of Prince of Wales, whose own school was being demolished and rebuilt. Eventually, the impressive three-storey building was purchased by developer Harry Stinson to be converted into condominium apartments (Arnold, 2016); the school lot had been declared surplus by the city, as there was no parkland or other amenities associated with the property.

Sanford (2006): The existing JK-8 Sanford school was identified by the HWDSB for closure in 2002; a replacement school, named Cathy Wever, was built on the playground of Sanford, and took in all existing Sanford students in September 2006. For several years afterwards, however, the Sanford building remained in use as a temporary holding school for students displaced during rebuilds of other schools. (Faulkner, 2006a)

Prince of Wales (2006): This JK-8 school (built 1922) was closed in June 2006, and its students went to Gibson (JK-3) and Sanford (4-8) as holding schools until the new Prince of Wales building was completed; however, the Prince of Wales student body also included a number of students who lived in the over-capacity catchment for Dr. J. Edgar Davey school, and it is known that some of its catchment's students transferred instead to the new Cathy Wever school. For a month in September 2006, the Prince of Wales building was used as a temporary holding school for students of Hillcrest and Hillsdale whose own replacement building's opening was delayed. It was eventually demolished and a new school built on the same lot; the new Prince of Wales school finally opened to students on 25 March 2009. (Faulkner, 2006b, 2007c; Kruchak, 2006)

Stinson and **Queen Victoria (2006)**: Both schools were already identified for closure by superintendent Jim Murray by 2003; originally a new school was going to be built on the site of the old Stinson school, but that year parents lobbied the Board to build the new school on the more centrally located Queen Victoria site (Cox, 2003a, 2003b). In June 2006 Queen Victoria was closed, to be demolished that fall to make way for a new JK-8 school on the same lot; during the long period of construction, JK-3 Queen Victoria and Stinson students were temporarily housed in the Stinson building, while their grades 4-8 students were temporarily housed in the Sanford school building that had been emptied when Sanford students began going to the new Cathy Wever school. (There are some press mentions of some JK-3 Queen Victoria and Stinson students also being temporarily housed at Gibson school.) The new JK-8 Queen Victoria school was meant to open in September 2008, but delays meant that students remained in holding schools until March 2009 when the new Queen Victoria building opened. (Faulkner, 2006d, 2006f, 2008)

Hillsdale (2006): Hillsdale (JK-5) and its middle school Hillcrest (Grades 6-8) were both identified for closure by 2003; a new, combined JK-8 school was built on the Hillcrest and Hillsdale site while the existing schools continued to be occupied through June 2006. Due to construction delays and the scheduled demolition of the old Hillcrest in July 2006 and the old Hillsdale during September 2006, during all of September and the first half of October all students had to temporarily be housed in the old Prince of Wales school on Lottridge until the new building could be occupied. The new combined JK-8 school kept the combined primary and middle school catchments of the old schools. (Cox, 2003a; Faulkner, 2008; Hamilton-Wentworth District School Board, 2021)

Vern Ames (2007): Vern Ames primary school's middle school, Lawfield, was damaged by an F1 tornado on 9 Nov 2005. The HWDSB rebuilt Lawfield as a JK-8 school, in the same park as the original Lawfield and Vern Ames; when the newly-constructed JK-8 Lawfield school opened in November 2017, it absorbed the JK-5 cohort from Vern Ames school, which then closed. (Faulkner, 2006e, 2007b; Fragomeni, 2007)

Central Park, **Dundana**, and **Dundas District (2007)**: Central Park, a JK-6 school, and Dundas District, a grades 7-8 school, were two Dundas schools originally meant to be closed by the start of the 2006-7 school year and replaced with a new JK-8 school built elsewherre in the Central Park catchment. This new school, eventually named Sir William Osler, was also to take over the north catchment of Dundana school (the area surrounding Veterans Park). Ultimately, Central Park and Dundas District had to remain open until 12 November 2007, as construction of the new Osler school was delayed by a failed property purchase and a labour dispute during construction. This reorganization occurred while the HWDSB was also considering the closure of one or both of Dundas' highschools, Highland and Parkside. (Faulkner, 2007d, 2007a; Pona, 2007)

King George (2012): In April 2011, the HWDSB presented a recommendation to the King George Accommodation Review Committee to close King George school, and transfer its JK-6 students to Prince of Wales and Memorial, partially due to an estimate of $3.5 million in improvements being required for the King George building; King George had only received $19,917 in capital improvements between 2000 and 2010. Trustees voted unanimously to approve the closure in February 2012, and reviewed the boundaries to more evenly distribute students: Queen Mary was added as a third neighbouring school to absorb other King George students. By June of that year, on the school's 100th birthday, the school was emptied of students; it began to be used as accommodation for HWDSB facilities maintenance staff formerly housed at Crestwood, while that building was demolished to make way for a new HWDSB headquarters (Pecoskie, 2011, 2012d, 2012a). The HWDSB had originally intended to demolish King George, and the neighbouring Parkview special-needs highschool, and use that site for the location of a new highschool, if the demolition of Scott Park Highschool didn't go through; but city councillors blocked this plan by adding King George to the heritage registry (Pecoskie, 2013c, 2013a). The King George building remained standing as of 2020.

Prince Philip (2014): In April 2012, trustees voted to close this school and move all its students to G.R. Allan, another JK-5 school in Westdale by September 2013; however, delays in completing upgrades to G.R. Allan meant that the closure of Prince Philip was delayed until the summer of 2014 (Leitner, 2012; Pecoskie, 2012c). In February 2015, the property was sold to Hamilton's French school board, Conseil Scolaire Viamonde (Dundas Star News, 2015); nevertheless, by 2020 the building was demolished to add to the greenspace of Alexander Park.

Eastmount Park and **Linden Park (2015)**: Linden Park school was originally recommended for closure in 1999, but was not closed for over a decade (Fisher, 2000). Trustees voted in June 2014 to close Eastmount Park (built in 1959) and Linden Park (built in 1957, on the same lot as Hill Park highschool), as well as their middle school at Cardinal Heights (grades 6-8). (Their highschool, Hill Park, was recommended for closure separately.) Eastmount Park school's catchment was completely absorbed by George L. Armstrong school; Linden Park's catchment was split between Franklin Road (JK-8) and Queensdale (which was JK-5, but became JK-8) schools (Leitner, 2014; Nolan, 2014; Reilly, 2015). Eastmount Park stood empty after being bought by the city in August 2016 (Hamilton Mountain News, 2016); it was torn down in 2020, and as of March 2021 the city was contemplating selling off parts of the property to pay for improvements to the surrounding park (Hamilton Mountain News, 2021).

Roxborough Park and **Woodward (2015)**: The ARC plan for east Hamilton originally also called for the closure of Parkdale and Rosedale schools, and the replacement of Viscount Montgomery with a brand new JK-8 school; however, the board trustees instead voted in June 2014 to only close Roxborough Park (built in 1960) and Woodward (built in 1952), raising $4 million to put toward capital improvements at the remaining east-end schools. (Trustee Judith Bishop noted this would be insufficient to update the remaining schools, since there were $17 million in repairs needed at Viscount Montgomery, Parkdale and W.H. Ballard alone; but the board voted to step back from its original aggressive plan, to reduce the impact of closures on the neighbourhood.) Woodward and Roxborough Park both closed in the summer of 2015; Woodward's catchment was completely absorbed by Hillcrest school, while Roxborough Park's catchment was split between Hillcrest, Parkdale and Viscount Montgomery schools. (Leitner, 2014; Reilly, 2015) Roxborough Park remained standing as of early 2017, but was demolished in early 2018 to make way for apartments and social housing units; its greenspace remained as of 2020. Woodward was demolished in September 2019, to make way for a townhouse development; its neighbouring park, however, presently remains.

4.3 Method

4.3.1 Selection of the dataset

A glance at the above list of closures identifies two different scenarios that occur:

- In some cases (Sanford, Prince of Wales, Hillsdale, Queen Victoria, Maple Lane/Grange, Central Park/Dundana/Dundas District), an existing school is "closed," but replaced by a new school on the same site; while it may be enlightening in the future to study whether major capital improvements to schools have an effect on house prices, school accessibility itself does not change in this set.
- In other cases (Gibson, Stinson, King George, Prince Philip, Eastmount Park, Linden Park, Roxborough Park and Woodward), an existing school is closed and its students are transferred to another school outside the closed school's catchment. In such a case, the distance to school definitely changes for all houses in that closed school's catchment (although, in a minority of cases, distance to the new school may actually decrease).

For this chapter's study, it was decided to concentrate only on the second scenario: school closures that changed neighbourhoods' "distance to school" measure. Of these closures, five happened in downtown Hamilton between John Street and the Red Hill valley. The closure of the two Mountain primary schools, Eastmount Park and Linden Park, happened at the same time as the closure of that area's highschool, Hill Park: this would introduce bias into a "distance to school has changed due to school closure" categorical variable, since there is also a change in the distance to highschool (with grade 9-12 students generally having to change either to Sherwood, their official catchment highschool, or Barton, which was temporarily kept open for years and was open for enrollment for all former Hill Park students).

We have seen in a previous chapter that distance to highschool is also valued, so omitting highschools from our study may introduce bias into the coefficient estimation for primary schools; but including highschools would introduce its own bias, since the only highschools whose catchments close in this period are Hill Park on the central mountain, Parkside in Dundas, and the confusing case of Barton on the east mountain (which spent six years being "closed but not really" while its replacement was being built).

Of course, there is the even more confusing case of Sir John A. Macdonald and Delta highschools, both of which the HWDSB voted to close in 2012 (Pecoskie, 2012b); one could argue the closure of Delta to be a source of bias for estimation of the value of King George primary school, which was closed in 2012 and whose catchment (part of the Delta highschool catchment) forms part of our study. However, the replacement of these two highschools was quickly bogged down: city council's addition of the closed King George to the heritage registry (Pecoskie, 2013a) made the HWDSB unable to demolish it and build a new highschool on that property. Only by March 2015, with the HWDSB's successful defense of a court challenge over the expropriation of the former Scott Park highschool property (itself closed in 2001 and previously sold to a developer), did the board finally own a site where they could proceed to build a new highschool (Nolan, 2015). Sir John A. Macdonald and Delta both continued to operate until summer 2019, when both were finally closed and all students were transferred to the new Bernie Custis Secondary School on the former Scott Park site. Given the fact that prospective homeowners in the King George primary catchment had no information about where their new highschool would be built (or even certainty about when or if their present highschool would be closed, given the HWDSB's high-profile backtracking on the closure of Sherwood highschool on the east mountain in this same period) until many years after the closure of King George, it was decided to ignore any informational effect that the proposed future closure of Delta would have on homebuyers in the King George primary catchment.

The future closure of Delta, parenthetically, would have no impact on the 2015 closures of Roxborough Park and Woodward, since those two primary schools were both part of the Sir Winston Churchill highschool catchment.

However, the closure of Hill Park highschool at the same time as Eastmount Park and Linden Park, combined with the fact that the Mountain has a somewhat different history, social character and urban form from the downtown, suggested that the two mountain primary schools be removed from this study. Including the Mountain would introduce a bias of its own, since its neighbourhoods are different in age, housing and social characteristics: we can try to minimize any possible difference between treatment and control populations by limiting our study to the downtown area. Omitting these two schools suggested that Prince Philip also be omitted: a study can just be done of the downtown contiguous area, and ignore the disconnected Prince Philip catchment in West Hamilton - where, after all, a greater tendency for conversions to student housing may change the reaction of the housing market to school closure, and where the proximity of McMaster University and its impacts on social characteristics may also greatly impact the neighbourhood.

A map showing the locations of the five closed primary school catchments chosen for this chapter's study is shown in Figure 4.1: clockwise from bottom left, the catchments shown are Stinson, Gibson, King George, Woodward, and Roxborough Park.

Thus, it was decided to trim down the RAHB dataset from the full 2005-2017 set to the contiguous residential urban area west-to-east from John Street to the Red Hill Valley, and south-to-north from north of the escarpment and Lawrence Road to the bay. A map showing the 2005-2017 transactions for this chapter's study area is given in Figure 4.2; this area provides just over 19,000 transactions for analysis.

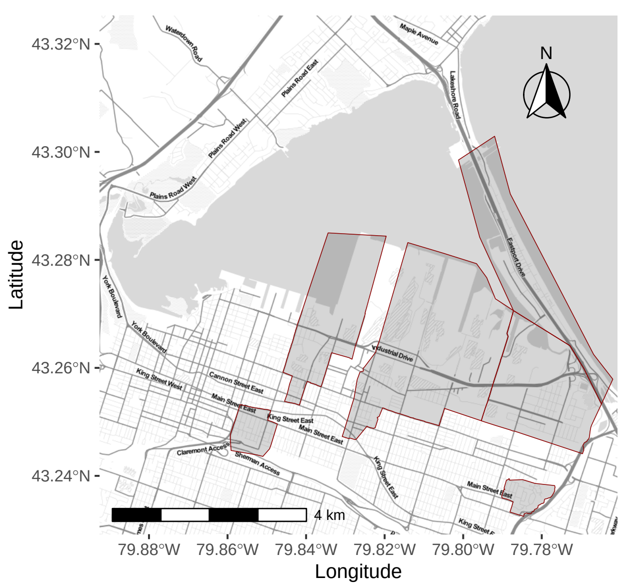

Figure 4.1: Map of closed downtown primary school catchments, 2006-2015

Figure 4.2: Map of study area, all transactions shown

As can be seen in Figure 4.3, the houses inside the downtown study area are significantly different from those elsewhere in the city; for the entire 2005-2017 period, the mean transacted house price in the study area was approximately 48% of that outside the study area. Similarly, Figure 4.4 shows the significant difference in lot areas between Hamilton houses included in the study area and those excluded.

4.3.2 Time categorical variable

An ordered "year and quarter" variable was generated for the study area dataset. This allows for one single regression equation for all years, with a quarterly time index; of course, this means the regression will assume homogeneity of coefficients from 2005 to 2017.

Adding a year-and-quarter variable to our dataset allows us to analyze our dataset in terms of number of houses sold per year and per quarter: this data is presented in Figure 4.5 and Figure 4.6. As can be seen clearly in Figure 4.5, volume of home sales in the east downtown decreased by 13.2 percent in 2008 alone, as the impact of the US financial panic began to be felt; volume reached a total year-over-year volume decrease of 22.4 percent by 2010. Downtown

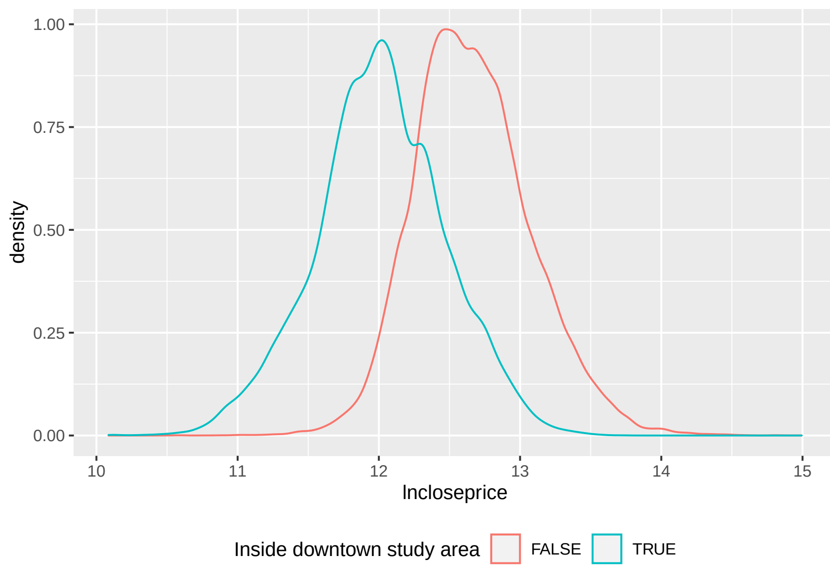

Figure 4.3: Comparative histogram of log price, 2005-2017

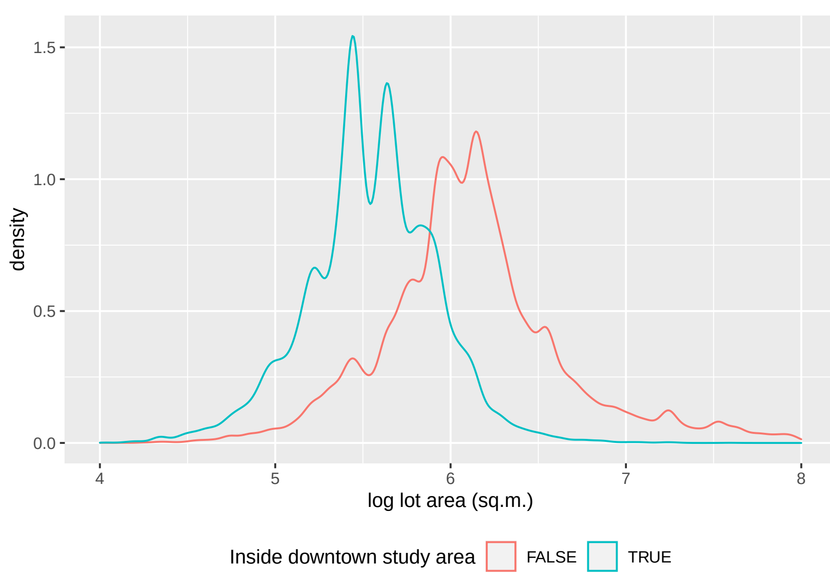

Figure 4.4: Comparative histogram of log lot area

housing sales volumes did not recover to their 2005-2007 levels until 2015, which was a full two years after the recovery in the North American equities indexes.

Analysis of Figure 4.6 shows an obvious periodicity in home sales: Quarter 1 (January to March) always has the lowest volume of sales for any quarter; either Quarter 2 (April to June) or Quarter 3 (July to September) is typically the quarter with highest sales.

4.3.3 Census variables

Census data from 2006-2016 (again at the DA level), as corrected previously in chapter 2, was added to the transactions of our downtown study area. Because our dataset includes transactions in inter-census years, the census data was arithmetically interpolated for each year; 2017 transactions were simply given the 2016 census values, and 2005 transactions were given 2006 census values.

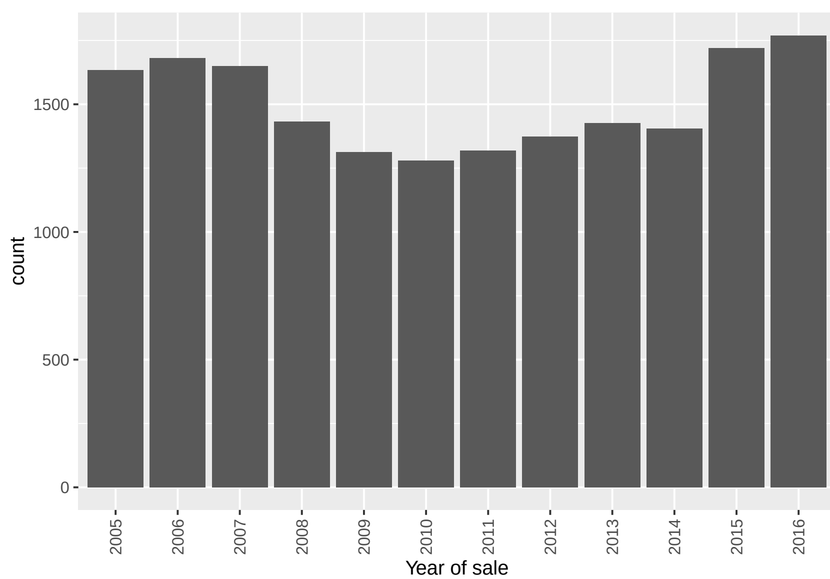

Figure 4.5: Study area home sales by year, 2005 to 2016

4.3.4 School closure categorical time variables

Each house in the 2005-2017 dataset was given a categorical variable as to whether or not it transacted in temporal proximity to a school closure, as follows:

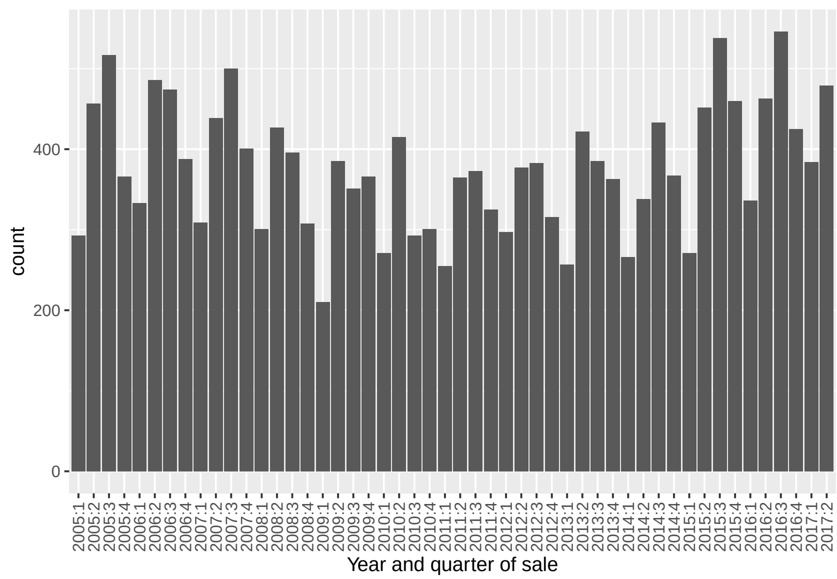

Figure 4.6: Study area home sales by year and quarter, 2005 Q1 to 2016 Q2

- All houses in our dataset initially had their “closetime” variable set to “None,” the default comparator for this categorical variable.
- The time of closure for Stinson and Gibson schools was assumed as 2006 Q3 (i.e. the July-September quarter). All houses in our dataset that were located within the 2005-2006 catchments for Stinson and Gibson, and which transacted from 6 quarters before to 7 quarters after their school’s closure, had their “closetime” variable reset to a value per the format in Table 4.1.
- The time of closure for King George school was assumed as 2012 Q3. All houses in our dataset that were located within the 2011-2012 catchment for King George, and which transacted from 6 quarters before to 7 quarters after this school’s closure, had their “closetime” variable reset to a value per the format in Table 4.1.
- The time of closure for Woodward and Roxborough Park schools was assumed as 2015 Q3. All houses in our dataset that were located within the 2014-2015 catchment for these schools, and which transacted from 6 quarters before to 7 quarters after the closures, had their “closetime” variable reset to a value per the format in Table 4.1.

This yields a dataset of downtown study area houses, transacted 2005-2017, where those houses which transact in a primary school catchment about to close or recently closed are identified with a separate time-based categorical variable that may allow us to see the specific price effect of an imminent or recent school closure on these houses' prices relative to all the other houses transacting that quarter.

This series of closure categoricals is extended several quarters before closure to attempt to investigate whether the simple announcement of a future closure can affect house prices. The initial announcement of an Accommodation Review investigating King George's closure came in April 2011, five quarters before its closure in September 2012; the announcement of an Accommodation Review for Woodward and Roxborough Park was made in June 2014, again five quarters before their closures: it would be remarkable, then, if there were any price effect found more than 5 quarters before the closures of these three schools, since that information would not be generally available for home buyers.

Stinson and Gibson do complicate this: as noted, both were selected for closure already in 2003, and their closures seem to only have been held up until 2006 by limitations in sequencing downtown new school construction. However, RAHB dataset limitations pre-2005, identified in chapter 2, and the lack of HWDSB school catchment maps before 2005, together force the analysis to be limited to the period 2005 forward.

The time-based closure categorical variable is extended to seven quarters after closure in order to see if the negative price effect of a school closure tapers off as time goes on. This should certainly be a possibility, since school accessibility is a good that may not valued by every market participant. So, the loss of school accessibility in one area may make its homes less attractive to prospective homebuyers with children, or at least homebuyers who want their children to walk to school; this reduction in demand for part of the buying population would make home prices drop to clear the market. But then, with these homes now selling at a discount to other homes, they may become more attractive to other homebuyers - those who prefer Catholic primary schools, those whose children have already finished grade 5, those who drive their children to school, or even those buyers who have no need for school accessibility whatsoever.

Basically, one could reasonably assume that the stochastic shock of a reduction of school accessibility would wear off over time; it would be ideal to extend this time-based closure categorical variable beyond seven quarters, but unfortunately this dataset of home transactions ends in July 2017.

4.3.5 School closure delta distance

In addition, for those houses whose school closes from six quarters after to seven quarters before the transaction, a change in distance to primary school was calculated, for later analysis: it may be that houses who see their catchment primary school close, but who were nearer to another school anyway, would not see a drop in prices. (E.g., all Gibson students were moved to Cathy Wever school: for a fair number of them, they actually saw a *decrease* in the distance they had to walk to get to school.)

This, of course, is complicated by the complexities of school closure. For example, while all the houses in the closed Stinson primary catchment become Queen Victoria houses in September 2006, the students in that catchment would temporarily still go to Stinson for grades JK-3 and then Sanford for grades 4-8, until the new JK-8 Queen Victoria school opened in March 2009: so, their "change in distance to school," for a large number of a Stinson catchment child's primary school years, is quite different from the calculated change in distance to the new Queen Victoria.

4.4 Exploratory analysis

The distribution of distance-to-school changes for houses in closed primary catchments can be seen in Figure 4.7: the median change in distance to future primary school for houses which transacted within 6 or 7 quarters of a primary school closure is positive and approximately 500m. 491 homes in our dataset see a decrease in distance to the new primary school; 2888 homes see an increase in distance.

Interestingly, across all closed primary catchments from 2005 to 2017, slightly more houses sell in the four quarters after a closure than in the four quarters before and during a closure, as is illustrated in Figure 4.8. In particular, it is most striking that between the five closed catchments of our study area, there are a total of 89 sales in the June-to-September quarter a year after a closure (i.e. “Qplus4”), and 77 sales in the summer quarter before (“Qminus4”), but only 66 sales in the June-to-September quarter in which the primary school is closed (i.e. “Qnow”): that summer quarter is typically supposed to see the highest sales of the year, so it may be that a primary school closure interferes with houses’ ability to sell in that summer quarter. Then again, the April-to-June quarter before closure (“Qminus1”) sees much higher sales than spring quarters before or after (88 sales, versus 69 and 78); it may be that house sales are pulled forward one quarter to avoid selling during the summer of a school closure.

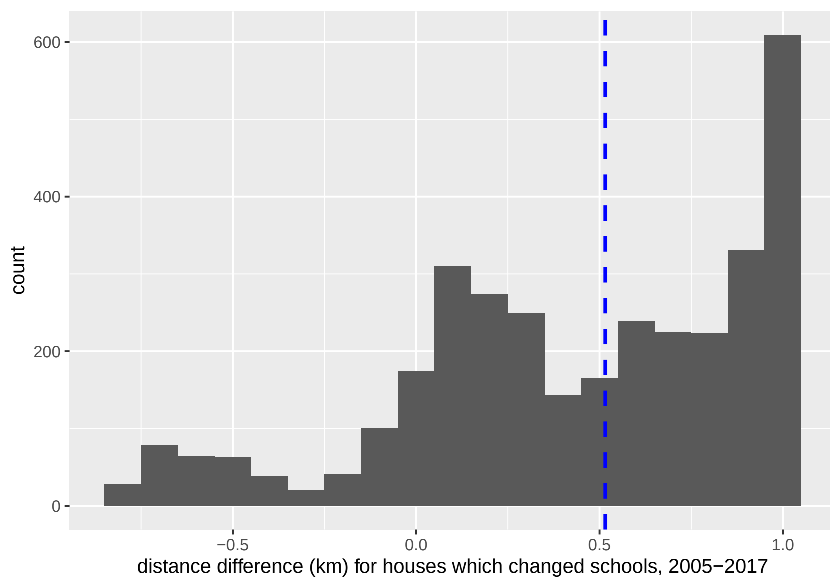

Figure 4.7: Comparative histogram of distance changes

A reason to limit this analysis to the downtown area between John Street and Red Hill is to ensure that we have comparable treatment and control populations: the estimate of categorical variable for the effect of a school closure should be independent of the other unmeasured effects on home prices in the neighbourhood. But are these two downtown populations comparable?

A comparative histogram of house prices (see Figure 4.9) shows this not really to be the case: the treatment group (houses which transact within 6 quarters before or 7 quarters after a school closure) sells for approximately 18% less

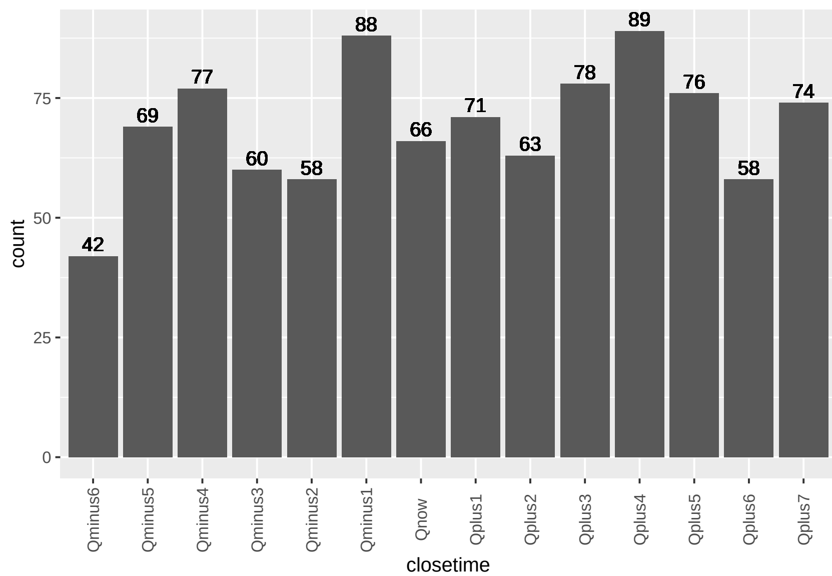

Figure 4.8: Closure catchment home sales by quarter relative to school closure, 2005 to 2017

than the control group.

This may be due to the different social characteristics of the neighbourhoods: transaction-level comparative histograms of social characteristics (Figure 4.10) indicate that the neighbourhoods that experienced school closures had (during the closure period) relatively lower income, higher unemployment rate, higher percentage of visible minority, and higher percentages dwelling in apartments, compared to the rest of the study area. This may be expected, given that a glance at Figure 4.11 shows most houses in our treatment dataset are north of Main Street, where social indicators in Hamilton are generally lower; in fact, Figure 4.1 shows that almost all the area north of Barton Street saw catchment closures. The north-south gradient of social characteristics in downtown Hamilton will thus still be captured in our study: and we know from earlier chapters that positive social characteristics are generally positively valued in house price.

This indicates, then, an important limitation to this chapter's results: since our treatment population is not a representative sample of the total population, there will be some bias in our results. Then again, this should be expected in any spatial natural experiment.

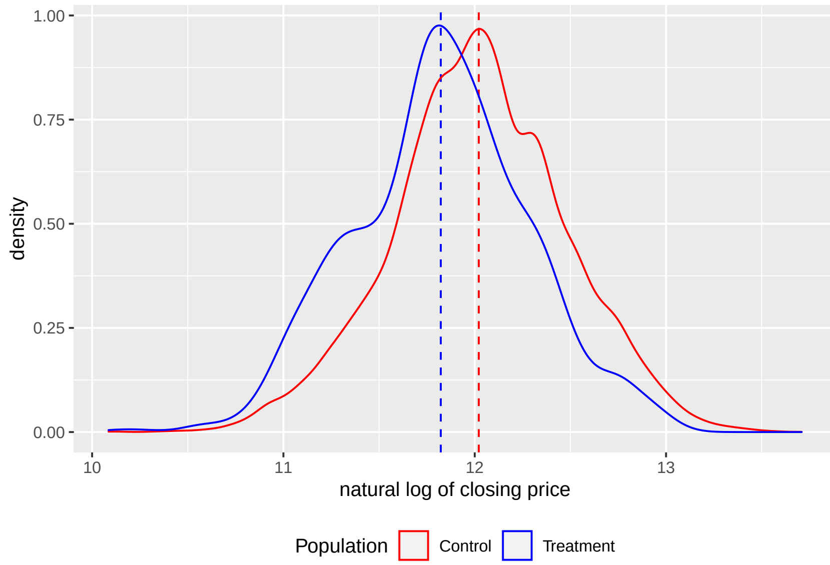

Figure 4.9: Closure catchment home sales by quarter relative to school closure, 2005 to 2017

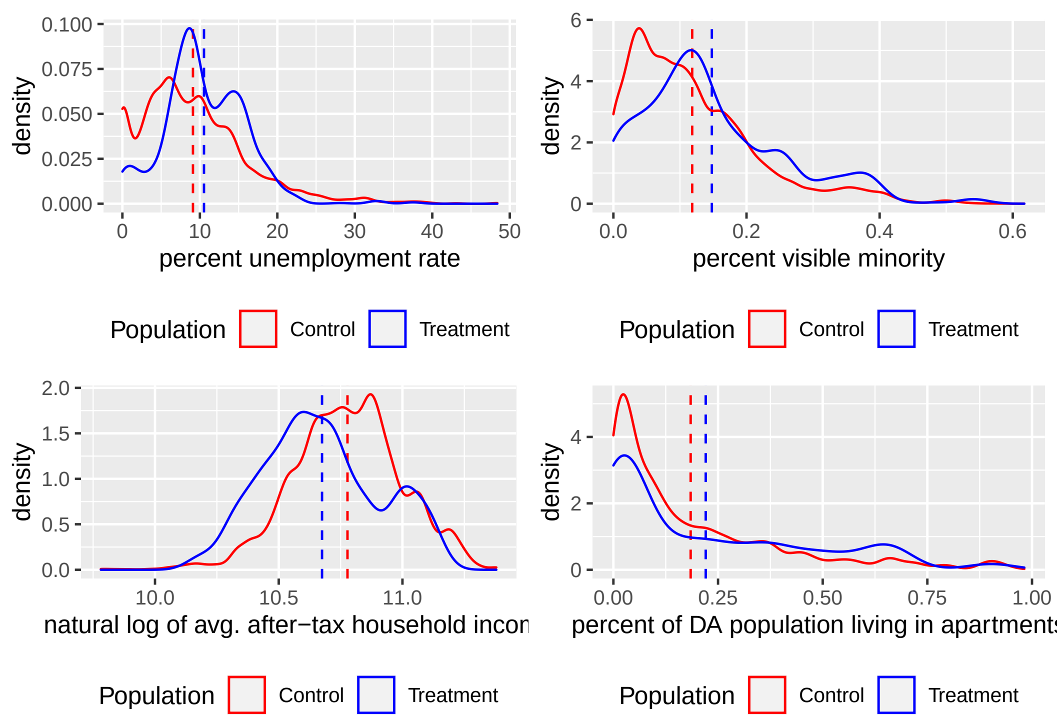

Figure 4.10: Comparison of social characteristics

4.5 A non-spatial time-series hedonic regression

The above work provides a dataset of all house transactions in downtown Hamilton which occur from January 2005 to July 2017; the houses which transact within a few quarters of a primary school closure are flagged with a categorical variable that identifies that transaction's temporal relation to a school closure, allowing us to add that categorical to a regression to look for statistical significance. A map of all the house transactions in this dataset that take place from six quarters before to seven quarters after a primary school closure, with change in distance to school, is presented in Figure 4.11.

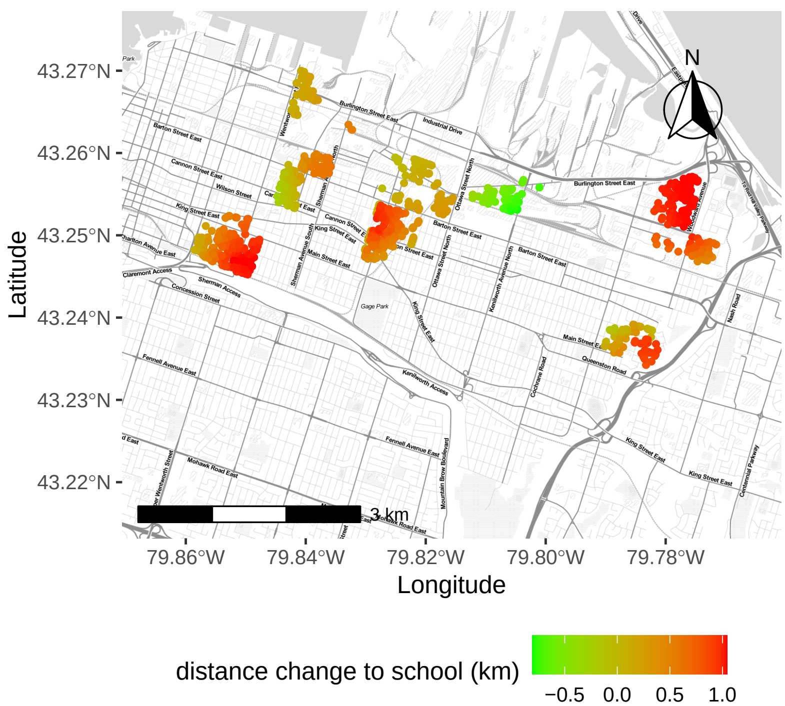

Figure 4.11: Map of closed catchment transactions (Qminus6 to Qplus7) shown

A simple hedonic regression - including a year and quarter variable for house price appreciation, a variable for the change in distance to primary school for schools in closed primary catchments, and a quarterly closetime variable for the houses in our treatment group - was performed. The estimated coefficients for housing (omitting estimates for house form-factor coefficients) are presented

Table 4.2: Non-spatial regression results

term	β	s.e.	p
(Intercept)	7.290	0.143	***
Approximate.Age0TO5	-0.123	0.044	**
Approximate.Age6TO15	-0.288	0.038	***
Approximate.Age16TO30	-0.316	0.041	***
Approximate.Age31TO50	-0.419	0.034	***
Approximate.Age51TO99	-0.459	0.032	***
Approximate.AgeUnknown	-0.503	0.032	***
AC.YN-NoCentralAir	-0.121	0.004	***
Basement-Unfinished	0.109	0.017	***
Basement-Partial	0.156	0.017	***
Basement-Finished	0.211	0.017	***
Baths.Full	0.040	0.003	***
Baths.Half	0.048	0.004	***
Beds.Total	0.004	0.002	*
Garage.Spaces.Number	0.011	0.003	***
Heat.FA-ForcedAir	-0.048	0.006	***
House.Type1-Semi	-0.112	0.007	***
House.Type2-Town	-0.149	0.012	***
Parking.Spaces.Total	0.020	0.002	***
Room.Count	0.016	0.001	***
lnlotareasqm	0.228	0.005	***
PCT0TO14	-0.333	0.057	***
PCT65UP	0.198	0.046	***
PCTAPT	0.322	0.012	***
PCTIMMIG	0.044	0.025	
PCABRGID	-0.136	0.039	***
PCVISMIN	-0.183	0.025	***
PCMAJRPR	-0.142	0.022	***
PC2564HS	-0.510	0.017	***
PCJWACTV	0.151	0.017	***
PCMOV1Y	-0.094	0.025	***
lnAVGATHHINC	0.311	0.012	***
UNEMPRAT	-0.002	0.000	***
diffdist	0.053	0.006	***

Note:

R-squared = 0.77 ; F statistic = 622.14

*** $p < 0.001$

** $0.001 < p < 0.01$

* $0.01 < p < 0.05$

in Table 4.2. A simple comparison with all three city-wide census-year results of the first regression in chapter 3 finds the following:

- this downtown sample area sees a somewhat higher price offered for parking spaces and lot area;
- a higher (positive) price is paid for percent of DA population living in apartments, which may be due to the richer opportunity landscape available in downtown high-density neighbourhoods compared to the non-downtown Hamilton urban area;
- a somewhat higher price is paid for neighbourhoods in this study area with a higher average after-tax household income.

Figure 4.12 is a Pearson correlation matrix presenting correlations between census variables and school distance change, for those houses transacting in the fourteen quarters surrounding and including a school closure: it shows distance difference has a correlation of less than 0.1 with every census variable (and the three above-stated house variables) except percent visible minority. Whatever reason there is for houses in closed catchments to trade at a relatively higher price when the distance from that house to primary school significantly increases, it seems to be orthogonal to the other census and house variables.

The coefficients for the quarterly closetime variable are presented in Table 4.3: it is obvious that there is a statistically significant ($p<0.001$) and arithmetically large price penalty for houses sold during eight of the nine quarters surrounding and including the quarter of a primary school closure. Indeed, houses selling in the fall (October to December) before a primary school closure are estimated to transact at an approximately 14.3% lower price.

In addition, a strange and statistically significant result is found for the distance difference variable: houses in closed primary catchments which see an *increase* in distance to their replacement school are valued *more*. There is a 5.16% increase in price per kilometre of increase in distance to the new school. This price increase is low enough that it does not cause a *net* increase in any house price after a school closure: as noted in Figure 4.7, maximum distance change in our dataset is 1km, so all houses in a closed catchment end up negatively impacted in price. The furthest houses are just impacted less on average.

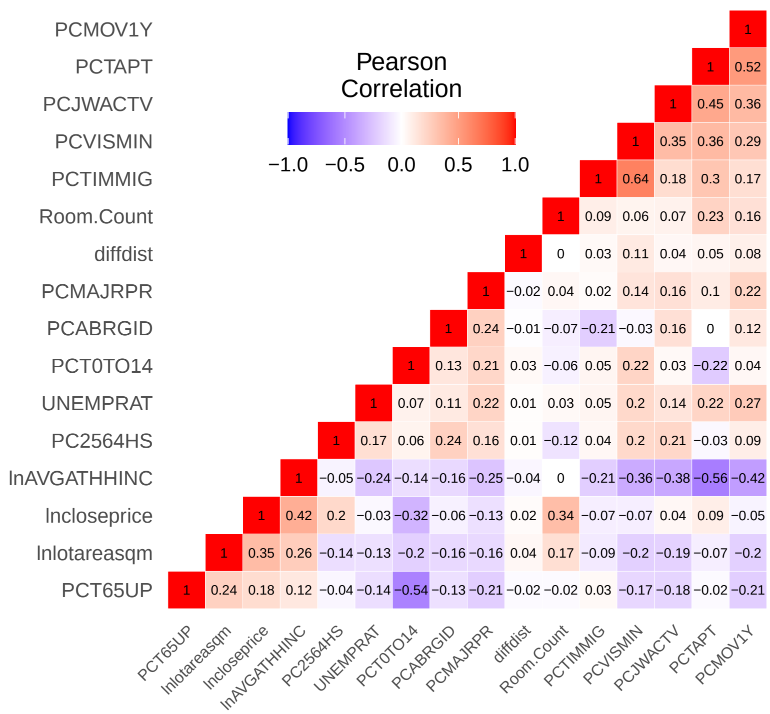

Figure 4.12: Pearson correlation matrix, distance change vs other variables

This result seems not to be due to any correlation with social variables: a regression of distance difference on census variables using the sub-dataset of houses shown in Figure 4.11, while yielding several estimates with statistical significance, only yields a multiple R-squared of 0.032. Addition of the house variables log close price, room count and lot area, all statistically significant variables, still only raises the adjusted R-squared of such a regression to 0.037. Put simply, house value and social variables have only an insignificant correlation to change in distance to school, for houses transacting during the 14 quarters before and after school closure.

Of course, HWDSB primary schools in Hamilton have fairly permeable catchments: a student can go to an out-of-catchment primary school as long as there is space available for them. These school closures were partially motivated in the first place by below-capacity enrolment in the downtown, so there would have been spaces available. One possibility, then, is that houses in closed catchments which see their school distances reduced will trade at a penalty because of loss of choice: the most local school is removed from the choice set of households. A price *increase* for houses which see their school distances increased, however, can't be explained this way: those houses have also lost options from their choice set.

Perhaps, instead, this counter-intuitive result is explained by a shift in buyer type, among a population that is heterogeneous in income and transportation? Assume that higher-income households are more likely to drive their children to school (as suggested by the results summarized in Sirard & Slater (2008)): then, after a closure, perhaps houses that are farther from their new school are more likely to be bought by people with higher incomes who are more likely to drive their children to school anyway, while houses closer to their new school are more likely to be bought by families with lower incomes who are less able to drive their children to school? If higher-income buyers also overbid more for houses, this may explain the sign for the distance-difference coefficient: once a school closure is announced, houses further away from the school begin to be bought by people with higher incomes, who don't care about walking distance to school and who are willing to overbid for houses suffering a price penalty due to the closure. If this trend continues over subsequent years, we might see

Table 4.3: Non-spatial regression results, closetime variable

term	β	s.e.	p
closetimeQminus6	-0.064	0.035	
closetimeQminus5	-0.074	0.027	**
closetimeQminus4	-0.107	0.026	***
closetimeQminus3	-0.154	0.029	***
closetimeQminus2	-0.027	0.030	
closetimeQminus1	-0.101	0.024	***
closetimeQnow	-0.126	0.028	***
closetimeQplus1	-0.131	0.027	***
closetimeQplus2	-0.097	0.029	***
closetimeQplus3	-0.129	0.026	***
closetimeQplus4	-0.139	0.024	***
closetimeQplus5	-0.081	0.026	**
closetimeQplus6	-0.064	0.030	*
closetimeQplus7	-0.087	0.026	***

Note:
*** $p < 0.001$
** $0.001 < p < 0.01$
* $0.01 < p < 0.05$

it in the subsequent ten years of census data.

This hypothesis was investigated using the September 2006 school closures of Gibson and Stinson Street schools. The census dissemination areas which contained all houses transacted from 2005-2007 in these two closed school catchments were selected: the mean change in school distance among transacted houses of each DA was calculated. In addition, each of these DAs had its 2006-to-2016 change in average after-tax household income and change in percentage population 0 to 14 calculated. Then, distance difference was regressed on the delta income and delta child population variables, to find each's marginal "effect." The results are shown in Table 4.4.

Unfortunately, since this regression was conducted at the dissemination area level of support, for only two school closures, this regression suffers from a low n of only 24 dissemination areas (23 with houses, and 1 additional DA

Table 4.4: Distance to school difference: marginal correlates

term	β	s.e.	p
(Intercept)	0.334	0.127	*
deltaincome	0.368	0.445	
deltapopkids	-2.353	2.081	

Note:
R-squared = 0.096 ; F statistic = 1.116
*** $p < 0.001$
** $0.001 < p < 0.01$
* $0.01 < p < 0.05$

with only apartments); coefficient estimates for change in average after-tax household income (positive) and change in population 0-14 (negative) are thus not statistically significant, though it is notable that both are approximately one standard deviation from 0 in the direction that supports our hypothesis. The low R-squared (unadjusted) of 0.096 is not a concern, since ten years of house transactions should not be expected to be highly explanatory for a neighbourhood's change in average after-tax income or child population over those ten years.

It may be, then, that a school closure does result in people with higher incomes (and higher propensity for driving), and people with fewer school-age children, differentially buying the houses that end up further away from the new school; our sample, unfortunately, is not large enough for us to obtain a statistically significant result. Nevertheless, scatterplots of the two relationships, as shown in Figure 4.13, are still persuasive.

4.6 Moving to a spatial model

4.6.1 Scale of spatial autocorrelation

Similar to the previous chapter, in order to do a spatial regression we must first solidify our assumption about the scale of spatial correlation, again by examining the residual variogram: since our full dataset covers transactions from 2005 thru 2017, this variogram will be based on a regression of log prices

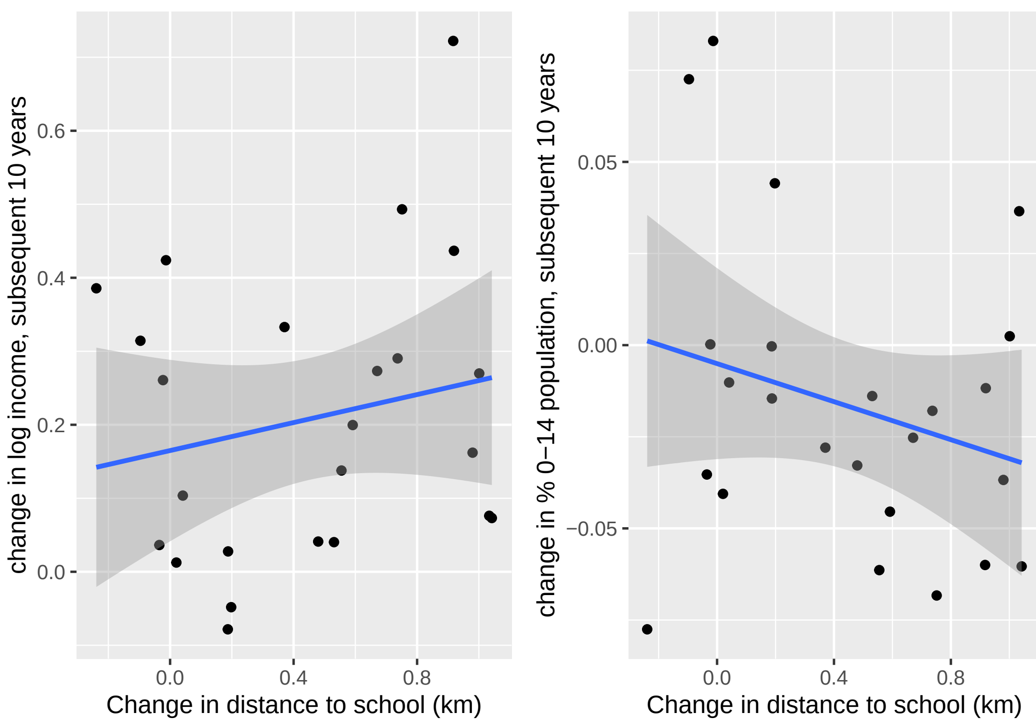

Figure 4.13: Scatterplots of delta distance

detrended to Q1 of 2005. In addition, due to the large number of repeat sales (7202 of 19027 total transactions from 2005 to 2017), the dataset for spatial analysis has been reduced to one single transaction per property, randomly chosen.

The short-distance section of the variogram for our downtown study area is shown in Figure 4.14; it reveals a sill at approximately 120-150m. Within that radius, the average house will have from 27 to 40 single-transaction nearest neighbours; based on the variogram, 30 nearest neighbours is chosen for the scale of spatial correlation.

(A full-distance residual variogram, not shown, suggests a further sill of 0.48 at approximately 2000m; however, since this distance is approximately equal to the total north-south extent of the study area, it can be discounted as simply the effect of the long-distance north-south trend in house prices.)

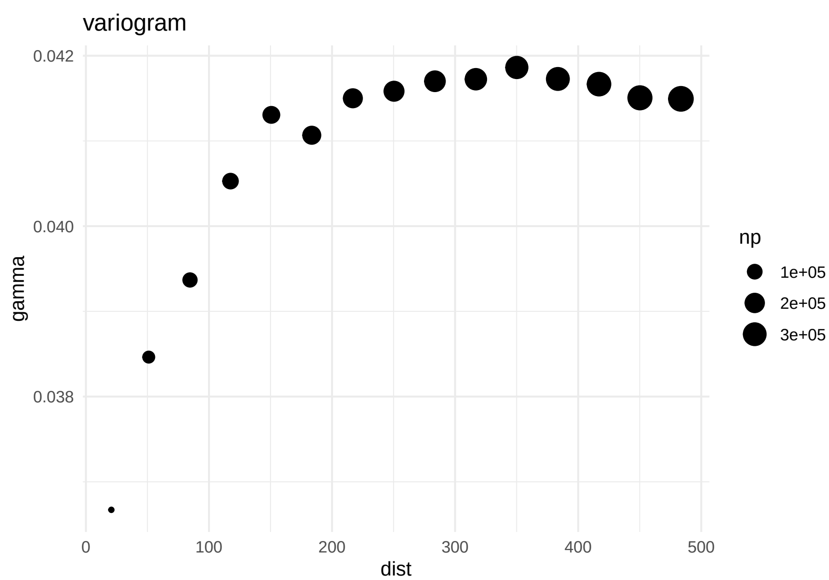

Figure 4.14: Short-distance variogram

A global Moran's I test was performed for the residuals of a regression with the study area dataset's repeat sales randomly removed, and with k=30 for nearest neighbours; Lagrange multiplier diagnostics were also performed, and all results are summarized in Table 4.6. The Lagrange multiplier values are problematic: with values so high, it will be unlikely that either a simple lag-error or lag-Y model would be able to sufficiently mop up residual autocorrelation. In such a case, as in the previous chapter, a spatial Durbin model could be tried; however, this may complicate the interpretation of the quarterly categorical variable for school closure.

4.6.2 Including a census dissemination area spatial dummy variable

Including dissemination area dummy variables is one alternate way to try to account for the fraction of spatial autocorrelation that is due to unobserved neighbourhood effects. This was not a productive path to go down in the previous chapter: since that chapter's variable of interest (distance to school) was very dependent on DA location, the value for school distance's coefficient would be mostly absorbed by the DA dummy variables, and so the distance estimate would be highly biased toward 0. But let it be assumed for now that the quarter-of-school-closure variable is mostly orthogonal to DA location: do DA dummy variables mop up residual spatial autocorrelation in the error term?

The closetime coefficient estimates from a DA dummy variable regression are presented in Table 4.5: it can be seen that, in fact, including DA dummies in this regression does significantly bias each estimate for the closetime variables toward zero, compared to the results in the non-dummy regression of Table 4.3. A chi-squared test of independence performed on the DA dummy variable and the closetime categorical variable yields a result of $p << 0.01$; it seems that this correlation will indeed bias the estimates of interest. Nevertheless, the exercise does still confirm that school closures still do have a statistically significant negative effect on house prices.

The Moran and Lagrange test results for the residuals of a DA dummy regression (with house prices detrended to 2015 Q1 values, and further with repeat sales now removed) are presented in Table 4.6: the results in the right column indicate that including dummies for census DA as a spatial control does seem to eliminate any spatial autocorrelation of the error term. The rise in adjusted R-squared suggests a more explanatory regression, although the degradation of the F statistic caused by including DA dummies is concerning. But in any case, even after including DA dummies, the Lagrange tests still indicate that autocorrelation of Y remains. A spatial regression must be the next step.

Table 4.5: Non-spatial regression results, closetime variable, with DA dummies

term	β	s.e.	p
closetimeQminus6	-0.040	0.032	
closetimeQminus5	-0.076	0.025	**
closetimeQminus4	-0.073	0.024	**
closetimeQminus3	-0.081	0.027	**
closetimeQminus2	-0.009	0.027	
closetimeQminus1	-0.050	0.022	*
closetimeQnow	-0.099	0.025	***
closetimeQplus1	-0.075	0.025	**
closetimeQplus2	-0.054	0.026	*
closetimeQplus3	-0.084	0.024	***
closetimeQplus4	-0.116	0.022	***
closetimeQplus5	-0.043	0.024	
closetimeQplus6	-0.038	0.027	
closetimeQplus7	-0.063	0.024	**

Note:
*** $p < 0.001$
** $0.001 < p < 0.01$
* $0.01 < p < 0.05$

Table 4.6: Tests of spatial autocorrelation, with and without DA dummy

	without DA dummy	with DA dummy
Moran std dev	67.6112	23.9641
RLM error	1900.2752	1.2560
RLM lag	1115.1423	182.7048
adjusted R^2 of regression	0.6228	0.6923
regression F value	355.9755	114.1838

4.6.3 SAR (Lag Y) regression

The robust Lagrange Multiplier tests for the DA dummy regression in Table 4.6 indicate that it may be possible to remove all autocorrelation with a simple SAR (lag-Y) regression that includes dummy variables for dissemination areas. The estimates for house and neighbourhood variables for this regression are presented in Table 4.7. The reported ρ value for this regression is reasonable; the Lagrange multiplier test for residual autocorrelation returns a value greater than 0.05, suggesting that this lag Y regression with DA dummy variables has sufficiently eliminated residual spatial autocorrelation.

As should be expected when including DA dummies in a regression, it can be seen that most census variables become statistically insignificant - although interestingly income, percent homes needing major repair, percent people 25-64 with no highschool and percent taking active transit to work are still significant at $p < 0.05$ or better. (Of course, if these census variables can change significantly over time in a nighbourhood, and our dataset covers a long enough period, then those particular census variables may well have sufficiently orthogonal components to dissemination area variables to allow independent estimation of their coefficients.)

It is interesting that the finding of a positive effect of change of school distance on house price remains at approximately the same value - but now, of course, with a large increase in standard error that greatly erodes its significance level. The negative sign for the marginal effect of change in household income is also interesting; that might suggest that the change in neighbourhood education level and upkeep of houses that associates with income is what is a more important determinant of the neighbourhood price premium, not simply a change in income itself.

The new estimates for school closure variables are presented in Table 4.8. It can be seen that some quarter estimates (Qminus6, Qplus2, Qplus6 - which are, interestingly, three of the four winter quarters being estimated) vary significantly from the estimates found in the non-spatial regression results of Table 4.3; the other estimates, however, are within 1 standard error of the nonspatial hedonic regression's results, and do not seem to show any systematic

Table 4.7: SAR regression with DA dummy variable: results

Variable	spatial coefficients			spatial lag coefficients			total spatial effects		
term	β	s.e.	p	β	s.e.	p	$\frac{\delta y}{\delta x_k}$	s.e.	p
(Intercept)	7.304								
Approximate.Age0TO5	-0.110	0.047	*	-0.059	0.026	*	-0.169	0.074	*
Approximate.Age6TO15	-0.228	0.041	***	-0.123	0.026	***	-0.350	0.065	***
Approximate.Age16TO30	-0.238	0.043	***	-0.128	0.026	***	-0.366	0.067	***
Approximate.Age31TO50	-0.360	0.037	***	-0.194	0.027	***	-0.554	0.060	***
Approximate.Age51TO99	-0.403	0.035	***	-0.217	0.027	***	-0.619	0.058	***
Approximate.AgeUnknown	-0.430	0.034	***	-0.231	0.028	***	-0.661	0.057	***
AC.YN-NoCentralAir	-0.102	0.004	***	-0.055	0.005	***	-0.156	0.008	***
Basement-Unfinished	0.115	0.018	***	0.062	0.011	***	0.176	0.029	***
Basement-Partial	0.141	0.019	***	0.076	0.012	***	0.218	0.030	***
Basement-Finished	0.197	0.019	***	0.106	0.014	***	0.302	0.030	***
Baths.Full	0.046	0.003	***	0.025	0.003	***	0.071	0.006	***
Baths.Half	0.048	0.004	***	0.026	0.003	***	0.073	0.007	***
Beds.Total	0.009	0.002	***	0.005	0.001	***	0.014	0.004	***
Garage.Spaces.Number	0.009	0.004	*	0.005	0.002	*	0.013	0.006	*
Heat.FA-ForcedAir	-0.015	0.007	*	-0.008	0.004		-0.023	0.011	*
HouseType-Semi	-0.117	0.008	***	-0.063	0.007	***	-0.181	0.014	***
HouseType-Town	-0.169	0.014	***	-0.091	0.011	***	-0.260	0.023	***
Parking.Spaces.Total	0.024	0.002	***	0.013	0.002	***	0.037	0.004	***
Room.Count	0.014	0.001	***	0.008	0.001	***	0.022	0.002	***
PCT0TO14	-0.122	0.126		-0.066	0.068		-0.188	0.195	
PCT65UP	-0.200	0.137		-0.107	0.075		-0.307	0.212	
PCTAPT	0.016	0.105		0.009	0.057		0.024	0.161	
PCTIMMIG	-0.051	0.046		-0.028	0.025		-0.079	0.071	
PCABRGID	-0.028	0.065		-0.015	0.035		-0.043	0.100	
PCVISMIN	-0.027	0.044		-0.015	0.024		-0.042	0.067	
PCMAJRPR	-0.088	0.035	**	-0.047	0.019	*	-0.135	0.054	**
PC2564HS	-0.247	0.020	***	-0.133	0.016	***	-0.380	0.033	***
PCJWACTV	0.053	0.027	*	0.029	0.015		0.082	0.042	*
PCMOV1Y	0.016	0.040		0.008	0.021		0.024	0.061	
lnAVGATHHINC	-0.045	0.021	*	-0.024	0.012	*	-0.069	0.032	*
UNEMPRAT	0.001	0.000		0.000	0.000		0.001	0.001	
diffdist	0.032	0.017		0.017	0.010		0.049	0.027	
rho	0.353								

Table 4.8: SAR regression with DA dummy variable: closetime variable results

Variable	spatial coefficients			spatial lag coefficients			total spatial effects		
term	β	s.e.	p	β	s.e.	p	$\frac{\delta y}{\delta x_k}$	s.e.	p
closetimeQminus6	0.022	0.039		0.012	0.021		0.034	0.061	
closetimeQminus5	-0.088	0.029	**	-0.048	0.016	**	-0.136	0.045	**
closetimeQminus4	-0.076	0.025	**	-0.041	0.014	**	-0.117	0.039	**
closetimeQminus3	-0.101	0.032	**	-0.055	0.018	**	-0.156	0.050	**
closetimeQminus2	-0.012	0.031		-0.006	0.017		-0.018	0.047	
closetimeQminus1	-0.056	0.025	*	-0.030	0.014	*	-0.086	0.039	*
closetimeQnow	-0.082	0.028	**	-0.044	0.016	**	-0.126	0.044	**
closetimeQplus1	-0.072	0.030	*	-0.039	0.017	*	-0.111	0.047	*
closetimeQplus2	-0.023	0.031		-0.013	0.017		-0.036	0.048	
closetimeQplus3	-0.068	0.031	*	-0.036	0.017	*	-0.104	0.047	*
closetimeQplus4	-0.101	0.024	***	-0.054	0.014	***	-0.155	0.038	***
closetimeQplus5	-0.027	0.028		-0.015	0.015		-0.042	0.044	
closetimeQplus6	-0.103	0.033	**	-0.056	0.018	**	-0.159	0.051	**
closetimeQplus7	-0.074	0.030	*	-0.040	0.017	*	-0.114	0.046	*

trend toward or away from 0.

As should be expected from a spatial regression, however, the ranges of standard errors of the estimates in Table 4.8 have increased, from 0.24-0.35 to 0.38-0.61.

A choropleth map of estimates of dissemination area dummy variable total spatial effects, which were not reported in Table 4.7, is shown in Figure 4.15. (Estimates which were not significant at p=0.1 are reported as 0 in this map.) It is obvious that there is a significant negative penalty (25%-40%) for being located near to (and especially north of) the intersection of Barton & Sherman; the two most penalized dissemination areas are both north of the CN rail line that runs north of and parallel to Barton. There are also significant (>10%) bonuses to being located in a few neighbourhoods close to the escarpment, and (strangely) in the area close to John and Barton.

4.7 Results and Discussion

Again, the results suggest that a school closure does negatively affect house prices - at least for housing in the east downtown of Hamilton. This negative effect begins at least several quarters before a school closure, suggesting that

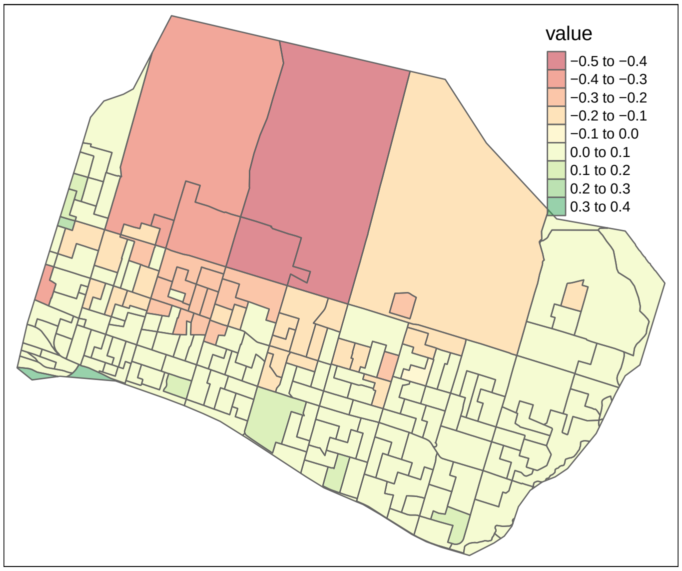

Figure 4.15: Choropleth map of dissemination area dummy estimates ($p<0.1$)

simply the anticipation of an announced closure is enough to impact prices. As can be seen in Tables 4.3, 4.5 and 4.8, this impact is also found to continue for at least seven quarters after a closure. As well, these results also suggest that the negative price impact seems to be minimized during the low-demand winter quarter.

The precise price effect, of course, is not so easily estimated: correlation of DA and closure means that the estimates for the closetime variable may still be biased toward 0. An attempt was made to rectify this by dropping dissemination area dummies and performing a further regression using a spatial Durbin model as in the previous chapter, but that regression failed: a Lagrange test on the residuals still indicated significant spatial autocorrelation of the residuals, and its AIC score was inferior to the AIC score of the lag Y regression above.

Further, while the results were statistically insignificant due to low n, an interesting possible explanation was found for why the distance-change variable is positive: it may be that the closure of a school drives a re-sorting of the neighbourhood in subsequent years, with higher-income fewer-child families - who are more likely to drive children to schools, and more likely to overpay for housing - preferentially buying the houses that are rendered more distant to primary schools after their local school is closed. In all cases, the distance-change variable is low enough that it does not ever outweigh the negative effect of a closure on prices; rather, houses which see the highest positive distance change (one kilometre) are only seeing their school closure price penalty reduced by approximately five percent of house price.

This treatment-effect analysis, while having its own limitations, manages to eliminate the problem stated in the previous chapter: our estimate here factors out the value of the school-related amenities that remain after a primary school's closure - the park, play equipment, and so on - and manages to concentrate on the value of the school building's provision of primary schooling.

Our quarter variables, unfortunately, cannot be extended further into the future to see if this negative house price effect dwindles over several years; this is limited by the closures of Roxborough Park and Woodward happening in 2015, and our housing and census datasets ending soon thereafter. For example, regressions could be performed simply on the Gibson and Stinson 2006 closures to look for a longer-run dwindling of the price penalty for a school closure: however, we end up with a sample size that is insufficient for finding statistical significance.

This method of analysis also is problematic in that the treatment population is not a representative subset of the total population, though effort is made to reduce differences by limiting Chapter 4 to downtown Hamilton. An attempt was made to fix this by splitting the treatment groups apart and narrowing the subset of control houses: for example, the Stinson and Gibson catchments were investigated as one group, with all houses inside their catchments used as treatment and houses only within 500m of their catchment boundary (i.e., not the entire rest of the downtown) included as controls. This would seem to be more likely to match treatment and control; however, while closetime variable coefficients were found to be approximately the same as in our full regression for the quarters preceding the Summer 2006 closures, the loss of n (from using 2 catchments instead of 5) meant the coefficient estimates were much less statistically significant. In addition, quarters after closure saw no significant deviation from zero for prices - possibly because of the incipient 2007-2009 housing market weakness. Simply, attempting to eliminate differences between treatment and control populations by focusing on one area made the regression vulnerable to a temporal peculiarity.

Nevertheless, given all the imperfections in this chapter's method, it does still seem confirmed that school accessibility is valued in house prices; and importantly, that this accessibility to school is valued separately from accessibility to the school's park and playground.

Conclusion

Non-spatial hedonic regressions performed on the RAHB urban transaction dataset for census years 2006, 2011 and 2016 do generally yield statistically significant coefficients for distance to primary school and highschool; the effect found is on the order of a 1% penalty to house price per km distance to school. Interpretation of the value for middle schools is complicated by the fact that some Hamilton grade schools are JK-5/6-8 while others are JK-8.

A parenthetical additional finding in the non-spatial regressions is that homeowners pay significantly more (over 10%) to buy houses in neighbourhoods where a larger fraction of the working population uses active transit (bus, walking or cycling) to get to work; it may be hypothesized that this is due to the denser opportunity landscape in neighbourhoods that have better bus service available.

However, as noted, non-spatial hedonic regression suffers from under-estimation of standard error and inability to eliminate bias in estimates. Repeating the exercise with a spatial Durbin model, a higher estimate is found for the value of school accessibility, with the expected sign for all estimates; but higher standard errors make these results less statistically significant.

Still, a regression that simply tries to study the value of school distance on house prices suffers from two main problems. First, it is unable to differentiate between the value of the school and the value of its related amenities - park, play structures, sports fields and so on. Secondly, it does not explicitly indicate any price effect of a school *closing*, nor how that will play out over time.

The final chapter addresses this by pursuing a treatment effect analysis on houses transacting in the central downtown, in the years 2005 to 2017, to look for a price effect before and after primary school closures. A non-spatial hedonic regression finds a negative price penalty of between 2% and 15% in every quarter, from six quarters before a closure to seven quarters after a closure. Nearly all estimates are significant at a 5% level or better. This value explicitly factors out the value of school-area appurtenances which remain after a closure.

In addition, this regression finds that houses which face the largest change in distance to primary school see the negative price impact lessened. A DA-level regression on the Gibson and Stinson catchments, for the ten years after their closures, while yielding statistically insignificant results due to low n, does seem to suggest that a reason for this might be that neighbourhoods which lose school accessibility are differentially bought up by higher-income homebuyers with fewer children.

The non-spatial regression was repeated with dissemination area dummy variables to attempt to reduce the effect of unobserved, spatially autocorrelated neighbourhood variables; it also finds a negative impact of primary school closures on house prices, although the effect is lessened to a 0%-12% penalty. A spatial (SAR, or lag Y) model using DA dummy variables to eliminate error autocorrelation still confirms, for the most part, a negative effect of primary school closure on house prices, but again with larger estimates of standard error. Interestingly, all regression results seem to agree that the price penalty for school closure is least during the relatively sedate winter quarter. In any case, use of dissemination area dummy variables may be biasing the results of our DA dummy regressions; a chi-squared test of independence indicates that correlation of DA identifier and school closure remains a problem.

The results of these studies do seem to suggest, then, that a primary school closure can indeed have a negative effect on house prices. The dollar amount can be significant: for a $300,000 house, the price effect of a school closure can easily run into the tens of thousands of dollars. More interestingly, the effect of a school closure might even begin a process of neighbourhood re-sorting, where the closure discount attracts buyers with higher incomes who value school proximity less.

References

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, *20*(1), 1–17.

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, *26*(2), 153–166.

Arnold, S. (2016). Councillor backs school reuse project as crucial for Barton Street. *Hamilton Spectator*, 28 March, A4.

Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, *115*(4), 588–638.

Begin, M., Caplan, G., & others. (1994). For the love of learning: Report of the Royal Commission on Learning.

Bertero, E., & Mayer, C. (1990). Structure and performance: Global interdependence of stock markets around the crash of October 1987. *European Economic Review*, *34*(6), 1155–1180.

Bishop, J. (2002a). A case for airing school renewal. *Hamilton Spectator*, 5 December, A15.

Bishop, J. (2002b). Board is acting prudently. *Hamilton Spectator*, 15 August, A12.

Bishop, J. (2002c). Unless funding rises, more cuts inevitable. *Hamilton Spectator*, 23 March, D5.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics*, *114*(2), 577–599.

Blettner, R. A. (1969). Mass appraisals via multiple regression analysis. *The Appraisal Journal*, *37*(4), 513–521.

Chin, H. C., & Foong, K. W. (2006). Influence of school accessibility on housing values. *Journal of Urban Planning and Development*, *132*(3), 120–129.

Colwell, P. F., & Guntermann, K. L. (1984). The value of neighborhood schools. *Economics of Education Review*, *3*(3), 177–182.

Cox, C. (2003a). Chopping schools. *Hamilton Spectator*, 10 July, A3.

Cox, C. (2003b). Open new school at Queen Vic: parents. *Hamilton Spectator*, 21 July, A4.

Cox, C. (2003c). Trustees get their say on closures. *Hamilton Spectator*, 14 July, A4.

Cox, C. (2004). School's out forever; last day of classes spells the end for five elementary schools. *Hamilton Spectator*, 25 June, A6.

Des Rosiers, F., Lagana, A., & Theriault, M. (2011). Size and proximity effects of primary schools on surrounding house values. *Journal of Property Research*, *18*(2), 149–168.

Ding, X., Bollinger, C., Clark, M., & Hoyt, W. H. (2020). *How do school district boundary changes and new school proposals affect housing prices* (Working Paper No. 8069). Munich: Center for Economic Studies & Ifo Institute.

Downes, T. A., & Zabel, J. E. (2002). The impact of school characteristics on house prices: Chicago 1987–1991. *Journal of Urban Economics*, *52*(1), 1–25.

Dundas Star News. (2015). French school board plan for Prince Philip still not clear. *Dundas Star News*, 7 May, 1.

Eisenlauer, J. F. (1968). Mass versus individual appraisals. *The Appraisal Journal*, *36*(4), 532–40.

Emerson, F. C. (1972). Valuation of residential amenities: An econometric approach. *The Appraisal Journal*, *40*, 268–278.

Faulkner, R. (2006a). A beauty of a school; Cathy Wever nearly ready. *Hamilton Spectator*, 3 August, A08.

Faulkner, R. (2006b). A full house for Sanford; Students from Prince of Wales may occupy old school, adding up to 1,300 on one site. *Hamilton Spectator*, 16 January, A06.

Faulkner, R. (2006c). Inner city kids to move to new schools; Sanford to become hub for downtown's neediest students as part of board's 2004-10 renewal plan. *Hamilton Spectator*, 25 April, A04.

Faulkner, R. (2006d). This school's history. *Hamilton Spectator*, 28 June, A01.

Faulkner, R. (2006e). Tornado-displaced students to stay at Seneca another year. *Hamilton Spectator*, 1 April, A06.

Faulkner, R. (2006f). 'Total chaos' ahead with schools' juggling act. *Hamilton Spectator*, 26 April, A04.

Faulkner, R. (2007a). A new chapter for students. *Hamilton Spectator*, 13 November, A03.

Faulkner, R. (2007b). Board delays school openings. *Hamilton Spectator*, 14 June, A07.

Faulkner, R. (2007c). Changes mark new school year. *Hamilton Spectator*, 4 September, A09.

Faulkner, R. (2007d). Schools get draft budget. *Hamilton Spectator*, 14 April, A11.

Faulkner, R. (2008). Lower city schools will now open next year. *Hamilton Spectator*, 14 February, A3.

Figlio, D. N., & Lucas, M. E. (2004). What's in a grade? School report cards and the housing market. *American Economic Review*, *94*(3), 591–604.

Fisher, D. (2000). It makes no sense to us to close Linden Park school; Education. *Hamilton Spectator*, 20 January, A10.

Florax, R. J., Folmer, H., & Rey, S. J. (2003). Specification searches in spatial econometrics: the relevance of Hendry's methodology. *Regional Science and Urban Economics*, *33*(5), 557–579.

Fragomeni, C. (2007). New Lawfield school opens: Positive spin to tornado damage. *Hamilton Spectator*, 20 November, A03.

Haas, G. C. (1922). *A statistical analysis of farm sales in Blue Earth county, Minnesota, as a basis for farm land appraisal.* Masters thesis, University of Minnesota.

Hamilton Mountain News. (2016). City buys Eastmount Park school. *Hamilton Mountain News*, 25 August, A6.

Hamilton Mountain News. (2021). Part of Eastmount Park on Hamilton Mountain could be sold to pay for park improvements. *Hamilton Mountain News*, 23 March.

Hamilton-Wentworth District School Board. (2021, May). Hillcrest: history. *www.hwdsb.on.ca*. Hamilton-Wentworth District School Board. Retrieved from https://www.hwdsb.on.ca/hillcrest/about/history/

Haurin, D. R., & Brasington, D. (1996). School quality and real house prices: Inter-and intrametropolitan effects. *Journal of Housing Economics*, *5*(4), 351–368.

Hayes, K. J., & Taylor, L. L. (1996). Neighborhood school characteristics: What signals quality to homebuyers? *Economic Review-Federal Reserve Bank of Dallas*, 2–9.

Hill, R. J. (2013). Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. *Journal of Economic Surveys*, *27*(5), 879–914.

Johnson, D. R. (2005). *Signposts of success: Interpreting Ontario's elementary school test scores*. Toronto: CD Howe Institute.

Kane, T. J., Staiger, D. O., Samms, G., Hill, E. W., & Weimer, D. L. (2003). School accountability ratings and housing values [with comments]. *Brookings-Wharton Papers on Urban Affairs*, 83–137.

King, M. A., & Wadhwani, S. (1990). Transmission of volatility between stock markets. *The Review of Financial Studies*, *3*(1), 5–33.

Kruchak, M. (2006). The schools... *Hamilton Spectator*, 28 June, A08.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, *74*(2), 132–157.

Leitner, R. (2012). Funding delay puts Prince Philip school closure on hold. *Hamilton Mountain News*, 8 November, A1.

Leitner, R. (2013). Hamilton School Board continues land selloff to raise school cash. *Hamilton Mountain News*, 4 June, A1.

Leitner, R. (2014). 'Hybrid plan' scales back east-Hamilton school closures. *Hamilton Mountain News*, 3 June.

Lesage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. CRC Taylor & Francis Group.

Machin, S., & Salvanes, K. G. (2016). Valuing school quality via a school choice reform. *The Scandinavian Journal of Economics*, *118*(1), 3–24.

MacLellan, D. (2007). The Fewer School Boards Act and the Toronto District School Board: Educational restructuring 1997-2003. In *Annual Meeting of the Canadian Political Science Association, Saskatoon*.

McKibbin, C. W. (1940). Is the public school a nuisance? *The Appraisal Journal*, *8*(1), 260–263.

Metz, N. E. (2015). Effect of distance to schooling on home prices. *Review of Regional Studies*, *45*(2), 151–171.

Murray, J. (2002). Needed: A vision for school renewal. *Hamilton Spectator*, 31 October, A13.

Nguyen-Hoang, P., & Yinger, J. (2011). The capitalization of school quality into house values: A review. *Journal of Housing Economics*, *20*(1), 30–48.

Nolan, D. (2014). 3 schools set to close on Central Mountain. *Hamilton Spectator*, 10 June.

Nolan, D. (2015). Scott Park won't go back to developer. *Hamilton Spectator*, 17 March, A1.

Oates, W. E. (1969). The effects of property taxes and local public spending on property values: An empirical study of tax capitalization and the Tiebout hypothesis. *Journal of Political Economy*, *77*(6), 957–971.

Ontario Ministry of Finance. (1996). *Ontario School Board Reduction Task Force: Final Report*. Toronto: Ontario School Board Reduction Task Force.

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, *32*(3), 289–320.

Owusu-Edusei, K., Espey, M., & Lin, H. (2007). Does close count? School proximity, school quality, and residential property values. *Journal of Agricultural and Applied Economics*, *39*(1), 211–221.

Pace, R. K., & LeSage, J. P. (2008). A spatial Hausman test. *Economics Letters*, *101*(3), 282–284.

Pecoskie, T. (2011). Public board recommends closing King George school. *Hamilton Spectator*, 11 April, A1.

Pecoskie, T. (2012a). Board invests thousands in shuttered school. *Hamilton Spectator*, 12 July, A3.

Pecoskie, T. (2012b). Closing: Delta, Parkview and Sir John A. Macdonald. *Hamilton Spectator*, 10 May, A1.

Pecoskie, T. (2012c). Prince Philip school to close. *Hamilton Spectator*, 17 April, A2.

Pecoskie, T. (2012d). Trustees decide to close King George public school. *Hamilton Spectator*, 14 February, A4.

Pecoskie, T. (2013a). City tries to block King George demolition. *Hamilton Spectator*, 19 September, A3.

Pecoskie, T. (2013b). Parent wants veteran trustee out; Claims conflict of interest over Prince Philip closure. *Hamilton Spectator*, 23 February, A4.

Pecoskie, T. (2013c). Parkview part of board's 'Plan B'. *Hamilton Spectator*, 13 September, A1.

Piras, G. (2014). Impact estimates for static spatial panel data models in R. *Letters in Spatial and Resource Sciences*, *7*(3), 213–223.

Pona, N. (2007). Ex-students, staff give Central Park sad send-off. *Hamilton Spectator*, 14 May, A10.

Progressive Conservative Party of Ontario. (1994). *The Common Sense Revolution: Common sense for a change*. Progressive Conservative Party of Ontario.

Prokaska, L. (2002). Hamilton native will run board. *Hamilton Spectator*, 31 August, A1.

Rajapaksa, D., Gono, M., Wilson, C., Managi, S., Lee, B., & Hoang, V. N. (2020). The demand for education: The impacts of good schools on property values in Brisbane, Australia. *Land Use Policy*, *97*, 7. http://doi.org/10.1016/j.landusepol.2020.104748

Reilly, E. (2015). Five Hamilton schools prepare to shut their doors. *Hamilton Spectator*, 2 May.

Ries, J., & Somerville, T. (2010). School quality and residential property values: evidence from Vancouver rezoning. *The Review of Economics and Statistics*, *92*(4), 928–944.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55.

Ross, S., & Yinger, J. (1999). Sorting and voting: A review of the literature on urban public finance. *Handbook of Regional and Urban Economics*, *3*, 2001–2060.

Sah, V., Conroy, S. J., & Narwold, A. (2016). Estimating school proximity effects on housing prices: The importance of robust spatial controls in hedonic estimations. *The Journal of Real Estate Finance and Economics*, *53*(1), 50–76.

Sirard, J. R., & Slater, M. E. (2008). Walking and bicycling to school: A review. *American Journal of Lifestyle Medicine*, *2*(5), 372–396.

Small, K. A., & Steimetz, S. S. (2012). Spatial hedonics and the willingness to pay for residential amenities. *Journal of Regional Science*, *52*(4), 635–647.

Smedmor, C. (2002a). How 300,000 dollars was squandered. *Hamilton Spectator*, 20 August, A11.

Smedmor, C. (2002b). *Investigation Report To The Minister Of Education, Province Of Ontario Regarding The Hamilton-Wentworth District School Board.* Toronto: Ministry of Education, Ontario.

Werner, K. (2011). Grange school project an object of perseverance. *Ancaster News*, 19 July, 1.

Werner, K. (2013). City, board discussing options for old school. *Hamilton Mountain News*, 13 August, A4.

Werner, K. (2018). Hamilton councillors approve Ancaster development. *Hamilton Mountain News*, 5 July, A6.

Wheeler, C. (1999). First day of school brings many changes. *Hamilton Spectator*, 7 September, A1.