Long Term Neighborhood Effects of Religious Diversity*

Chuhang Yin†

November 13, 2019

Abstract

This paper investigates how religious exposure in childhood affects neighborhood choice later in life. Using a panel of individuals with reported religious beliefs, socioeconomic status and residential locations over twenty years, I estimate a residential sorting model where preference for neighborhood religious composition varies by the largest religious group in childhood neighborhood (defined as "religious background") and current income. Housing services, air quality, green space and socioeconomic composition also affect neighborhood choice. Results suggest religious homophily is the strongest for low income individuals with Roman Catholic and non-Christian backgrounds in Glasgow, Scotland. Regardless of income, those with Protestant and secular backgrounds prefer living in secular neighborhoods while those with Catholic and other backgrounds prefer living in religious neighborhoods (for all religions). The heterogeneity in religious preferences contributes to inequality across religious backgrounds as reflected in average neighborhood income. Absence of such preference heterogeneity would lead to reduced segregation and a 3.8% increase in average neighborhood income for those with Catholic and other backgrounds at the lowest income quartile. This paper contributes to a growing literature on long term neighborhood effects and quantifies the cost of historical segregation on current urban structure.

*I am grateful to my dissertation committee members, Christopher Timmins, Daniel Xu, Patrick Bayer, V. Joseph Hotz and Arnaud Maurel for their guidance and support. I thank Charles Becker, Yanyou Chen, Allan Collard-Wexler, Guanpeng Dong, Daniel Garrett, Xian Jiang, Jonathan Moreno, and Gwilym Pryce for helpful comments, and members of my Daily Writing Slacker Group for their support. This paper has benefited from seminar participants at Duke, Sheffield Methods Institute, University of Glasgow, and University of Liverpool. I gratefully acknowledge the Economic and Social Research Council (ESRC) who funded the Understanding Inequalities project through which this research was conducted (Grant Reference ES/P009301/1) and Center for the Study of Philanthropy and Voluntarism at Duke University for a Dissertation Fellowship. The help provided by staff of the Longitudinal Studies Centre - Scotland (LSCS) is acknowledged. The LSCS is supported by the ESRC/JISC, the Scottish Funding Council, the Chief Scientist's Office and the Scottish Government. The author alone is responsible for the interpretation of the data. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

†Ph.D. candidate, Duke University Department of Economics. Email: chuhang.yin@duke.edu

1 Introduction

Religious diversity often has long-lasting effects on individual outcomes as it is deeply connected with political beliefs and community engagement. While diversity is often championed as a propellant for innovation and economic growth, it can also lead to conflicts and discrimination between groups especially when group identities overlap with and reinforce ethnic or class differences (Dowd (2014), McCauley (2017)). In Scotland, religion is also correlated with personal well-being: recent research has found that Catholics in Scotland have higher mortality, are less educated and are less likely to own property than Protestants even after accounting for income (Wright et al. (2017)).

There are several potential explanations for the disadvantage faced by Catholics: discriminatory hiring practices, sectarian attitudes leading to residential segregation, and preference heterogeneity resulting in residential sorting. Previous studies have investigated the prevalence of sectarian attitudes and labor market discrimination through surveys (Walls & Williams (2003), Walls & Williams (2004)), but there is little research on the extent and causes of residential segregation by religion due to the lack of geo-coded individual-level panel data. Given that neighborhoods often play a key role in forming social connections, especially when one is young, preference for a certain religion could lead to clusters of minority religions in deprived neighborhoods. If Catholics are historically segregated into worse neighborhoods, they would be more likely to stay deprived due to limited education and employment opportunities, lack of public safety, and deteriorating public health conditions. Furthermore, if the characteristics of one's childhood neighborhood affects religious preferences in adulthood, religious segregation and socioeconomic inequality can reinforce each other across generations.

This paper answers the following question: how does religious composition in one's childhood neighborhood affect her residential choice later in life? To answer this question, I first develop and estimate a residential sorting model using new individual-level panel data in Glasgow, Scotland. I allow the marginal utility for neighborhood religious composition and marginal utility of income to vary by the individual's current income and childhood religious background, which is defined as the largest religious group in childhood neighborhood. This definition of childhood religious background, as opposed to the reported "religion of upbringing", reflects the observed and unobserved characteristics of one's place of upbringing but does not directly reflect one's preferences as an adult for other amenities. The religious composition of childhood neighborhood may affect revealed preferences for religious composition through channels such as social interactions, exposure to Catholic schooling, and family attachment.

The proposed model captures two key channels of religious sorting while allowing flexible valuations of other neighborhood amenities. First, religious background can affect the marginal utility of income and make some neighborhoods less affordable (in a utility sense) than others. Second, individuals with different religious backgrounds may have different values of religious homophily. Apart from religious composition and housing services, neighborhood amenities also include air quality (PM10), coverage of green space, deprivation level, and an unobserved neighborhood quality that is constant across individuals. Valuations

of these neighborhood amenities are allowed to vary by individual religious background and current income through heterogeneity in the marginal utility of income.

The inclusion of a neighborhood unobserved quality in the sorting model requires a two-stage contraction mapping method for estimation S. T. Berry (1994). In the first stage, I recover neighborhood mean utilities and the individual-specific components in the marginal utility parameters by matching predicted and observed population shares across neighborhoods (adjusted by area size). In the second stage, I decompose neighborhood mean utilities through a Two Stage Least Square regression to recover the marginal utility of amenities for the baseline group. I use instrumental variables to address the potential endogeneity of price of housing services, deprivation, and religious composition. My model is similar to a class of horizontal locational sorting models commonly used in urban and environmental economics (Sieg et al. (2004), Bayer & Timmins (2007), Timmins & Murdock (2007)). Religious homophily by childhood background is identified through variation in religious composition of the chosen neighborhoods that are specific to individual religious background, after accounting for unobserved neighborhood quality and valuation of other amenities.

Results from the sorting model suggest that childhood religious background, in addition to income, is an important determinant of preferences for neighborhood religious composition. Those with Catholic/other childhood background have much stronger religious homophily than those with Protestant/secular childhood background. Religious homophily declines with income, but the speed of decline is slower for those with Catholic/other childhood background. Differences in valuation of neighborhood religions by background reflect the combined impact of childhood neighborhoods, including religious preferences in the family, exposure to religions through interactions with neighbors, social networks, and other aspects that affect preferences for neighborhood diversity that are not captured in adulthood income.

I quantify the importance of religious preferences through two exercises. First, I simulate population flows and neighborhood composition assuming individuals value other neighborhood religions the same amount as their religion of upbringing and that marginal utility of religion does not vary across income. Changes in segregation patterns and average neighborhood income by religious background and income quartile are measured. Second, I estimate an alternative sorting model where neighborhood religious composition is assumed irrelevant but valuation for housing and environmental amenities vary by income. Since religious composition is an important neighborhood amenity, ignoring it in the sorting estimation might lead to inconsistent estimates of the valuation of other neighborhood amenities.

The heterogeneity in religious preferences by childhood background and income exacerbates neighborhood inequality across religious groups. Counterfactual simulations suggest that the absence of such preference heterogeneity would lead to a substantial reduction in segregation across religious backgrounds. Furthermore, the average neighborhood income gap between Protestant/secular and Catholic/other backgrounds would be closed completely. Those with Catholic/other background in the lowest (individual) income quartile would experience the biggest increase in average neighborhood income of 3.8%. Ignoring preferences for religious composition of neighborhoods leads to an over-estimation of low income individuals' valuation for

neighborhood air pollution.

This paper adopts an inter-generational view to examine the long term neighborhood effects of exposure to different religions. It contributes to a large literature on neighborhood effects (Durlauf (2004)) which examines the importance of neighborhoods (in childhood or adulthood) on a broad set of individual outcomes including school performance, employment and income (Chetty et al. (2016), Chetty & Hendren (2018)), health (Leventhal & Brooks-Gunn (2003)), criminal behavior (Damm & Dustmann (2014), Chyn (2018)), and general well-being (Ludwig et al. (2012)). An important set of papers assesses the impact of the Moving to Opportunity (MTO) project, an experiment that offers randomly selected families housing vouchers to move from high-poverty housing projects to lower-poverty neighborhoods in five U.S. cities in the mid-1990s. These papers find different results by the age of move, gender, and time horizon (Kling et al. (2005), Kling et al. (2007), Chetty et al. (2016)) with the most recent study by Chetty et al. (2016) highlighting the positive long term effects on children below the age of 13 at the time of move. Another set of papers use quasi-random variation in neighborhood quality and find varying degrees of how characteristics of childhood neighborhood affect adult income (Oreopoulos (2003), Chetty & Hendren (2018)) and criminal activity (Chyn (2018), Damm & Dustmann (2014)). I contribute to this literature by assessing the long term neighborhood effects of exposure to diversity through the lens of preference heterogeneity. This inter-generational variation is rarely exploited in previous studies due to the lack of individual panels spanning a long time horizon. In a counterfactual exercise, I quantify the improvement in neighborhood outcomes if preference heterogeneity across religious backgrounds are eliminated and calculate the subsequent impact on individual incomes using findings from Chetty & Hendren (2018).

I assess the degree of religion-based residential segregation, a rarely investigated aspect of segregation in the previous literature, and highlight how childhood neighborhood reinforces residential segregation. A large literature in economics and sociology investigates the causes and consequences of residential segregation (Massey & Denton (1993), Williams & Collins (2001), Bayer & McMillan (2005)). Massey & Denton (1993) argues that residential segregation exposes African Americans to higher levels of neighborhood deterioration, creates communities with high welfare dependency, and contributes to their persistent socioeconomic disadvantages. In the economics literature, Bayer & McMillan (2005) shows that racial sorting explains a substantial fraction of the gap between whites and blacks in the consumption of neighborhood amenities, and that the adverse effects of racial sorting for African Americans are related to the relatively small proportion of African Americans in U.S. metropolitan areas. A few papers investigate the time trend in religion-based residential segregation . Using census data, Doherty & Poole (1997) finds increasing segregation (reflected in Segregation Index and Isolation Index) between Catholics and Protestants across neighborhoods in Belfast, Northern Ireland from 1971 to 1991. Gale (2013) describes the level of Muslim segregation in Birmingham, England, and points out a gradual movement of Muslims away from concentrated inner urban areas. Dean & Pryce (2017) uses housing sales data and finds evidence of perceived religious homophily between neighborhoods in Glasgow. My study contributes to this literature by providing interpretable measures of how

neighborhood religious composition is valued relative to other local amenities.

I provide new evidence on how religious diversity affects social integration using previously unavailable geo-coded panel data with religion information. There are three prominent theories in political science and sociology on how diversity affects social connections: "contact theory", where people from heterogenous groups trust each other more after they come into contact (Stouffer (1949)); "conflict theory", where diversity fosters out-of-group trust and in-group solidarity (Blumer (1958), Bobo (1999)); and "constrict theory", where diversity reduces both in-group and out-of-group trust (Putnam (2007)). This literature has mixed findings supporting all of the three theories (Putnam (2007), Bail et al. (2018), Bazzi et al. (2019)). Most of the studies use self-reported data to measure integration, but Bazzi et al. (2019) use self-reported ethnicity as well as revealed preference measures on language use, intermarriage and children's name choices. In contrast to previous papers, my paper uses revealed preference in residential location choice and quantifies the valuation for neighborhood religious composition across subgroups of the population. My findings can also be applied more generally to study other forms of integration, such as ethnic and linguistic fragmentation.

The rest of this paper is structured as follows. Section 2 introduces the importance of religion in Glasgow, Scotland. Section 3 describes the data and provides suggestive evidence of religion-based residential sorting. Section 4 develops the sorting model and proposes an estimation strategy. Section 5 outlines the main results from the sorting model, followed by sections 6 and 7 on the costs of segregating attitudes and consequences of mis-specifying the model. Section 8 concludes.

2 Background

Glasgow is a port city on the River Clyde in Southwest Scotland. As the most populous city in Scotland, it has been a cultural hub with individuals from diverse backgrounds. It is highly homogeneous racially (88% white from 2011 Census) but there are a multitude of religions. Historically, tension between Protestants and Catholics was a defining feature of the city. Although Scotland had been predominantly Protestant since the Reformation in 1560, an influx of Irish immigrants in the 19th and 20th Centuries established a large Catholic community in Glasgow City. The arrival of Irish immigrants, who are predominantly Catholic, increased competition for jobs in local industries such as ship building and mining, triggering anti-Catholic attitudes and discriminating recruitment policies (McAspurren (2005)). Many firms categorized job applications by religious affiliation, and Catholics were rarely given skilled posts (Bell et al. (1999)). Consequently, the skilled workforce of small engineering firms around Glasgow and the shipyards was largely Protestant until the 1960s.

In the early 20th Century, sectarianism was entrenched in the education system and the organization of communities even though immigration from Ireland declined. In the education system, children attended Church of Scotland schools or Catholic schools based on their religions. The Education (Scotland) Act 1918 offered the Catholic schools the opportunity to sell or lease their schools to the education authority

and to be fully incorporated into the national system. This nationalization of the Catholic school system in 1918 expanded educational opportunities for Catholics, but Catholic schools have been under continued scrutiny for fueling sectarianism through segregating children along religious lines (Flint (2012), McKinney (2015))[1]. Residential segregation manifested through the clustering of Protestants in the more affluent areas of Northwest Glasgow while Catholics living predominantly in the poorer Glasgow East End.

One unique aspect of Protestant - Catholic tension in Glasgow is the rivalry between the Rangers and Celtic football clubs (Services (2015)). The Celtic Football Club was formed in 1888 "for and by Catholics" in Glasgow's East End[2], while the Rangers was founded within the Protestant community in the 1890s. Football matches between the two, collectively referred to as the Old Firm, have led to many violent incidents including deaths in Glasgow City. Verbal abuse and threats were common during football seasons and beyond, especially in neighborhoods dominated by one religion. The following is a description of a particularly violent death in October 1999 from *The Guardian*.

> On the day that he would die, Mark Scott's mother urged him not to wear his Celtic top in case it brought him trouble. Zipping his jacket to cover the green and white hoops, the 16-year-old schoolboy had laughed. "Don't worry, Mum," he said. "They don't do that kind of thing any more." But they did, and hours later Mark had his throat cut by a man who picked him at random from a group of Celtic supporters as they walked home from a match through a Protestant area of Glasgow. His jacket was still zipped.

Since the late 20th Century, tension between the two religious groups has taken on forms of perceived job discrimination, verbal abuse, and religion-motivated hate crimes. According to the Scottish Crime and Justice Survey, Roman Catholicism has been the most cited religion in reported "religiously aggravated" charges from 2010 to 2014. In 2013-14, 63% of charges included reference to behaviour that was derogatory towards Catholicism (Services (2015)).

Neighborhood deprivation is highly correlated with religious composition: Catholics live in more deprived neighborhoods than Protestants on average. To measure neighborhood deprivation, I use Scottish Index of Multiple Deprivation (SIMD) published by the Scottish Government. SIMD is a combined measure of neighborhood deprivation using statistics on income, employment, education, health, access to services, crime and housing. The higher the SIMD score, the more deprived the neighborhood is. It is constructed at the Scottish Census datazone level, with each datazone containing 500 to 1000 residents (comparable to U.S. Census block groups). Since SIMD is a weighted measure of deprivation statistics with varying units, we can only interpret the relative levels between datazones but not the magnitude of difference[3].

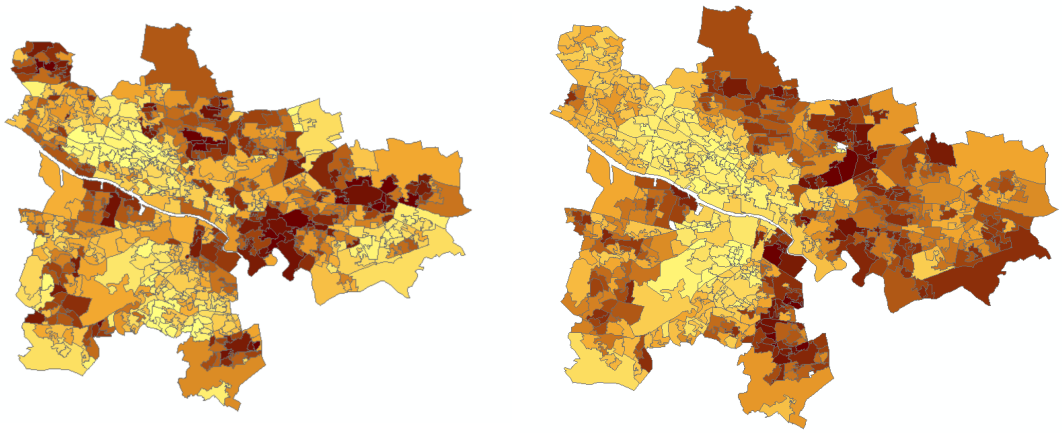
[1]Currently, 13 of the 37 secondary schools in Glasgow are Catholic schools. They have larger catchment areas than non-denominational secondary schools. There is mixed evidence on how Catholic schools perform academically compared with non-denominational schools (Shields & Gunson (2017), McKinney et al. (2014)).

[2]Bill Murray, *The Old Firm: sectarianism, sports and society in Scotland* (Edinburgh, 1984), pp 60-61.

[3]More details about SIMD can be found at https://www.isdscotland.org/Products-and-Services/GPD-Support/Deprivation/SIMD/.

Figure 1 shows the correlation between deprivation and percentage Catholic in 2011. The left panel shows the SIMD score of datazones in Glasgow City with the darker color representing higher deprivation, while the right panel shows the percentage of residents belonging to the Roman Catholic Church. Evidently, the neighborhoods with high concentration of Catholic residents coincide with higher deprivation.

Figure 1: Percentage Catholic vs Deprivation Level in Glasgow City, 2011



(a) SIMD Score

(b) % Residents Catholic

Source: 100% Scottish Census, 2011.

Further evidence of the correlation between Catholic neighborhoods and higher levels of deprivation is provided in table 1. Each column presents the OLS regression results of a specific dimension of deprivation score (higher score indicating higher deprivation) on percentages of residents belonging to Church of Scotland, Other Christian churches, and Roman Catholic Church. Note that the estimates imply correlations instead of causality. The results suggest that neighborhoods with a higher percentage of Roman Catholic residents are more deprived than other neighborhoods in all aspects including health, education, employment, income and housing. This relative deprivation persists after accounting for income deprivation (column two). Neighborhoods with more Church of Scotland residents have lower crime.

There has been a debate on whether sectarian attitudes have decreased in Glasgow City in recent decades. A survey in 2003 suggested that although more than two thirds of Glaswegians still considered sectarianism a problem, few could give an example from their experiences (*Sectarianism in Glasgow* (2003)). Previous studies have found mixed evidence of work-related discrimination against Catholics, but there has been limited evidence on whether Catholics and Protestants sort into different neighborhoods based on preference for their own religion. Since neighborhoods are important in forming social interactions that can affect one's social identity, understanding sectarianism in terms of neighborhood composition is crucial to assessing the current status of religious integration. Furthermore, quantifying the importance of childhood experiences in affecting preferences for neighborhood diversity will inform policy makers of the long term consequences of

Table 1: Scottish Index of Multiple Deprivation and % Catholic in Glasgow City Datazones 2011

SIMD Dimension	All	All	Health	Education	Employment	Income	Housing	Crime
% Catholic	106.04***	9.88***	4.80***	4.24***	56.45***	54.67***	0.19	18.96
	[8.48]	[2.22]	[0.35]	[0.36]	[4.47]	[4.68]	[7.99]	[337.36]
% Protestant	-6.98	-3.18	0.80*	-1.64***	3.10	-2.16	-88.43***	-2223.09***
	[11.25]	[2.69]	[0.46]	[0.48]	[5.93]	[6.21]	[10.61]	[448.32]
% Religion NA	286.64***	36.01***	10.52***	9.98***	145.41***	142.49***	107.59***	3372.93**
	[42.07]	[10.34]	[1.73]	[1.78]	[22.16]	[23.22]	[39.65]	[1672.50]
Constant	-11.90**	-7.45***	-1.63***	-0.86***	-6.38**	-2.53	59.69***	1005.00***
	[4.73]	[1.13]	[0.19]	[0.20]	[2.49]	[2.61]	[4.46]	[187.73]
Income SIMD		1.76***						
		[0.02]						
N	690	690	690	690	690	690	690	685
R-squared	0.23	0.96	0.26	0.20	0.23	0.20	0.11	0.04

[1] Source: ISD Scotland, 100% Scottish Census 2011.

integration policies.

3 Data

I use individual level data from the Scottish Longitudinal Study (SLS), which links Scottish censuses from 1991, 2001 and 2011 for a representative 5.3% of the Scottish population[4]. SLS includes people born on one of twenty selected dates in each calendar year. These dates are also used to update the sample in new census years. Between census years, people can join the study if they are born on one of the twenty selected dates or if they move to Scotland and their date of birth is one of the dates. SLS members leave the study either by death or by moving out of Scotland. Each wave contains information on demographics, educational attainment, employment, family structure and housing. Importantly, the postcode of enumeration is reported so a model of residential choices can be estimated.

SLS is particularly suited to study long-term neighborhood effects of exposure to religions for two reasons. First, it contains information on religion of upbringing (2001 only) and practicing religion (2001 and 2011), which is rare among Censuses. Categories of religious beliefs include Church of Scotland, Roman Catholic, other Christian, Buddhist, Hindu, Jewish, Muslim, Sikh, other religions, no religion, and unanswered. I define an aggregate category of "other religion" including Buddhist, Hindu, Jewish, Muslim, Sikh and all other

[4]SLS also contains vital events data (births, deaths, marriages), NHS Central Register data (migration into or out of Scotland), and education data (Schools Census and SQA data). It can be linked to further data sets, especially NHS health data. Details of SLS are available at https://sls.lscs.ac.uk/.

religions due to the small sample size of these groups. Second, individuals in SLS are observed in both their childhood and adulthood neighborhoods. Such longitudinal on individuals and neighborhoods facilitates a richer residential sorting model that accounts for the impact of childhood environment on outcomes later in life.

This paper studies the residential location choice of two SLS samples in 2011. The first sample contains all SLS members who are observed in 1991, 2001 and 2011 and who are between the ages of 6 and 18 in 1991. I restrict the age of the individuals to be no younger than 6 in the first year they are observed (the "childhood" year), so interactions with other residents in the neighborhood are meaningful enough to affect their preferences. The second sample contains all SLS members who are observed in 2001 and 2011 (but not in 1991) and who are between 6 and 18 in 2001 (the "childhood" year).

Exposure to religions in childhood is defined as the percentage of residents in each religion in childhood (1991 for sample 1, 2001 for sample 2) datazone. For sample 2, I use religious composition statistics from the 100% Scottish Census in 2001. For sample 1, I calculate the religious composition of neighborhoods in 1991 using a panel of SLS individuals that are observed in 1991 and 2001 as religion is not reported in the 1991 Scottish Census. For this panel, I impute each individual's religion in 1991 to be their religion of upbringing in 2001 if they were no older than 18 in 1991 and to be their practicing religion in 2001 otherwise. Datazone religious composition variables are then calculated with these imputed religious beliefs and residential locations in 1991[5].

I define an individual's religious background to be the largest group of religion in her childhood datazone. This neighborhood based definition has two advantages over the individually reported "religion of upbringing" in SLS. Firstly, reported "religion of upbringing" reflects the religious values and overall family preferences, so it is likely to be correlated with unobserved preferences for other neighborhood amenities in adulthood. In contrast, the religious composition of one's childhood neighborhood reflects the religious preferences of her neighbors and provides variation in religious exposure that could affect one's future religious preferences through neighborhood interactions. Secondly, the neighborhood definition of childhood religious background allows me to study the intergenerational instead of cross sectional neighborhood effects of religious exposure.

In the main analysis, I combine two samples and examine the residential location decisions in Glasgow City in 2011. Restricting the residential choice to Glasgow City is necessary because the data do not contain enough information on employment and earnings to account for inter-city job sorting, an important determinant for residential choice (Roy (1951), Roback (1982), Diamond (2016)). I also drop individuals who are restricted in residential locations in 2011: social housing (UK term for public housing) renters, residents in communal establishments, and full-time students.

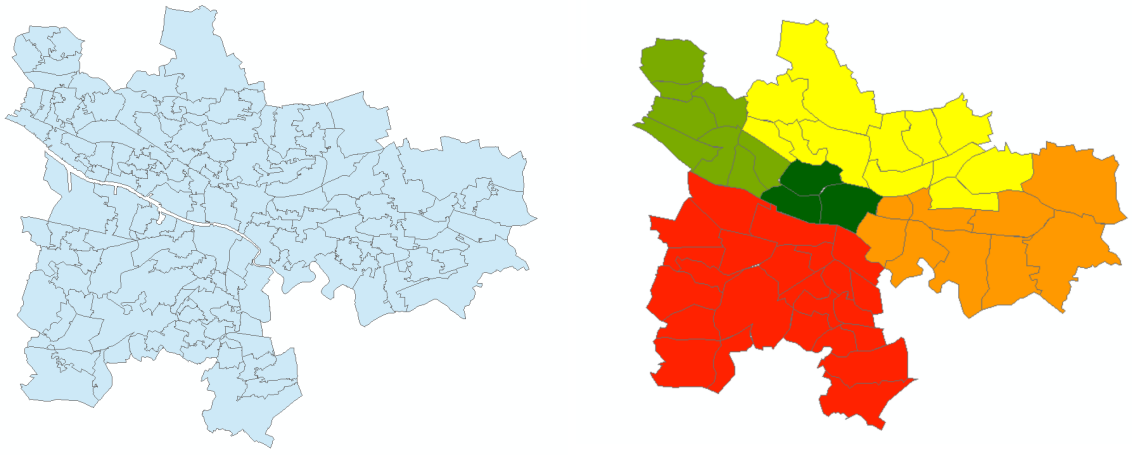
An important limitation of the SLS data is its lack of income information. Instead of income, the Scottish

[5]Since SLS contains 5.3% of the Scottish population, some datazones in 1991 have few observations. Therefore the religious composition variables are subject to small-unit bias (Carrington & Troske (1997), Ransom (2000)). Previous work has incorporated small-unit bias into the calculation of segregation measures Rathelot (2012), but it is challenging and beyond the scope of this paper to account for it in a residential sorting framework.

Census derives a four-digit occupation code based on reported employment information, and assigns an economic class category (Ns-SeC) to each individual. I estimate the log gross weekly wages of individuals using age, gender, education attainment and the occupation code using income regression results from the Quarterly Labour Force Survey (QLFS). Details of this procedure are described in Appendix A. This predicted income measure is used in the sorting model to control for current socioeconomic status in neighborhood choice.

I define neighborhoods as Intermediate Zones using time consistent boundaries from the 2001 Scottish Census. Intermediate Zones contain on average 4000 residents, comparable to U.S. census tracts. The residential location data were extracted at the datazone level, but my model will not be identified if datazones were used as neighborhoods due to the level of preference heterogeneity and the presence of neighborhood unobservables. Therefore, datazones are aggregated to the next level - Intermediate Zones. Additionally, I define five subareas of Glasgow City that plausibly constitute distinct local housing markets[6]. The subareas are locally known to have distinct demographics, culture and housing patterns. They will be used to define instrumental variables for housing price. Figure 2 displays the Intermediate Zones and subareas.

Figure 2: 2001 Intermediate Zones and Subareas in Glasgow City



(a) Intermediate Zones

(b) Subareas

Source: Scottish Census Geography 2001, People Make Glasgow Website.

I construct a comprehensive data set of Intermediate Zone amenities for the sorting analysis. Measures of school performance and business counts by sector are extracted from Scottish Neighborhood Statistics[7]. Modelled air quality variables from early 2000s to 2015 are acquired from the Department for Environment, Food & Rural Affairs (DEFRA). They contain concentration measures of PM10, PM2.5, NO2, NOx, CO,

[6]My definition broadly follows this website https://peoplemakeglasgow.com/neighbourhoods.

[7]http://statistics.gov.scot/

SO2, ozone and benzene for every $1km * 1km$ grid in the UK. These measures are aggregated to the Intermediate Zone level by overlaying the boundaries of Intermediate Zones on the grids and weighting the measures by the percentages of overlapping areas. Proximity measures to landfill sites are constructed using information on the location and capacity of waste sites from the Scottish Environmental Protection Agency (SEPA). Proportions of land used as urban green space, agricultural land, and open space without vegetation for each datazone in 2000 and 2012 are constructed using the CORINE Land Cover inventory[8].

Data on housing sales and housing attributes are acquired and merged from multiple sources. Housing sales in Glasgow City from 1991 to 2013 are acquired from the Glasgow City Council website[9]. This data set records the transaction price and postcode of each sale, and covers most areas within the city. However, it does not contain any information on housing attributes. To address this limitation, I complement the housing sales with datazone level housing attributes reported in the 100% Censuses in 2001 and 2011[10] including type of housing, number of rooms, and availability of central heating. Additionally, housing sales from 1996 to 2015 are acquired from a private mortgage company, which contains transaction price, postcode, and detailed housing attributes such as type of housing, number of bedrooms/bathrooms, and size. Combining these three sources, I estimate the joint distribution of housing prices and attributes using an Iterative Proportional Fitting (IPF) procedure (Norman (1999)) described in Appendix B.

Descriptive Statistics

Table 2 summarizes the key neighborhood amenities at the Intermediate Zone level in 2011. In terms of religious composition, more than 30% of residents are secular, while the percentage of Catholics is comparable with the percentage of Protestants (Church of Scotland and Other Christian). Around 7% of people belong to non-Christian religions.

Table 3 shows the changes in the religious composition of Glasgow City residents from 2001 to 2011. The decade witnessed a marked decline in the percentage of people belonging to Church of Scotland, a slight decline in the percentage of Catholics, a large increase in the secular population and a modest increase in the percentage of Muslims.

My main analysis is based on two childhood religious backgrounds: 1) Protestant (Church of Scotland and other Christian) or secular, and 2) Roman Catholic or other religion. I combine Church of Scotland and other Christian because the latter is a small group and historically had a similar philosophy with the former. A small sub sample of people with secular background are combined with those having Protestant background. There are few people with non-Christian childhood religious backgrounds, and their preferences cannot be recovered due to limited sample size. Therefore, they are combined with those having a Roman Catholic background. Out of the main sample of 3,100 individuals, 2,341 (75.52%) have Protestant/secular background, while 759 (24.48%) have Catholic/other background. I plan to conduct a robustness check of

[8] http://land.copernicus.eu/pan-european/corine-land-cover

[9] https://data.glasgow.gov.uk/dataset/glasgow-house-sales-1991-2013

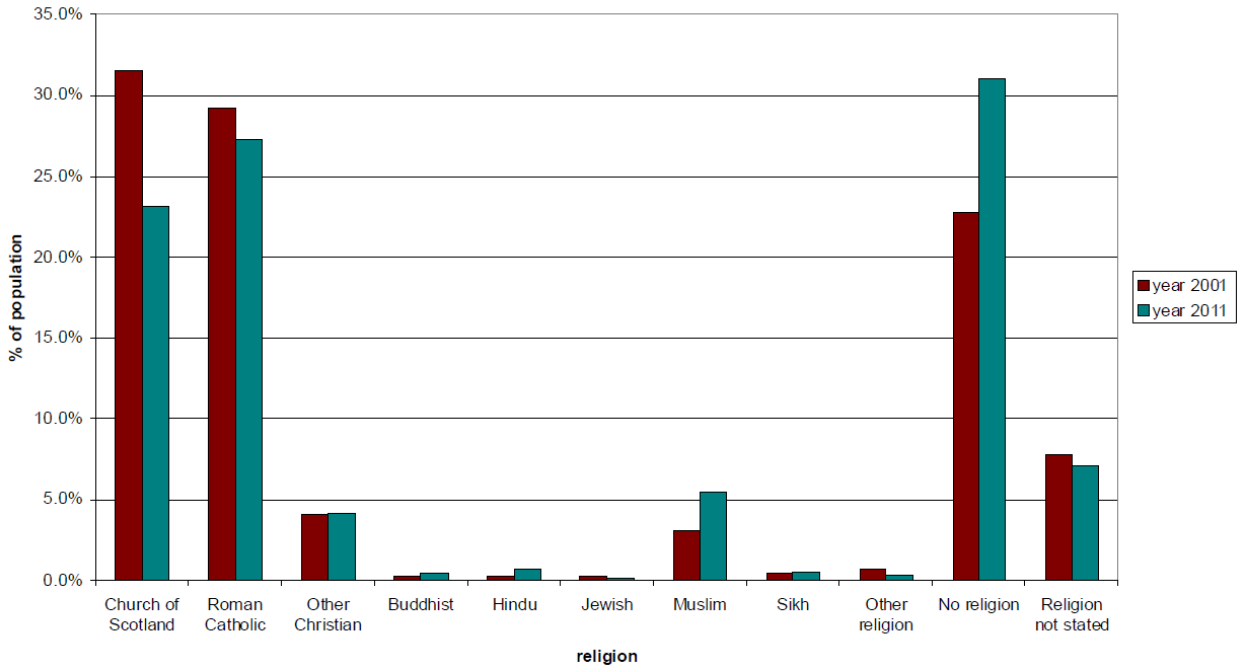
[10] http://www.scotlandscensus.gov.uk/ods-web/data-warehouse.html

Table 2: Intermediate Zone Amenities, 2011

Variable	Mean	Std. Dev.	Min.	Max.
Log price adjusted by housing attributes[1] (2015 £)	8.685	0.176	8.276	8.977
PM10	13.996	1.32	11.499	17.947
% Land as urban green space	0.095	0.136	0	0.742
% Church of Scotland	0.237	0.063	0.072	0.372
% Roman Catholic	0.276	0.077	0.087	0.448
% Other Christian	0.04	0.016	0.016	0.086
% Other religion	0.071	0.08	0.007	0.515
% No religion	0.305	0.073	0.168	0.514
% Religion unreported	0.071	0.01	0.047	0.112
% S4 pupils with 5 awards at SCQF level 5+	0.297	0.17	0.024	0.764
Number of business sites	148.534	355.049	15	3930
N = 133				

[1] The method to calculate this unit housing price is described in detail in the methodology section.

Figure 3: Religions in Glasgow City, 2001 and 2011



Source: Glasgow City Council, calculated from 2001 and 2011 censuses.

my findings where I drop those who were brought up in non-Christian or secular neighborhoods.

Table 3 summarizes the transition patterns between religious backgrounds and current practicing religions in the sample used for analysis. The transition table highlights two important patterns. First, childhood religious background does not have a one-to-one correspondence to current religion. Those who were brought up in Protestant/secular neighborhoods have 50% probability of being non religious as an adult, while more than 40% of those were who brought up in Catholic/other neighborhoods still practice their religion of upbringing. Second, there is more heterogeneity in current religion among those brought up in Protestant/secular neighborhoods than those brought up in Catholic/non-Christian neighborhoods.

Table 3: Religious Background and Current Religion (2011)

Background	% CoS	% RC	% OthChri	% Other	% None	Total
Protestant/Secular	22.08	20.21	2.14	2.99	52.28	2341
Catholic/Non-Christian	14.76	42.42	1.45	6.32	35.05	759
Total	629	795	61	118	1497	3100

[1] Source: SLS samples 1 and 2. N = 3100.

It is important to understand moving patterns from childhood to adulthood in the sample. In the combined sample, 77.7% of individuals live in a different neighborhood (Intermediate Zone) in adulthood relative to childhood, and 46.7% moved from outside Glasgow. Moving patterns vary by religious background. Among individuals who were brought up in Protestant/secular neighborhoods, 81.3% moved across neighborhoods and 56.8% moved from outside Glasgow. In contrast, 66.7% of individuals brought up in Catholic/other neighborhoods moved neighborhoods and 15.8% moved from outside Glasgow. These differences are likely driven by the concentrated Catholic presence in Glasgow. In the results section, I show that my main findings are not completely due to such differences.

I show that having Catholic or other childhood religious background is correlated with worse neighborhood amenities in adulthood, even after accounting for current income. Table 4 presents OLS regression results of neighborhood amenities in 2011 (PM10, number of business sites, SIMD score) on individuals' current income, an indicator variable of having Catholic/other religious background, an interaction term of the two, and a constant. Having Catholic/other background is significantly correlated with higher exposure to PM10 pollution, even after accounting for imputed income. Correlations between religious background and number of business sites as well as SIMD score exhibit nonlinear patterns: those with Catholic or other religious background live in neighborhoods with fewer business sites and higher deprivation, but the correlations lowers as income increases.

The descriptive statistics in this section show two key patterns in the data. First, there is persistent residential clustering by religion. Second, having grown up in a Catholic/non-Christian neighborhood is negatively correlated with current neighborhood quality. It is essential to develop a neighborhood sorting

Table 4: Religious Background and Neighborhood Characteristics

Dependent variable:	**PM10**	**PM10**	**business sites**	**business sites**	**SIMD**	**SIMD**
Log gross weekly wage	0.020	0.039	0.172	-18.9	-0.062	0.694
	[0.048]	[0.054]	[13.191]	[15.08]	[0.723]	[0.827]
Religious background is RC/other	0.116**	0.55	17.943	-420**	0.883	18.267**
	[0.057]	[0.611]	[15.942]	[169]	[0.874]	[9.294]
Log gross weekly wage*RC/other background		-0.08		80.596***		-3.198*
		[0.112]		[31.024]		[1.702]
Constant	13.875***	13.770***	156**	261***	34.645***	30.47***
	[0.264]	[0.302]	[73.314]	[83.698]	[4.021]	[4.593]
R-squared	0.001	0.001	0	0.003	0	0.001
N	3100	3100	3100	3100	3100	3100

[1] Source: SLS and constructed Intermediate Zone amenities.

model to understand the mechanisms driving such patterns. In addition, having a structural model is necessary to evaluate the welfare effects of public policies to foster integration.

4 Methodology

4.1 Model

My model assumes individuals instead of households are decision makers to be consistent with the structure of SLS. I model the residential choice decisions of those between the age of 16 and 38 in 2011 who are not full time students or social renters. Individuals choose among 133 neighborhoods (Intermediate Zones) in Glasgow City to maximize their indirect utility given budget constraints. The key feature of my model is that childhood religious background (Church of Scotand/Other Christian, or Roman Catholic/Other), in addition to predicted current income, affects the marginal utility of current neighborhood religious composition. Furthermore, I allow the marginal utility of income to also vary by income and religious background, therefore adding more flexibility to the willingness to pay for neighborhood religious composition.

The indirect utility for individual i living in Intermediate Zone j is:

$$V_{i,j} = \alpha^X X_j + \alpha_i^W W_j + \alpha_i^P P_j + \phi_j + \nu_{i,j} \tag{1}$$

where

- X_j is a vector of exogenous neighborhood amenities including pollution (PM10), percentage of land as green space, and SIMD score.
- W_j is a vector of variables describing the religious composition of Intermediate Zone j, including percentage of residents belonging to Church of Scotland/other Christian, Roman Catholic/other, and not reported. Percentage of residents without religion is omitted to avoid collinearity.

- P_j is the unit price of housing services in Intermediate Zone j, denoted in terms of annual rental flow. I assume everyone is a renter, with owners paying the annualized value of their house. Modeling the decision to own versus rent is beyond the scope of this paper. Since the interest here is not to study the dynamic capital gains for owning versus renting, annualized rent is a reasonable measure of housing cost for owners[11].
- ϕ_j is unobserved quality of Intermediate Zone j.
- $\nu_{i,j}$ is an individual-level idiosyncratic utility shock.

Marginal utility of income and of religious composition vary linearly with religious background and predicted income

$$\alpha_y^i = \beta_y^0 + \sum_{k=1}^{K} \beta_y^k G_{i,k}, y = Z, P$$

where $G_{i,k}$ includes an indicator of having Catholic/other religion background, estimated log gross weekly income, and interaction of the two. Therefore, I allow for differences in marginal utility (for income and for neighborhood religious composition) between religious backgrounds and across the income spectrum. For example, those growing up in Catholics who are poor or rich in 2011 are allowed to value living next to Catholics differently.

Estimating preference heterogeneity based on religious background instead of current practicing religion provides two advantages. First, childhood religious background is arguably exogenous (affected by parents' preferences but not the individuals') conditional on current socioeconomic status, therefore inferences based on this variable is less subject to unobserved individual preferences. Second, the definition of religious background allows us to conduct inter-generational policy evaluation using estimated preference parameters.

It is important to note that religious composition of childhood neighborhood might reflect other aspects of an individual's upbringing that can affect their preferences towards religions in adulthood. For example, growing up in a Catholic neighborhood might mean a higher chance of going to a Catholic school, which could increase one's preference to live with other Catholics as an adult. The estimated effect of childhood religious composition on religious preferences as an adult should be viewed as a constellation of related factors in the childhood neighborhood.

The indirect utility can be re-written in terms of neighborhood mean utility and an individual-neighborhood specific term:

$$V_{i,j} = \delta_j + \lambda_j^i + \nu_{i,j} \tag{2}$$

where

$$\delta_j = \beta_X^0 \cdot X_j + \beta_W^0 \cdot W_j - \beta_H^0 \ln P_j + \phi_j \tag{3}$$

$$\lambda_j^i = (\sum_{k=1}^{K} \beta_W^k G_{i,k}) W_j - (\sum_{k=1}^{K} \beta_H^k G_{i,k}) \ln P_j \tag{4}$$

[11] In 2011, 45.6% of the residents in Glasgow City are house owners.

Assuming Type I Extreme Value distribution of $\nu_{i,j}$, the probability of individual i choosing neighborhood j is

$$Pr_j^i = P(V_{i,j} \geq V_{i,q} \forall q \neq j) = \frac{\exp(\delta_j + \lambda_j^i)}{\sum_{l=1}^{J} \exp(\delta_l + \lambda_l^i)} \tag{5}$$

Maximizing the probability that each household makes its optimal neighborhood yields the following log-likelihood function

$$L = \sum_i \sum_j 1_j^i \ln(P_j^i) \tag{6}$$

where 1_j^i is an indicator variable that equals 1 if individual i is observed to be living in Intermediate Zone j and 0 otherwise.

This sorting model makes four key assumptions. First, individuals make static decisions with full mobility. While a dynamic model would capture the over time trade-off against neighborhood amenities, it is infeasible to estimate because housing is the only type of wealth (imperfectly) observed in my data. Second, individuals derive utility from housing services (reflected in rent), air quality (PM10), % urban green space, religious composition, neighborhood deprivation level, and neighborhood unobserved quality. Third, religious composition variables consist of percentage Protestant (Church of Scotland or other Christian), percentage Catholic/non-Christian, and percentage religion unreported. The religion categories are pooled to be consistent with the childhood religious background variable. In addition, such groupings are necessary because the limited sample size of individuals belonging to other Christian churches and other religions prevent the estimation of preference parameters in a framework with unobserved neighborhood quality. Since people belonging to other Christian or religions are a minority of the sample, they should not be driving the main estimation results. Fourth, marginal utility of income and preferences for neighborhood religious composition differ by childhood religious background and estimated log income, while marginal utility for PM10 and percentage green space are uniform across individuals.

The sorting model accounts for two channels of sorting by income and religious backgrounds. First, variation in the marginal utility of income makes the expensive neighborhoods more affordable to some than others. Second, variation in the marginal utility of religious composition causes people to sort into neighborhoods with certain religious compositions. I plan to account for two additional channels of sorting: one through family support by including a disutility for moving out of one's childhood Intermediate Zone; the other through proximity to employment by adding the number of business sites in neighborhood amenities.

4.2 Estimation

I estimate the model using the pooled sample of SLS individuals through a two step process similar to S. Berry et al. (1995). In the first stage, I use S. T. Berry (1994)'s method to recover neighborhood mean utilities δ from the observed datazone-level population shares in SLS, and then use Maximum Likelihood to recover preference parameters that fit the detailed individual-level data. The first-stage estimation algorithm is as follows:

1. Given an initial guess of β, use the inversion method (S. T. Berry (1994)) to find the mean utilities δ_j's that minimize the difference between predicted Intermediate Zone-level population shares and observed shares in SLS. I adjust the population shares by the area of Intermediate Zones to account for population density.

2. Substitute the old δ using the optimal δ from step 1 and calculate individual choice probabilities Pr_j^i.

3. Find the value of β's that maximize the likelihood function. Loop over the previous steps to find the parameters that maximize the likelihood function.

I estimate the price of housing services by neighborhood using a simulated sample of housing prices and attributes from the Iterative Proportional Fitting process described in Appendix B. Assuming an interest rate of 5%, the annualized rent is 5% of the house transaction price. I regress the log annual rent on housing features Z_i and Intermediate Zone dummies D_j.

$$\ln AnnualRent_{h,j} = constant + \sigma_j \cdot D_j + \gamma \cdot Z_h + \xi_{h,j} \tag{7}$$

where h denotes house, j denotes Intermediate Zone, Z_h includes type of housing and number of rooms, and $\xi_{h,j}$ is a random shock. The coefficient estimate on D_j, σ_j, is the log unit price of housing services for Intermediate Zone j.

The second stage of the estimation decomposes mean utilities δ into observable and unobservable components at the Intermediate Zone level according to equation 3, using instrumental variables for log price of housing P_j and religious composition variables.

Instruments for Unit Price of Housing

The unit price of housing is likely to be correlated with unobserved neighborhood quality, which leads to inconsistent estimates of marginal utility parameters. I construct a set of instrumental variables from the average housing features in nearby neighborhoods in the same subarea of Glasgow City. The housing features are the breakdown of housing types (detached/semidetached/terraced/flat) and average number of rooms. Since the unit price of housing is not only determined by the demand from residents but also the supply of housing relative to nearby neighborhoods, these instruments affect the price of housing but should not be correlated with the unobserved quality of the neighborhood in question. These instruments are similar in spirit with the exogenous cost shifter instruments in the Industrial Organization literature (S. Berry et al. (1995), S. T. Berry & Haile (2014)).

Instruments for Religious Composition

The religious composition of a neighborhood can be correlated with unobserved neighborhood quality. For example, Catholics can be disproportionately employed in more polluted industries and therefore live in

neighborhoods with poorer environment. In this example, if the poorer environment is not fully accounted for by PM10 included in my second stage regression, then the coefficient estimates on religious composition will be inconsistent.

Instruments for 2011 religious composition variables should be significant predictors of the religious composition of a neighborhood after accounting for other amenities but uncorrelated with unobserved neighborhood quality. To satisfy this requirement, I propose a set of instruments based on the locations of historical churches in Glasgow City. I instrument for the percentage of individuals belonging to each religion (Church of Scotland/Other Christian, Roman Catholic/Other, religion not reported) using the coverage of Church of Scotland, Roman Catholic and other Christian churches that were built before 1940. Since these churches were built decades ago, they are unlikely to be correlated with forces driving current neighborhood quality. Proximity to a church of a given religion (e.g. Roman Catholic) only affects the religious breakdown of a neighborhood through attracting more individuals from the given religion (Roman Catholic) and reducing the percentage of individuals in the other religions (e.g. Church of Scotland).

A map of the locations of Church of Scotland, Roman Catholic and other Christian churches within Glasgow City and their opening and closing years is constructed from the Glasgow Churches Together website[12]. Religious landmarks belonging to other religions (not Church of Scotland, Roman Catholic or other Christian) are not included in this analysis because they are scarce and built much more recently. From the map I calculate the percentages of each Intermediate Zone within a 500-meter radius of a Church of Scotland, Roman Catholic, and other Christian churches that have existed for at least 40 years (i.e. built no later than 1970, as robustness check) or at least 70 years (i.e. built no later than 1940) for 2011.

Neighborhood Deprivation and Instruments

To test the robustness of my results, I include the area-weighted SIMD score of Intermediate Zones as an additional amenity in the second stage decomposition as an alternative specification. SIMD scores for Intermediate Zones are calculated as the area-weighted SIMD score of all constituent datazones, the level at which SIMD scores are originally defined. SIMD score is likely to be correlated with the neighborhood quality unobserved to the econometrician, so I adopt an instrumental variable strategy to address the endogeneity problem.

A valid instrument should satisfy two conditions: first, it should be significantly correlated with neighborhood deprivation after accounting for other exogenous neighborhood amenities (PM10 and percentage of green space); second, it should be orthogonal to the unobserved component of neighborhood quality. I construct two instruments from the 1971 British Census: percentage of households living in social housing in 1971, and the percentage of households living in overcrowded dwellings in 1971.

From the early 19th Century to the mid 20th Century, cities in Scotland grew as massive influx of migrants from the Highlands and Ireland moved to the cities in search for work. The population of Glasgow increased

[12]http://glasgowchurches.org.uk/church-listings/

from 77,000 to 275,000 from 1840 to 1841. After World War II, returning soldiers created more demand for housing and relocation. The Labor government elected in 1945 committed to an extensive plan of housing construction to address the problems of overcrowding and squalor in British cities. In Glasgow, four large peripheral social housing estates, each containing their own shops, schools and churches, were developed in the late 1940s and 1950s to relocate residents from inner city slums. In the 1960s, new towns surrounding the city of Glasgow were developed to further reduce the overcrowding in the city center. By the end of the 1960s, 63% of all housing in Glasgow was social housing. The allocation mechanism focused on creating high quality, affordable housing for the working class; therefore, living in social housing was not associated with poorer socioeconomic status. However, in the 1970s the newly built new towns and peripheral estates started deteriorating due to high density and poor ventilation. Furthermore, residents in the new estates often suffered high unemployment, limited amenities, and poor access to services due to the isolation of their locations. Policy shifts in the 1980s and 1990s also increased the stigmatization of neighborhoods with high percentage of social housing. The "Right to Buy" policy introduced in 1980 allowed residents to purchase social housing at a steep discount and accelerated the process of housing privatization. Since devolution, social housing has prioritized households "in housing need" instead of the average working class family. In 2010-11, 47% of the social housing lets in Glasgow were offered to homeless individuals. There is also a starker contrast between social and private rents: in 2014 the average social rent was only 53% of the average private rent (Serin et al. (2018)).

The percentage of social housing and overcrowded households in 1971 are likely to be exogenous to unobserved neighborhood quality in 2011 because of the changes in the nature of social housing and, as a result, the composition of neighborhoods. However, the negative consequences of the fast expansion of social housing from the 1940s to the 1960s suggests that neighborhoods with higher percentage of social housing and more overcrowding in the 1970s can be less appealing over time and therefore may be more deprived forty years later. So both the exogeneity condition and the relevance condition should be satisfied.

5 Results

Instrumental Variables

Table 5 presents first stage regressions for unit housing price, religious composition variables, and SIMD score. The instruments are significant predictors for all endogenous variables.

Table 5: First Stage Regressions

Dependent Variable	Log annual rent	% CoS	%RC	% Religion NA	SIMD score
PM10	-0.075***	0.001	-0.002	0.002**	0.390
	[0.010]	[0.003]	[0.005]	[0.001]	[0.971]
% Urban green space	0.000	-0.000	-0.000	0.000*	0.015
	[0.001]	[0.000]	[0.000]	[0.000]	[0.081]
% Detached in nearby IZs	-0.032*	0.016***	0.007	-0.004***	0.565
	[0.019]	[0.006]	[0.009]	[0.001]	[1.728]
% Semidetached in nearby IZs	0.003	-0.004	0.005	0.002***	0.503
	[0.008]	[0.002]	[0.004]	[0.001]	[0.707]
% Terraced in nearby IZs	-0.018**	0.014***	-0.016***	-0.001**	-0.449
	[0.008]	[0.003]	[0.004]	[0.001]	[0.754]
Mean number of rooms in nearby IZs	0.498***	-0.222***	0.248***	0.013	9.298
	[0.123]	[0.039]	[0.059]	[0.009]	[11.433]
% near pre-1940 CoS churches	-0.071	-0.001	-0.063***	0.010***	8.262*
	[0.049]	[0.015]	[0.023]	[0.003]	[4.522]
% near pre-1940 RC churches	-0.028	-0.083***	0.090***	-0.008*	10.249*
	[0.060]	[0.019]	[0.028]	[0.004]	[5.541]
% near pre-1940 other Christian churches	0.004	-0.036*	0.013	0.001	-3.887
	[0.064]	[0.020]	[0.031]	[0.005]	[5.995]
% HHs social housing in 1971	-0.037	0.056***	0.015	0.005*	25.547***
	[0.045]	[0.014]	[0.022]	[0.003]	[4.214]
% HHs overcrowded in 1971	-0.206*	-0.166***	-0.013	0.023***	79.776***
	[0.123]	[0.039]	[0.059]	[0.009]	[11.489]
N	133	133	133	133	133
R-squared	0.560	0.575	0.453	0.288	0.643
F-stat	14.019	14.865	9.097	4.445	19.829

[1] Source: housing attributes from 100% Scottish Census tabulations (1971, 2011); neighborhood amenities constructed from DEFRA, Corine Land Survey, Glasgow Churches website; SIMD score from ISD Scotland.

Baseline Preference Estimates

In this section I show the main results from the sorting model in three specifications (different in the second stage decomposition):

- Amenities include PM10, % green space, housing price and religious composition.
- Amenities also SIMD as a control variable without instruments.

- Amenities also include SIMD with instruments.

Table 6 presents the first stage results. Those with higher incomes prefer to live in secular neighborhoods and the rate of change is faster for percentage of Protestants in the neighborhood. Those with Catholic or other childhood religious backgrounds value living close to neighbors who practice their religions. The marginal utility of income is lower for those with higher incomes (reflected by lower marginal disutility of housing price), consistent with previous findings (Layard et al. (2008)).

Table 6: First Stage Preference Estimates

	All Specifications
Log annual rent*RC/other background	-0.668
	[2.686]
Log annual rent*log income	0.397***
	[0.024]
Log annual rent*RC/other background*log income	0.188
	[0.495]
% Protestant*RC/other background	10.691
	[9.436]
% Protestant*log income	-4.762***
	[0.067]
% Protestant*RC/other background*log income	-0.797
	[1.712]
% RC/other *RC/other background	28.902***
	[7.616]
% RC/other *log income	-0.915***
	[0.054]
% RC/other *RC/other background*log income	-3.709***
	[1.389]

[1] Controls: % NA*RC background, % NA*log income, % NA*RC background*log income.

[2] Source: SLS and constructed Intermediate Zone amenities.

Table 7 presents second stage estimates from 2SLS regressions of Intermediate Zone mean utilities on neighborhood amenities for all scenarios and specifications. The first two columns present results with and without SIMD score to control for socioeconomic composition of neighborhoods. The third column presents results from my preferred specification where SIMD score is instrumented by percentage of households in public housing in 1971 and percentage of overcrowded households in 1971.

Unsurprisingly, the marginal utility of unit housing price is significantly negative (indicating positive marginal utility of income) in all specifications. PM10 is negatively valued while green space is positively valued. When religious composition from all age groups are included but SIMD score is not included, percentage Protestants is positively valued while percentage of Catholics and other religion residents is negatively valued (although the latter is statistically insignificant from zero). When SIMD score is included, estimates for marginal utilities of income and of percentage Protestant residents become larger, and estimated marginal utility of percentage Catholic/other residents becomes positive. Deprivation significantly lowers utility. When SIMD score is instrumented, marginal utility estimates for religious composition variables increase. These patterns suggest that when socioeconomic composition is omitted from an estimation, the positive correlation between religious composition variables and deprivation causes an underestimate of the marginal utility of the former. The magnitude of the underestimation differs across religions.

Table 7: Second Stage Preference Estimates

Specification	I	II	III
Log annual rent	-3.266**	-3.847***	-3.710***
	[1.371]	[1.021]	[1.050]
% Protestant	14.034***	20.562***	23.122***
	[3.092]	[1.915]	[1.666]
% Catholic/Other	-2.711	2.587*	4.570***
	[1.763]	[1.417]	[1.582]
% NA Religion	63.700***	44.061***	50.285***
	[21.445]	[5.951]	[6.758]
PM10	-0.208	-0.116	-0.075
	[0.155]	[0.102]	[0.100]
% Green space	0.008	0.009**	0.008*
	[0.006]	[0.004]	[0.004]
SIMD score		-0.023***	-0.031***
		[0.004]	[0.006]
Constant	23.723*	26.056**	22.715**
	[14.386]	[10.467]	[10.464]
IV for SIMD	No	No	Yes
R-squared	0.709	0.847	0.832
Overidentification Test: Sargan		2.98 (p = 0.70)	
Overidentification Test: Basmann		2.75 (p = 0.74)	
N	133	133	133

[1] Source: SLS and constructed Intermediate Zone amenities.

Preference Estimates with Alternative Assumptions

In this section, I estimate two alternative models to test the robustness of the main results. The first model assumes that the marginal utility of income only varies by income but not by religious background. The second model is similar to the first, but assumes there is an individual-invariant disutility of moving into a new neighborhood (Intermediate Zone) in adulthood. This model accounts for the role of social networks

(in childhood neighborhood) in neighborhood choice.

Table 8 shows the first stage estimates of both models. Model I produces similar estimates as the baseline, while model II estimates less homophily (and declining with income at the faster rate) for those with Catholic/other background. This result is consistent with the lower moving rate among those with Catholic/other background. However, the overall patterns of higher homophily for those with Catholic/other background still persist.

Table 8: First Stage Preference Estimates: Alternative Models

Alternative Model	I	II
Log annual rent *log income	0.422***	0.567***
	[0.021]	[0.022]
% Protestant *RC/other background	9.182	10.459
	[9.388]	[9.665]
% Protestant *log income	-4.789***	-5.225***
	[0.067]	[0.067]
% Protestant *RC/other background *log income	-0.526	-0.687
	[1.702]	[1.670]
% RC/other *RC/other background	23.954***	17.138**
	[7.158]	[7.133]
% RC/other *log income	-1.115**	-2.098***
	[0.054]	[0.053]
% RC/other *RC/other background *log income	-2.769**	-2.007
	[1.305]	[1.239]
1(Moved neighborhood)		-4.632***
		[0.050]

[1] Controls: % NA*RC background, % NA*log income, % NA*RC background*log income.

[2] Source: SLS and constructed Intermediate Zone amenities.

Table 9 presents second stage estimates for the alternative models. Model I produces similar results as the baseline, while model II produces higher estimates for the marginal utilities of income and of percentage Catholic/other religion neighbors.

Table 9: Second Stage Preference Estimates: Alternative Models

Model	I	II
Log annual rent	-3.760***	-4.502***
	[1.050]	[1.039]
% Protestant	23.256***	25.295***
	[1.665]	[1.648]
% Catholic/Other	5.622***	11.213***
	[1.581]	[1.565]
% NA Religion	64.817***	64.779***
	[6.755]	[6.684]
PM10	-0.074	-0.034
	[0.100]	[0.099]
% Green space	0.008*	0.008*
	[0.004]	[0.004]
SIMD score	-0.031***	-0.029***
	[0.006]	[0.006]
Constant	21.709**	25.035**
	[10.469]	[10.350]
IV for SIMD	Yes	Yes
R-squared	0.838	0.854
N	133	133

[1] Source: SLS and constructed Intermediate Zone amenities.

6 Cost of Segregating Preferences

Evaluating the costs of segregation is important for the design of integration policies. In 2018, the Scottish government launched a bold advertising campaign named "One Scotland", which aims to combat hate crime and discrimination. The campaign displays a series of letters addressed to "haters" and signed by Scotland in public places such as bus stops and railway stations, warning them that discrimination and abuse will not be tolerated. Additionally, a "One Scotland" website[13] containing educational materials on hate crime and personal stories from victims was created.

Anti-discrimination campaigns such as "One Scotland" can potentially reduce segregating religious attitudes across both religious backgrounds and socioeconomic spectrum. Although it is empirically challenging

[13] https://onescotland.org/

to assess the welfare effects of such policies, I shed light on the costs of segregating preferences by simulating population flows and neighborhood configurations if preference heterogeneity in religious composition (by childhood religious background) is eliminated.

To implement this welfare analysis, I need to solve the model and characterize the equilibrium allocation of individuals among neighborhoods given the amenities. Note that the amenities include endogenously determined housing price and religious composition of residents. To close the model, I assume that the housing stock is owned by absentee landlords and is given by

$$H_j^s = l_j P_j^\tau \tag{8}$$

where l_j is a neighborhood-specific constant reflecting differences in land endowments and other fixed factors, and τ is the constant supply elasticity. Following Sieg et al. (2004), I assume $\tau = 0, 1, 2$ and assess differences between the scenarios.

A locational equilibrium consists of a vector of housing prices (denoted in annualized rent) and an allocation of neighborhood amenities such that

- Every individual consumes the optimal amount of housing;
- Each individual lives in one neighborhood and no one wants to move to a different neighborhood;
- The housing market clears in every neighborhood.

To compute the equilibrium prices, I need to solve the nonlinear equations given by the 133 Intermediate Zone housing market clearing conditions. This process must also incorporate the fact that neighborhood religious composition changes with individual residential choice probabilities. An additional complication is that my sample is limited to individuals between 16 and 38 years old, while other age groups can affect housing prices through their residential locations. To address this complication, I assume the excluded age groups of individuals (0-15 y/o and over 49 y/o, due to census tabulation definitions) do not move between neighborhoods and use the religious composition of their respective age groups (taken from the 100% Scottish Census) to calculate the equivalent number of people in each group if they were observed in my sample.

As closed-form solutions of equilibria do not exist, I use a numerical algorithm to calculate locational equilibria described below.

1. Given initial population shares by Intermediate Zone and housing prices, calculate l_j for $j = 1, .., 133$.
2. Guess a new vector of prices given the step size (random in first iteration).
3. Calculate individual choice probabilities Pr_j^i based on new prices and initial neighborhood religious composition.
4. Calculate the population across neighborhoods, adding the missing age group.
5. Calculate the supply of housing by neighborhood.

6. Set the step size of price vector to be a positive function of excess housing demand.

7. Calculate the neighborhood religious composition variables adding the missing age group as observed in the census.

8. Loop steps 3 to 7 until the price vectors and neighborhood religious composition variables both converge.

The presence of agglomerative forces in religious composition means that there could be multiple equilibria in simulation exercises. To address this issue, I start from the observed prices and individual location choices and choose the first equilibrium meeting convergence criteria given preference estimates. For all counterfactual exercises, I start from this simulated equilibrium and use it as the baseline for comparison.

In the main counterfactual exercise, I assume individuals value other neighborhood religions the same amount as their religion of upbringing and average the marginal utility for religion across income by religious background. I am interested in measuring the amount of segregation between the two religious backgrounds and the average neighborhood income for each religious background and income quartile.

Segregation is measured by three alternative indices with different strengths and weaknesses: Dissimilarity Index (DI), Gini Index (G) and Interaction Index (Int) according to Massey & Denton (1988). Table 10 presents the formula and conceptual interpretation of these measures. While Dissimilarity Index and Gini Index describe how even the two religious backgrounds are distributed across neighborhoods (larger values for *more* segregation), Interaction Index describes the extent of exposure from the minority population's perspective (larger values for *less* segregation).

Table 10: Comparison of Segregation Measures

Measure	Formula	Interpretation
DI	$\frac{1}{2}\sum_{i=1}^{N}\lvert\frac{x_i}{X}-\frac{y_i}{Y}\rvert$	% minority who need to move to achieve even distribution
G	$\sum_{i=1}^{n}\sum_{j=1}^{n}[t_i t_j\lvert p_i-p_j\rvert/2T^2P(1-P)]$	Degree of minority and majority sharing the same areas
Int	${}_xP_y^*=\sum_{i=1}^{n}[\frac{x_i}{X}][\frac{y_i}{t_i}]$	Degree of minority's exposure to majority
Where:	x_i (y_i) is the population of minority (majority) group in area i.	
	X (Y) is the total population of minority (majority) group.	
	t_i (T) is the total population in area i (in all areas).	
	p_i (P) is the minority proportion of the population in area i (in all areas).	

Table 11 presents simulated baseline equilibrium and counterfactual results for segregation indices assuming the supply elasticity of housing $\tau = 1$. Results are similar for $\tau = 0, 2$. Segregation across religious backgrounds is reduced substantially according to Dissimilarity Index and Gini Index, but the reduction is smaller according to Interaction Index.

Table 11: Counterfactual and Baseline Segregation

	Baseline	CF	% Change
Dissimilarity Index	0.2127	0.0206	-0.9033
Gini Index	0.2948	0.0276	-0.9065
Interaction Index	0.7151	0.7548	0.0556

[1] Source: SLS and constructed Intermediate Zone amenities.

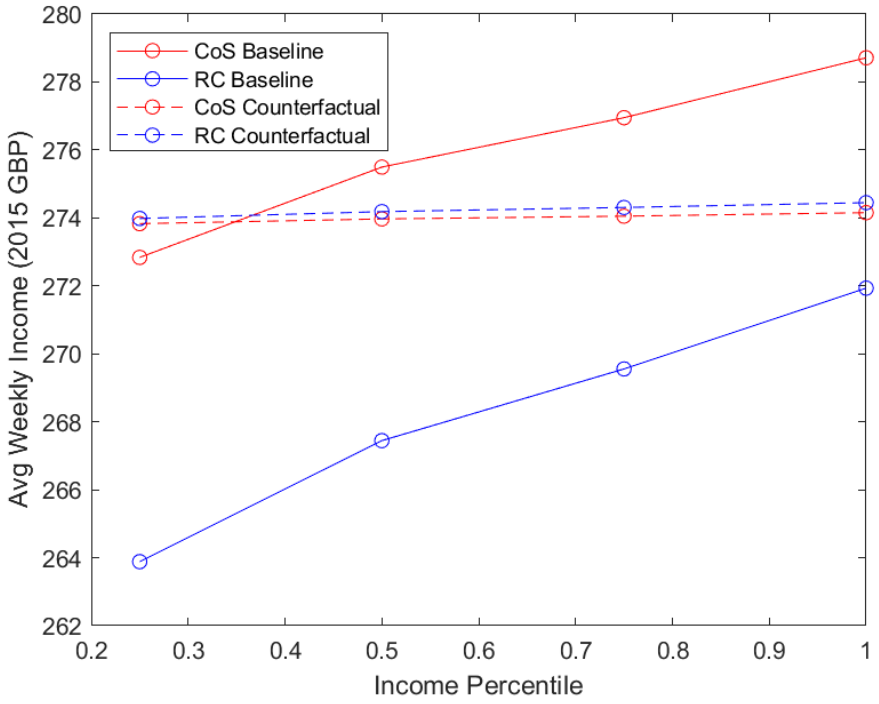
Figure 4 plots the average neighborhood income by income quartile and religious background for baseline and counterfactual. At the baseline, people with Protestant/secular background live in neighborhoods that are on average 2.5% to 3.4% richer than those with Catholic/other background depending on the income quartile of the individual. When segregating preferences are eliminated, the two religious backgrounds enjoy similar income levels in their neighborhoods. The improvement is the most salient at the lower socioeconomic spectrum: individuals with Catholic/non-Christian background at the lowest income quartile live in neighborhoods with 3.8% higher average income than before. The findings above suggest that anti-sectarian campaigns would likely benefit low-income people with Catholic/non-Christian childhood background much more than other population sub groups.

Exposure to better neighborhoods has been shown to have a long lasting positive impact on adult incomes. Chetty & Hendren (2018) uses tax data in the United States and finds that children who move to wealthier neighborhoods converge to the adult incomes of permanent residents in the destination at a rate of 4% per year of exposure. Assuming their estimated convergence rate applies to smaller neighborhoods (Intermediate Zones in Glasgow instead of Commute Zones in the United States in their paper) and is the same when individual incomes are denoted in absolute terms instead of rankings adjusted by parental incomes (see Chetty & Hendren (2018) for a detailed description on the construction of income rankings), their findings can be applied to the counterfactual simulation to calculate income changes in adulthood brought by exposure to wealthier neighborhoods in childhood. Under these assumptions, children with Catholic/other background at the lowest income quartile would live in neighborhoods with £530 (denoted in 2015£) higher average income if segregating religious attitudes were eliminated. If these children were born into the wealthier neighborhoods and stayed there for 20 years, the accumulated income gains would be £424 per year as an adult, or 80% of the initial income gap.

7 Consequences of Model Mis-Specification

Since religious composition is an important neighborhood amenity, ignoring them in the sorting estimation might lead to inconsistent estimates in the WTP for other neighborhood amenities. In this section I quantify the consequence of ignoring religious composition as a neighborhood amenity by estimating a sorting model

Figure 4: Average Neighborhood Income by Religious Background and Income, Scenario I



Source: SLS and constructed Intermediate Zone amenities.

where marginal utilities for income and environmental amenities (PM10 and percent green space) vary by income while controlling for deprivation.

Table 12 presents the preference estimates. While the marginal utility estimates for housing services and green space are similar to the baseline specification, results suggest that marginal utility from pollution increases with income. Table 13 presents the ratio of marginal utility estimates of PM10 and green space relative to housing services at the 25th, 50th and 75th income quartiles. This ratio is estimated to be positive for PM10 and negative for green space and higher for the poor. This result is potentially driven by the fact that low-income individuals with Catholic/other upbringing sacrifice air quality to live with others who practice their childhood religions. This result highlights the importance of including the cultural and identity into the valuation framework for neighborhood amenities.

Table 12: Preference Estimates without Religious Composition

Variable	Marginal Utility Estimate	Std Err
Log annual rent	-5.234**	[2.088]
Log annual rent*log income	1.364***	[0.020]
PM10	-0.659***	[0.200]
PM10 * log income	0.027***	[0.003]
% Green space	0.003	[0.005]
% Green space * log income	0.002***	[0.000]
SIMD	0.002	[0.800]
Constant	54.550***	[20.806]

[1] Source: SLS and constructed Intermediate Zone amenities.

Table 13: Marginal Utility of Local Amenities Relative to Housing Services without Religion Variables

Amenity	25Pct Inc	Median Inc	75Pct Inc
PM10	0.2813	0.2205	0.1829
% Green space	-0.0073	-0.0061	-0.0054

[1] Source: SLS and constructed Intermediate Zone amenities.

[2] Unit: per 1% increase in housing services.

8 Conclusion

This paper studies what religious composition of childhood neighborhood implies for religious preferences in adulthood neighborhood choice. Using a comprehensive panel of individual religious beliefs, socioeconomic

status and residential location choices from Glasgow, Scotland, I estimate a residential sorting model where preference for religious composition varies by religion of upbringing and current income. I find long-lasting impacts of religious upbringing on adult preferences across the socioeconomic spectrum. High income individuals prefer secular neighborhoods more than low income individuals. Those with Protestant and secular childhood backgrounds prefer to live in secular neighborhoods, while those with Catholic and other childhood backgrounds value living with religious neighbors (irrespective of religion). Importantly, low income individuals with Catholic and other religious backgrounds strongly value living close to people practicing these religions. This religious homophily could partially explain the higher neighborhood deprivation that Catholics experience relative to Protestants.

Counterfactual simulations highlight the costs of segregating attitudes and the importance of including religious attitudes in neighborhood sorting models. When preferences for all neighborhood religions are assumed to be the same conditional on religious background, segregation between religious backgrounds is substantially reduced and the neighborhood income gap between majority and minority religious backgrounds is eliminated. This equalizing effect is the strongest for people with Catholic and other religious backgrounds at the bottom income quartile: they live in neighborhoods that are on average 3.8% richer than the baseline. Not incorporating religious composition in the neighborhood choice framework can lead to inaccurate marginal utility for other amenities. Such mis-specification of the residential sorting model causes an over estimation of the poor's valuation of air pollution.

This paper has two main contributions. First, I adopt an inter-generational perspective and quantify the long term impacts of childhood exposure to religions through the lens of revealed preference in neighborhood choice. Second, I use the model estimates to quantify the how segregating attitudes reinforces inequality across religious backgrounds. This research can help policy makers evaluate the long term effects of integration policies. The methodology in this paper can also be combined with other papers (e.g. Dean et al. (2018)) to simulate changes in violence and crime patterns with shifting preferences for diversity.

Appendix A: Income Imputation in Scottish Longitudinal Study

Allowing preference for neighborhood diversity to depend on income offers a fuller range of preference heterogeneity than observed characteristics such as age and gender because income is a continuous measure. To complement the lack of income data in SLS, I impute wage incomes based on age, gender, education attainment and occupation. I use the UK Quarterly Labour Force Survey (QLFS) to estimate a multilevel random effects model of log gross weekly wages in 2001 and 2011. Then I use the estimates to predict wages from individual characteristics in SLS for respective years.

The UK QLFS began in 1992 and is the largest survey on employment and earnings in the UK. It interviews 50,000 households each quarter using a rotating panel approach: the households are divided into five groups and each group is interviewed in five consecutive quarters (also called "waves") covering different questions in each wave. At the end of each quarter, one group finishes five waves of survey and is replaced by a new group of households. For each quarter, I limit the sample to the households who are in waves 1 and 5, when each household member is asked detailed questions about employment status, industry, occupation, and earnings.

I limit the sample to individuals who are living in Scotland, in employment (as defined by ILO) and have a positive gross weekly pay from their main job, measured in the reference week leading up to the interview. Ideally the sample should be limited to those who are living in Glasgow at the time of survey, but such a restriction would lead to small sample sizes and limited variation in occupation. Furthermore, my estimation results are similar with and without the geographical limitation to the Strathclyde area. Since wage levels and returns to skills are unlikely to change substantially in a two-year window, I combine all quarters in 2001 and 2002 (2011 and 2012) to construct a data set for analysis regarding 2001 (2002)[14]. For each year, I winsorize the wage variable at the first and 99th percentiles to prevent outliers from driving estimation results.

Crucial to my income imputation process is the use of Standard Occupational Classification (SOC) code, a detailed measure of occupation developed by the UK Office of National Statistics in 1990 and revised in 2000 and 2010. The structure of SOC is hierarchical with four levels: unit groups ($n_{2000} = 353, n_{2010} = 369$) nested in minor groups ($n_{2000} = 81, n_{2010} = 90$) nested in 25 sub-major groups nested in 9 major groups. Each descending level provides more details in the occupational type within the upper category. At the highest level, the major groups include: Managers and Senior Officials, Professional Occupations, Associate Professional and Technical Occupations, Administrative and Secretarial Occupations, Skilled Trades Occupations, Personal Service Occupations, Sales and Customer Service Occupations, Process, Plant and Machine Operatives, and Elementary Occupations. As SOC codes are only reported starting from April 2001, my 2001 data set only contains April to December of the year.

I estimate a relationship between log gross weekly pay and individual characteristics using a multi-level random effects model similar to Clemens & Dibben (2014). Log gross weekly pay is assumed to depend

[14]Estimation results are similar if I restrict the analysis to only 2001 (2011).

on gender, age, whether the individual has a qualification (e.g. first degree, higher degree, professional qualification, etc) and occupation (defined by SOC2000 for 2001 analysis and by SOC2010 for 2011 analysis). Gender, age, age squared, and indicator of having qualification are assumed to have the same effect on wage across individuals, i.e. they constitute the "fixed effects" component of the estimation. Random effects are incorporated at two levels: SOC minor group and SOC unit group. Specifically, I estimate the log wage of individual i who is in SOC unit group k, SOC minor group j as the following:

$$lgWage_{ikj} = \beta_1 + \beta_2 male_{ikj} + \beta_3 age_{ikj} + \beta_4 age^2_{ikj} + \beta_5 qual_{ikj} + \xi_{1k} + \xi_{2k} \cdot age_{ikj} + \xi_{3j} + \xi_{4j} \cdot age_{ikj} + \epsilon_{ikj} \quad (9)$$

where the β parameters are fixed effects of constant, being male (relative to females), age, age^2, and having a qualification; ξ_{1k} is the random intercept term at the unit SOC level k, and ξ_{2k} is the random slope for age at the unit SOC level; ξ_{3j}, ξ_{4j} are the random intercept and slope for minor SOC groups; ϵ_{ikj} is a random error term for individual i.

Since random effects of the mixed model are not estimated directly, they are predicted using an empirical Bayes Best Linear Unbiased Estimator (BLUP) which shrinks towards the mean of the higher level (SOC minor group) if there are few cases in the more detailed group (SOC unit). The predicted random effects are then combined with the fixed effects coefficients to predict wages.

This specification takes full advantage of the hierarchical structure of SOC categorization while allowing randomness in the mean wage and the importance of age for wages within a category. Empirically, my model predicts over 76% of the variation in wages in 2001 and in 2011. Residuals from the prediction appear to be randomly distributed with respect to age.

Using the estimates from the year-specific regressions, I estimate the log weekly wage for each SLS individual in 2001 and in 2011. The estimation covers all SOC minor groups and 98% all the SOC units. For those individuals with SOC units not covered in the estimation sample, I predict their wages based on the fixed effects of age and gender and the random effects based on SOC minor groups and age.

Appendix B: Iterative Proportional Fitting

Ideally, the estimation of neighborhood housing prices should be implemented using micro data on all housing sales in Glasgow City with neighborhood information, transaction price and housing attributes. However, the micro data I have is a limited sub sample of all sales and does not cover sufficient geography. In this section, I develop an algorithm to combine the strengths of aggregate and micro housing data and provide an alternative set of hedonic estimates. I estimate the joint distributions of price and housing attributes by neighborhood so that they are consistent with observed marginal distributions. Using the estimated joint distribution, I simulate samples of housing sale observations and get neighborhood housing price estimates.

To estimate the joint distribution of housing features and prices for each datazone with observed sales, I use a statistical procedure called Iterative Proportional Fitting (IPF), or raking. To perform IPF, the researcher first needs to determine the joint distribution (cross tabulation) of interest and select a few variables (or combinations of variables) from the joint distribution that the process will be based on. These variables need to have observed population distributions. The IPF procedure then iteratively adjusts the probability of each cell in the cross tabulation until the marginal distributions of the selected variables (or combination of variables) align with the population distribution (Deming & Stephan (1940), Kalton (1983), Norman (1999)).

The statistical properties of the IPF have been thoroughly studied (Bishop & Fienberg (1969), Haberman (1984), Ireland & Kullback (1968)). It is shown to be the distribution closest to the starting cross tabulation matrix that is consistent with the observed marginal distributions (Ireland & Kullback (1968)). Therefore, the initial values for the joint distribution is crucial to the accuracy of the IPF results. In this case, I start from observed joint distributions in Nationwide sales at a geographical level higher than datazones.

The goal of the IPF exercise is to estimate the joint distribution of house price, number of rooms and housing type by datazone for Glasgow City in 2001 and in 2011. A few adjustments need to be made before we start the IPF process. First, I use housing sales data from Glasgow City Government in 2003, Nationwide sales in 2003, but use distributions of housing attributes in 2001 for the 2001 analysis. The number of days with high ozone varies considerably from year to year, but housing stock changes at a much slower rate due to construction costs and urban planning. Data on ozone is only available starting from 2003, therefore I estimate the joint distribution of housing prices and attributes in 2003 to be consistent. From 2001 to 2003 there are no substantial changes in the macro economy or in the housing market, so I assume the hedonic gradients in 2003 can represent those in 2001. Second, I combine the categories of observed housing sizes and housing types to ensure the cross tabulation does not contain too many zero cells, which is necessary to ensure the convergence of IPF (Wong (1992)). After experimenting with different categorizations, I decide to use five housing types (detached/semi-detached/terraced/flat/converted or mobile) and three housing sizes (no more than 2 rooms/3-4 rooms/5 or more rooms). Third, I discretize prices for the IPF procedure but estimate log normal parameters for price distributions in each year for the simulation process. In each year (2003 and 2011), I assign a discrete category (1-4) based on the quartiles of the inflation-adjusted housing prices. Fourth, I aggregate Nationwide observations and price distributions to six sub-areas of Glasgow City to ensure there are enough observations in each area. The sub-areas are created to follow the People Making Glasgow website[15].

I use IPF to match the joint distribution of datazones, housing price quartiles, types and sizes by sub-area and year. For each sub-area and year, I impose the following restrictions:

1. For each datazone with observed housing sales, the distribution of housing types matches that observed in the Census.

[15]https://peoplemakeglasgow.com/neighbourhoods

2. For each datazone with observed housing sales, the distribution of housing sizes matches that observed in the Census.

3. The distribution of price quartiles matches that observed in the housing sales data.

4. The joint tabulation of price quartiles, housing types and housing sizes matches that observed in Nationwide. [16].

For each year (2003 and 2011), I then weight the estimated joint distribution in each datazone by the share of Glasgow City sales in that datazone. Eventually, I have the joint distribution of datazones, housing price quartiles, type and sizes for the entire city in both years. From this distribution I use the following steps to calculate coefficients for housing attributes and prices as well as their standard errors:

1. Draw T samples, each containing N observations of housing price quartile, type and size. Draw a housing price based on the truncated log normal distribution using parameters estimated earlier.

2. In each sample t, match observations to neighborhood amenities based on datazone.

3. Using each sample t, regress log(house price) on house size, house type, and a datazone dummy variable. Recover the coefficients.

4. From the T samples, calculate the estimated coefficients as the mean of all coefficients and the 95% confidence intervals as the 5% and 95% percentiles.

[16]We impute the total number of rooms from the number of bedrooms and housing type in Nationwide to make sure they are consistent with the Census. The total number of rooms is assumed to be the number of bedrooms + 2 for houses, number of bedrooms + 1 for flats, and number of bedrooms for mobile or temporary housing

References

Bail, C. A., Argyle, L., Brown, T., Bumpus, J., Chen, H., Hunzaker, M. F., ... Volfovsky, A. (2018). Exposure to opposing views can increase political polarization: Evidence from a large-scale field experiment on social media.

Bayer, P., & McMillan, R. (2005). *Racial sorting and neighborhood quality* (Tech. Rep.). National Bureau of Economic Research.

Bayer, P., & Timmins, C. (2007). Estimating equilibrium models of sorting across locations. *The Economic Journal*, *117*(518), 353–374.

Bazzi, S., Gaduh, A., Rothenberg, A. D., & Wong, M. (2019). *Unity in diversity? how intergroup contact can foster nation building* (Tech. Rep.). National Bureau of Economic Research.

Bell, A., Bradley, M., & Kemp, A. (1999). New nation, old bigotry. *The Guardian*. Retrieved 15 August 1999, from `https://www.theguardian.com/uk/1999/aug/15/theobserver.uknews1`

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Berry, S. T., & Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, *82*(5), 1749–1797.

Bishop, Y. M., & Fienberg, S. E. (1969). Incomplete two-dimensional contingency tables. *Biometrics*, 119–128.

Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific Sociological Review*, *1*(1), 3–7.

Bobo, L. D. (1999). Prejudice as group position: Microfoundations of a sociological approach to racism and race relations. *Journal of Social Issues*, *55*(3), 445–472.

Carrington, W. J., & Troske, K. R. (1997). On measuring segregation in samples with small units. *Journal of Business & Economic Statistics*, *15*(4), 402–409.

Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, *133*(3), 1107–1162.

Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *The American Economic Review*, *106*(4), 855–902.

Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review*, *108*(10), 3028–56.

Clemens, T., & Dibben, C. (2014). A method for estimating wage, using standardised occupational classifications, for use in medical research in the place of self-reported income. *BMC medical research methodology*, *14*(1), 59.

Damm, A. P., & Dustmann, C. (2014). Does growing up in a high crime neighborhood affect youth criminal behavior? *American Economic Review*, *104*(6), 1806–32.

Dean, N., Dong, G., Piekut, A., & Pryce, G. (2018). Frontiers in residential segregation: Understanding neighbourhood boundaries and their impacts. *Tijdschrift voor economische en sociale geografie*.

Dean, N., & Pryce, G. (2017). Is the housing market blind to religion? a perceived substitutability approach to homophily and social integration. *Urban Studies*, *54*(13), 3058–3070.

Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, *11*(4), 427–444.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review*, *106*(3), 479–524.

Doherty, P., & Poole, M. A. (1997). Ethnic residential segregation in belfast, northern ireland, 1971–1991. *Geographical Review*, *87*(4), 520–536.

Dowd, R. (2014). Religious diversity and violent conflict: Lessons from nigeria. *Fletcher F. World Aff.*, *38*, 153.

Durlauf, S. N. (2004). Neighborhood effects. *Handbook of regional and urban economics*, *4*, 2173–2242.

Flint, J. (2012). Catholic schools and sectarianism in scotland: Educational places and the production and negotiation of urban space. *Policy Futures in Education*, *10*(5), 507–517.

Gale, R. (2013). Religious residential segregation and internal migration: the british muslim case. *Environment and Planning A*, *45*(4), 872–891.

Haberman, S. J. (1984). Adjustment by minimum discriminant information. *The Annals of Statistics*, 971–988.

Ireland, C. T., & Kullback, S. (1968). Contingency tables with given marginals. *Biometrika*, *55*(1), 179–188.

Kalton, G. (1983). Compensating for missing survey data.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, *75*(1), 83–119.

Kling, J. R., Ludwig, J., & Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics*, *120*(1), 87–130.

Layard, R., Mayraz, G., & Nickell, S. (2008). The marginal utility of income. *Journal of Public Economics*, *92*(8-9), 1846–1857.

Leventhal, T., & Brooks-Gunn, J. (2003). Moving to opportunity: an experimental study of neighborhood effects on mental health. *American Journal of Public Health*, *93*(9), 1576–1582.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2012). Neighborhood effects on the long-term well-being of low-income adults. *Science*, *337*(6101), 1505–1510.

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social forces*, *67*(2), 281–315.

Massey, D. S., & Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass*. Harvard University Press.

McAspurren, L. (2005). *Religious discrimination & sectarianism in scotland: A brief review of evidence (2002-2004)*. Scottish Executive Social Research Edinburgh.

McCauley, J. F. (2017). Disaggregating identities to study ethnic conflict. *Ethnopolitics*, *16*(1), 12–20.

McKinney, S. (2015). The historical and contemporary debate about the relation of catholic schools in scotland and the social problem of sectarianism. *Ricerche di Pedagogia e Didattica. Journal of Theories and Research in Education*, *10*(1), 13–45.

McKinney, S., Hall, S., & Lowden, K. (2014). *Catholic schools and attainment in scotland* (Tech. Rep.). Mimeo., University of Glasgow.

Norman, P. (1999). Putting iterative proportional fitting on the researcher's desk.

Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood. *The quarterly journal of economics*, *118*(4), 1533–1575.

Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century the 2006 johan skytte prize lecture. *Scandinavian political studies*, *30*(2), 137–174.

Ransom, M. R. (2000). Sampling distributions of segregation indexes. *Sociological Methods & Research*, *28*(4), 454–475.

Rathelot, R. (2012). Measuring segregation when units are small: a parametric approach. *Journal of Business & Economic Statistics*, *30*(4), 546–553.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of political Economy*, *90*(6), 1257–1278.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, *3*(2), 135–146.

Sectarianism in glasgow (Tech. Rep.). (2003). NFO Social Research. (Available at: http://www.glasgow.gov.uk/CHttpHandler.ashx?id=9735.)

Serin, B., , Kintrea, K., & Gibb, K. (2018). *Social housing in scotland.*

Services, J. A. (2015). *An examination of the evidence on sectarianism in scotland: Update–2015.*

Shields, J., & Gunson, R. (2017). Autonomy in the right place: School governance reform in scotland. *Institute of Public Policy Research (IPPR): Scotland*.

Sieg, H., Smith, V. K., Banzhaf, H. S., & Walsh, R. (2004). Estimating the general equilibrium benefits of large changes in spatially delineated public goods. *International Economic Review*, *45*(4), 1047–1077.

Stouffer, S. A. (1949). *The american soldier. 1. adjustment during army life.* University Press.

Timmins, C., & Murdock, J. (2007). A revealed preference approach to the measurement of congestion in travel cost models. *Journal of Environmental Economics and management*, *53*(2), 230–249.

Walls, P., & Williams, R. (2003). Sectarianism at work: accounts of employment discrimination against irish catholics in scotland. *Ethnic and Racial Studies*, *26*(4), 632–661.

Walls, P., & Williams, R. (2004). Accounting for irish catholic ill health in scotland: a qualitative exploration of some links between 'religion', class and health. *Sociology of health & illness*, *26*(5), 527–556.

Williams, D. R., & Collins, C. (2001). Racial residential segregation: a fundamental cause of racial disparities in health. *Public health reports*, *116*(5), 404–416.

Wong, D. W. (1992). The reliability of using the iterative proportional fitting procedure. *The Professional Geographer*, *44*(3), 340–348.

Wright, D. M., Rosato, M., Raab, G., Dibben, C., Boyle, P., & O'reilly, D. (2017). Does equality legislation reduce intergroup differences? religious affiliation, socio-economic status and mortality in scotland and northern ireland: A cohort study of 400,000 people. *Health & place*, *45*, 32–38.