Sexual orientation and neighborhood quality: Do same-sex couples make better communities?

Fu, Shihe

Research Institute of Economics and Management, Southwestern University of Finance and Economics, China

11 March 2008

Sexual orientation and neighborhood quality: Do same-sex couples make better communities?

Shihe Fu

Research Institute of Economics and Management

Southwestern University of Finance and Economics

55 Guanghua Cun Street, Chengdu 610074, China

Email: fush@swufe.edu.cn

Tel: (86)28-87099206

Fax: (86)28-87099300

October 31, 2007

Abstract

This study provides an initial empirical analysis on identifying the general relationship between housing values and the spatial distribution of same-sex couples across the U.S. The paper uses the 1990 and 2000 census 5% Public Use Microdata Samples and introduces the gay index into the social-amenity-based hedonic housing models. The results show significant correlation between the spatial concentration of same-sex couples and housing values; furthermore, housing values are higher in a city where the proportion of same-sex couples was higher a decade ago, suggesting that same-sex couples *make* better communities.

JEL Classifications: A14; J15; R14; R31

Keywords: Same-sex couples; Hedonic housing model; Gentrification; Gay index

> They [gay people] have paid for their identity, and in doing so have most certainly gentrified their areas. They have also survived and learnt to live their real life. At the same time, they have revived the colours of the painted facades, repaired the shaken foundations of the buildings, lit up the tempo of the street and helped to make the city beautiful and alive, all in an age that has been grim for most of urban America.
>
> —Manuel Castells, 1983, *The City and The Grassroots*, p.161

"While the spatial pattern of gay and lesbian concentration has affected property values, there has been relatively little scholarly literature on this subject" (Moss [40]). This paper contributes to the literature by identifying the general relationship between housing values and the spatial distribution of same-sex couples across the U.S. The paper uses both the 1990 and 2000 census 5% Public Use Microdata Samples (PUMSs) and estimates hedonic housing models with the social amenity and gay indexes. The tentative conclusion is that same-sex couples *make* better communities.

1 Introduction

During the 1960s, gay people began moving into decaying neighborhoods in inner cities as less advantaged citizens. Those residential neighborhoods with a concentration of gay people were called "gay ghettos" (Levine [36]).[1] Case studies demonstrate that, since then, gay people have done much to rehabilitate and restore their communities. Property values in those neighborhoods have increased much faster than average. For example, Castells and Murphy [14] and Castells [13] studied how gay men as gentrifiers developed the gay community in San Francisco. Knopp [34] studied the Marigny neighborhood in

[1] Levine's definition of a gay ghetto is an urban neighborhood that "contains gay institutions in number, a conspicuous and locally dominant gay subculture that is socially isolated from the larger community, and a residential population that is substantially gay." It does not include poverty status.

New Orleans. Marigny was experiencing disinvestment and slum landlordism in the 1960s when a small number of predominantly gay middle-class professionals started moving in. They organized a movement for historic preservation in the neighborhood and completed a large-scale gentrification. A similar story happened in Boston's South End. A few decades ago, the South End was a run-down neighborhood. But, since a few urban pioneers–many of them gay people–began moving into this historic neighborhood, it has become one of the hottest real estate markets in the Northeast. Table 1 lists the census tracts in the South End where the reported median housing values are available. In many of the census tracts the median housing values grew much faster than the overall nominal housing price index (HPI) in the Boston-Quincy Metropolitan Statistical Area (MSA).

Table 1
Median housing values in the selected tracts in the South End, Boston

	Median housing value ($)			Growth rate over a decade (%)	
Census tract ID	1980	1990	2000	1980-1990	1990-2000
25025070300	81100	356400	690200	339.46	93.66
25025070500	63300	302400	617600	377.73	104.23
25025070600	87100	421400	970000	383.81	130.19
25025070700	60900	456300	961500	649.26	110.72
25025070800	65000	288900	839300	344.46	190.52
25025070900	36700	228600	307100	533.89	34.34
25025071000	61300	165600	625000	170.15	277.42
HPI	38.01	103.22	163.40	171.56	58.30

Median housing values are from the Geolytics census CD 1980 and the neighborhood change database (1970-2000 tract data). HPI is the 4th quarter housing price index for the Boston-Quincy MSA, constructed by the Office of the Federal Housing Enterprise Oversight.

Five different theories are relevant to explain why and how the gay ghettos have been evolving into high-quality neighborhoods.

1. Spatial sorting. Many same-sex couple households are double-income, no-kids (DINK) families. Most single gay people do not have children either. This type of non-traditional family structure reduces lifetime demand for housing, children's education, and other goods, and frees some lifetime resources and time to be allocated elsewhere. If urban amenities are normal goods, then gay people will disproportionately sort into high-amenity locations. Black et al. [7] used this economic approach to explain why gay men live in San Francisco. Gay people may also choose to reside where the social milieu and political environment are tolerant and friendly towards gays (Murray [41]). The gay index, the proportion of the gay population at a location, even, has been used to measure the degree of openness and tolerance of the local social milieu, which is believed to be one of the crucial factors that attract talented people (Florida [20, 21]).

The sorting theory, however, cannot explain why, a few decades ago, gay people first sorted themselves into distressed ghettos where poverty, crime, and racial conflicts resulted in middle-class white flight; neither can it explain why some gay people choose to live in family-oriented, predominantly heterosexual neighborhoods instead of gay communities.

One difficulty raised by the possible presence of spatial sorting is that the correlation between the spatial distribution of gay people and property values could be spurious. For example, rich consumer amenities in a city can drive housing rents grow fast (Glaeser, Kolko, and Saiz [28]). This identification issue will be discussed in sections 4 and 5.

2. Gay politics. In the case study of San Francisco's gay community, Castells [13] argued that gay men struggled for survival. They formed space clusters to vote and to gain social recognition and political power. Knopp [34] documented a neighborhood-based political action group in the Marigny neighborhood of New Orleans. There, gay people founded the Faubourg Marigny Improvement Association in order to lobby the mayor and city council for land use regulations.

These case studies explained one important motivation of gay community development. However, this gay-neighborhood-based approach cannot be extended to communities where the gay population is not dominant.

3. Gentrification theory. Gentrification models (Palen and London [42]; Smith and Williams [45]) are suitable for case studies on gay communities, and can provide evidence to disentangle the sorting versus causality problem. The gentrification case studies by Castells [13] and Knopp [34], indeed, tell us that it is gay people who improved their neighborhoods, and not the case that gay people choose to move into high-quality communities. Note that gentrification usually refers to new upscale residents and capital investment flow into a decaying neighborhood.[2] However, gay people moved into decaying neighborhoods as a less advantaged group, probably not as real estate investors or speculators. An opposite argument would be that gay people expected future housing value appreciation and became risk-taking investors; but, this raises another question: Why didn't other people see the future profitability? The gentrification theory, then, can describe the dynamic process of gay community development; however, it is still not clear what the incentives of gay people as gentrifiers are.

4. Housing market discrimination towards gay people. Some studies on housing market discrimination towards African-American people show that, in the 1960s, black households paid more than white households for identical bundles of residential services (King and Mieszkowski [31]; Yinger [47]). The discrimination markups are mainly due to the supply restrictions, meaning that less housing units are available outside ghettos to black households when they are discriminated against.[3] Further, if black homeowners spend more for renovation and repair than white households of similar characteristics, the average increase in the market value of black-owned housing units will be higher than that of

[2] Recent studies on gentrification also employed this concept (see Brueckner and Rosenthal [11]; Kolko [35]).

[3] An example of discrimination against racial minority is steering–black homebuyers are shown houses in systematically different neighborhoods than those shown to comparable white homebuyers.

white-owned housing units (Kain and Quigley [30]).[4] Does the housing market discrimination towards African-Americans also apply to sexual orientation? Legal studies do provide evidence of housing market discrimination against non-heterosexual people.[5] However, no systematic empirical work has been done, since sexual orientation is not as easy to observe as faces, or colors of skin; it can even be concealed. Future studies on housing market discrimination based on sexual orientation will mainly depend on the availability of reliable data.

5. Intrinsic preference. Case studies on gay community development bring up an interesting question: Do gay people, compared with their heterosexual counterparts, make better neighborhoods? If this is true, what are the driving forces or motivations? Godfrey [29] argued that bohemian and gay people have strong aesthetic tastes that help them identify the charm and profitability of run-down housing units. His three-stage life-cycle theory of gentrification suggests that future property values are correlated with the past gay population.[6] Fellows [19] documented the lives of many gay men across the U.S., and concluded that gay men are very sensitive to beauty, and have long been impassioned pioneers, as keepers of culture from large cities to rural communities: restoring decrepit buildings, revitalizing blighted neighborhoods, etc. It is these stronger aesthetic tastes of gay people that have made their neighborhoods nicer and better. Unfortunately, the intrinsic preference theory is hard to test empirically, as it is very difficult to construct data to measure artistic tastes of people, artistic characteristics of buildings, and household expenditures on housing decoration

[4] The studies from the 1970s tend not to find evidence of discrimination markups; recent studies show that the African-American rent premium fell dramatically between 1940 and 1970 and had reversed entirely by 1990. See Ross [44] for the detailed review.

[5] A few states have fair housing laws that prohibit discrimination in housing based on sexual orientation, such as California, Massachusetts, Minnesota, New Jersey, Rhode Island, Vermont, and Wisconsin.

[6] The first stage is bohemian influx: single people, counter-cultural, gay and lesbians, artists, and feminist households, as urban pioneers, discover the special charms of run-down or dangerous neighborhoods, such as social diversity, subcultural identification, architectural heritage, or historical distinction, and make them livable and attractive. The second stage is middle class transition: local businessmen and middle class residents move in, and housing speculation begins. The third stage is bourgeois consolidation, when outside firms enter the local shopping area and residents become increasingly homogenous; rents and property values rise and push the low-income and original bohemians out to other areas.

and arts.[7] However, as we shall see later, the empirical results of this paper provide indirect evidence to support this theory.

If, indeed, gay people can make better neighborhoods, compared with heterosexuals, then, after controlling for housing characteristics and other neighborhood attributes, housing prices in a neighborhood with proportionately more gay people should be higher than those with proportionately less gay people. Additionally, housing prices in a neighborhood with proportionately more gay people initially should grow faster than those with proportionately less gay people.

However, no empirical research has been done on the general relationship between property values and the spatial distribution of gay population.[8] More surprisingly, even though economists have performed extensive studies on racial and gender minorities, few have been interested in studying sexual minorities, although Nobel laureate Gary Becker began to do so in the early 1980s (Becker [5]).[9]

This study aims to empirically identify the general relationship between property values and the spatial distribution of gay people, rather than focus on a particular gay community. The primary goal is to use both the 1990 and 2000 census 5% PUMSs to test whether gay people contribute to better communities. A gay index and a set of neighborhood attributes are constructed to proxy for local social amenities. Hedonic housing models, including individual housing

[7] Moorhouse and Smith [39] studied the market values of different architectural features of 19th century row houses in Boston's South End.

[8] One exception is a recent working paper by Florida and Mellander [22]. Their statistical analysis shows that even after controlling for income or wages, technology, and human capital, the artistic, bohemian, and gay population is still positively and significantly correlated with MSA median housing values.

[9] Black et al. [7] studied why gay men live in San Francisco, by using the 1990 PUMS. Florida [20] constructed a gay index to proxy for the openness and diversity of urban social milieu in order to study the spatial distribution of a creative class across metropolitan areas. A few labor economists studied the wage gap between homosexual and heterosexual people (Allergetto and Arthur [3]; Carpenter [12]; Black et al. [8]; Blandford [10]), and all found that gay men earn less. There are two very important case studies on spatial organization and the development of gay communities in San Francisco (Castells [13]) and in New Orleans (Knopp [33, 34]). Klawitter [32] explained why so few economists are interested in research on issues of sexual orientation, despite the cultural, political, and economic importance of the topic. The reasons include discrimination against sexual minorities, the absence of support for such research, and the scarcity of appropriate models and data.

characteristics, the gay index, and a set of other neighborhood attributes, are then estimated at the Public Use Microdata Area (PUMA) level and at the city level. The results show that the correlation between the gay index and property values is strong, significant, and robust; furthermore, property values grew faster in cities where a proportionately higher population of gay people existed one decade earlier. Therefore, this study provides the evidence that gay people "*cause*" or "*make*" better communities. To be more interesting, our empirical results are consistent with the intrinsic preference hypothesis in two very intuitive aspects: intrinsic aesthetic tastes motivate gay people to renovate their housing units not only everywhere (which generates locational premia), but, also, all the time (which generates faster growth in housing values).

Since the U.S. census data can identify only same-sex unmarried partners but not single gay people, our tentative conclusion would be, to be more precise and rigorous, that same-sex couples contribute to better communities. The empirical evidence found in this study is hoped to be of value for developers, urban planners, financial institutions, and related government sectors to make decisions on gay and gay-community-related issues, especially real estate markets in gay neighborhoods and the effect gay people have upon them.[10] This study is also hoped to stimulate further economic research on sexual orientation and gay communities.

The rest of the paper is organized as follows: section 2 reviews the theory of urban social amenities and the concept of gay index, and section 3 introduces the data sets. Section 4 presents the cross-section models and the results, and section 5 discusses further identification and causality issues. Section 6 reviews the interpretation of a causality link, and section 7 discusses further related issues and concludes.

[10] Moss [40] discussed the vital role that gays and lesbians play in renewing central-city neighborhoods.

2 The theory of urban social amenities and the gay index

In contrast to the five theories described in section 1, this study mainly draws its framework from the theory of urban social amenities developed in Fu [23], and extends it to incorporate the gay index.

Extensive studies show that urban amenities, such as accessibility, climate, the quality of views, school quality, crime, and racial segregation, can be capitalized into property values. Different from natural urban amenities, school quality, crime, and racial segregation are social consequences of social interactions among urban populations. Information spillovers, peer effects, and neighborhood effects play a very important role in shaping pupils' school achievement, criminal behavior, and labor market outcomes in cities. Such location-specific social interactions, where urban residents interact with each other, are referred to as urban social amenities. Urban social amenities can further be classified into three categories: human capital, social capital, and cultural capital. Here we briefly explain the intuition how each type of social amenities affects property values.

Human capital is the knowledge and skills embodied in individuals. The main linkage between human capital and property values is the spatial equilibrium mechanism. If, at a location, wages are higher because of knowledge spillovers, land and housing rents must adjust correspondingly to ensure that economically identical workers achieve the same utility level. The second mechanism is that the social benefit of education reduces the probability of engaging in socially costly activities, such as committing a crime (Lochner [38]), and makes residential neighborhoods safer. The third mechanism is that well-educated neighbors, themselves, are attractive consumption amenities (Glaeser, Kolko, and Saiz [28]). Following Fu [24], this paper constructs two variables to measure the quality and diversity of local human capital stock. Detailed variable definitions are given in section 4.

Social capital refers to the relations between people that can be used to reach other resources or facilitate certain actions of actors (Coleman [15]). Putnam [43] argued that social capital at the community level is a strong predictor of educational performance, crime rate, and other measures of neighborhood quality of life. Social capital, specifically, the strength, diversity, and content of network ties, has important effects on labor and housing market outcomes. In poor communities in inner cities, the strength of strong ties may deprive their residents of sources of useful information about employment opportunities elsewhere and ways to attain them (Stack [46]). DiPasquale and Glaeser [18] showed that home ownership can promote residents' investment in social capital, both through the direct incentive effect and the longer tenure. Here, we tentatively use the percentage of different types of households in a neighborhood to measure the stock of social capital at the community level. These types include home ownership rate, the percentage of households that moved into a location a certain number of years ago, and the percentage of households with children under four years of age.

Cultural capital refers to the values, norms, customs, and cultural traditions that serve to identify and bind together a given group of people. Much cultural capital is formed through interactions with people from the same culture. Race, language spoken, and religion are the main indicators of cultural capital. Studies on residential segregation and labor market racial discrimination show that cultural capital has important effects on housing and labor markets. The bounded solidarity in a homogenous racial community can be a powerful motivational force; in contrast, heterogeneity in terms of cultural background may decrease trustworthiness in social groups. Alesina, Baqir, and Easterly [2] showed that after controlling for other socioeconomic and demographic factors, the share of spending on productive public goods in U.S. metropolitan areas and counties are negatively related to the local ethnic fragmentation. This suggests that high diversity of races may have negative impact on local property values. This paper constructs two variables to measure cultural capital: the percentage of residents who spoke English well, and the racial fragmentation in terms of racial diversity.

The gay index is defined as the ratio of the total number of gay people at a location to the total number of residents at that location. Since the U.S. census data can identify only same-sex unmarried partners but not single non-heterosexual people, here, "gay people" refers to only people who were self-identified as same-sex unmarried partners, or, what we call, same-sex couples. Lesbians have been considered to have different social and economic behavior from gay men. For example, lesbians are more likely to adopt children, less likely to form residential clusters. Therefore, it makes sense to have three gay indexes: the percentage of male, female, and total same-sex couples at a location.

The interpretation of the gay index could be multi-fold. First, the gay index can proxy for the degree of openness and tolerance of the local social environment (Florida [20, 21]). From this perspective, the prevalence of gay population can be treated as a type of social amenity that can be included into hedonic housing models. Second, the inclusion of the gay index to the hedonic housing model can test how sexual orientation of residents is related to housing values, and in general, neighborhood quality. Third, since we do not have data on the intrinsic characteristics and social behaviors of gay people, the gay index could also capture all the effects of unobservable characteristics of gay people, such as aesthetic tastes. Last but not least, the gay index may be endogenous because of the spatial sorting problem. We will be discussing how to use instrumental variables (in section 4) and panel data (in section 5) to deal with this issue.

3 Data

The data sets used in this study are the 1990 and 2000 census 5% PUMSs, downloaded from the Integrated Public Use Microdata Series (IPUMS) web site (www.ipums.org). The data contain detailed information on individuals' personal characteristics, family structure, characteristics of housing units, and the geographic information of residence and workplace. The PUMA (city) sample used in this paper selects housing units and workers of age 16-65 in identified PUMAs (cities) in metropolitan areas.

The 1990 census data is the first census that includes sexual orientation information. The data contain a variable "relationship to the head of household," of which one value is "unmarried partner."[11] We identify same-sex unmarried partners as gay people, or same-sex couples, as we call them in this study, including both male and female same-sex couples. This is the only way to identify non-heterosexual orientation in the census data, and it has been employed in all the census-data-based studies on sexual orientation (Black et al. [6]). There is no way to identify single non-heterosexual people in the census data. Another point worth noting is that there is no way to identify whether people who filled out the census survey forms as same-sex unmarried partners were publicly acknowledged as such (i.e., "in or out of the closet"). Therefore, the census data is not ideal for studying housing market discrimination against same-sex couples.

The geographic hierarchy of the census PUMS is worthy of detailed explanation because geographic levels affect the estimation and interpretation of the hedonic housing model. There are four geographic levels in the PUMS: state, metropolitan area, city, and PUMA (or Super PUMA). A PUMA is a geographic area of at least 100,000 residents. If the population exceeded 200,000, then, the Census Bureau split the area into as many PUMAs of 100,000+ residents as possible. In the 1990 PUMS, PUMAs sometimes span across state lines; but for all the PUMAs that are nested within metropolitan areas, none crosses a state boundary. The PUMAs in the 2000 census do not cross state lines. In 1990 cities are identified when at least 99% of the PUMA residents lived in a given city and no more than 1% of the PUMA residents lived outside the city limits. In the 2000 PUMS, only cities meeting the minimum population threshold of 100,000 residents are identified. For example, only Boston, Cambridge, Lowell, Springfield, and Worcester are identified in Massachusetts.[12]

Since housing markets are very localized, the lower geographic level data can

[11] The questionnaire provides four choices if a person in a household is not related to the head of the household: roomer, boarder, or foster child; unmarried partner; housemate, roommate; other non-relative.

[12] An unpublished appendix with a detailed explanation of census geographic levels is available from the author, upon request.

better control for localized neighborhood externalities and local amenities. For example, a metropolitan-area fixed effect can control for the impacts of regional or macroeconomic conditions on local housing markets, while a census-tract fixed effect can even control for local zoning regulation, local school quality, and other tract-specific amenities. Since PUMA is the smallest geographic level in the 1990 and 2000 census 5% PUMS, and PUMAs are nested within a state, state fixed effects are used when a neighborhood is defined at the PUMA level.

There are 126 cities and 1,726 PUMAs in the 5% sample of the 1990 census PUMS, and 150 cities and 1,581 PUMAs in the 2000 census PUMS, identified in metropolitan areas. Table A-1 in the appendix presents some geographic summary statistics for the 2000 PUMS; Tables A-2 and A-3 list the top 10 cities and PUMAs in 2000, in terms of the same-sex couple index. For the summary statistics of demographic characteristics of gays and lesbians across states and MSAs in 2000, see Gates and Ost [25].

4 Cross-sectional correlation

This section uses the 5% sample of the 2000 census PUMS and estimates hedonic housing models at the PUMA and the city levels, respectively.

4.1 PUMA level models

In this subsection, hedonic housing models are estimated with a set of social amenity variables and the gay index, constructed at the PUMA level. Since PUMAs are nested within a state, to control for the differences in natural amenities and housing production efficiency at macrogeographic levels larger than PUMAs, state fixed effects are included.[13] Another advantage of using state fixed effects is that the differences in the legal environment, in terms of anti-discrimination against sexual orientation across states, are also controlled for.

[13] The PUMAs in the PUMA sample are also nested within metropolitan areas, but a few metropolitan areas do cross a state line.

The hedonic housing model at the PUMA level is specified as follows:

$$\log P_{nj} = \alpha + \lambda_s + \beta' X_n + \gamma' X_j + \epsilon_{nj}, \qquad (1)$$

where P_{nj} is the reported housing value of housing unit n at PUMA j; α is a constant; λ_s is a state fixed effect, representing natural amenities and legal environment that are state specific; X_n is the vector of characteristics of housing unit n, variables include the number of bedrooms and other rooms, building age, and a set of dummies for housing type: dummies for mobile, detached, attached, number of apartments is 2, 3-4, 5-9, 10-19, 20-49, and greater than 50, if lot size is greater than 10 acres, and if there is a business or medical office on it; X_j is the attributes vector of social amenities at PUMA j, including variables measuring human capital, social capital, cultural capital, and the gay index; β and γ are the coefficient vectors to be estimated; ϵ_{nj} is the disturbance term, probably spatially correlated and not identically distributed.

Two variables are used to proxy for different dimensions of local human capital stock at the PUMA level:

Average education: Percentage of residents with a college degree, or higher, at a PUMA, proxy for the quality of local human capital stock.

Occupation diversity: Proxy for the broadness of human capital, in terms of occupations at a PUMA. It equals one minus the Herfindahl index of occupations at a PUMA. The classification of occupations is listed in Table A-5 in the appendix. Let S_{oj} denote the ratio of residents of occupation o at PUMA j to the total residents at PUMA j, then

$$Occupation\ diversity = 1 - \sum_o S_{oj}^2. \qquad (2)$$

The PUMA level social capital is tentatively measured by two variables:

Home ownership rate: Percentage of households who were homeowners at a PUMA.

Five-year households: Percentage of residents at a PUMA who lived in the same house for at least five years.

Similarly, the PUMA level cultural capital is also tentatively measured by two variables:

English proficiency: Percentage of residents at a PUMA who spoke English well.

Racial fragmentation: Diversity index in terms of races. It equals one minus the Herfindahl index of races. The races are classified as White, Black, Hispanic, Asian and Pacific Islander, and others. Let S_{rj} denote the ratio of residents belonging to race r at PUMA j to the total number of residents at PUMA j, then

$$Racial\ fragmentation = 1 - \sum_r S_{rj}^2. \tag{3}$$

The variable of our focus is the gay index. The following three indexes are defined, but only one of them is used in a model:

SS index: Percentage of residents who were identified as same-sex unmarried partners at a PUMA;

SSM index: Percentage of residents who were identified as male same-sex unmarried partners at a PUMA;

SSF index: Percentage of residents who were identified as female same-sex unmarried partners at a PUMA.

If residents commute to the central business district (CBD) or subcenters to work, then commuting costs will be (partially) capitalized into residential land rents. Therefore, a variable "*Commuting time*", the average commuting time (minutes) to workplace in a residential PUMA, is also included.

To check the stability of the model specification and the robustness of the estimation, we also try other related variables such as the percentage of households that moved into a house within one year, within two years, with the presence of children under the age of four; the percentage of unemployed, white, or bohemian residents.[14]

[14] The bohemian index is constructed using the definition by Florida [20]. The selection of bohemian occupations included: authors; designers, musicians and composers; actors and directors; craft-artists, painters, sculptors, and artist printmakers; photographers; dancers; artists, performers, and related workers.

We use the Huber/White estimate of variance clustered by PUMAs to produce consistent standard errors. Table 2 presents the results of the hedonic housing model with the gay index and other social amenity variables. The coefficients of housing characteristic variables have the expected signs. Since housing characteristics are not of particular interest in this study, they are not reported in Table 2.

Table 2
Hedonic housing models with the gay index at the PUMA level (2000 PUMS)

	1	2	3	4
Variable	Coefficient	Coefficient	Coefficient	Coefficient
Commuting time	0.0370***	0.0384***	0.0322***	0.0354***
	16.28	17.35	14.22	15.86
Average education	1.9874***	1.9318***	1.7627***	1.7117***
	35.49	32.46	30.04	23.19
Occupation diversity	1.7360***	1.5228**	-0.6981	1.4536**
	2.59	2.27	-1.01	2.19
Ownership rate	-0.7681***	-0.6351***	-1.0087***	-0.6137***
	-10.05	-7.09	-13.33	-7.80
Five-year households	0.7481***	0.5856***	0.8979***	0.6593***
	6.07	4.26	7.69	5.44
English proficiency	0.8356***	0.9161***	0.7561***	0.8165***
	11.44	12.02	11.48	11.27
Racial fragmentation	-0.3665***	-0.3425***	-0.2636***	-0.3001***
	-8.18	-7.66	-6.11	-6.76
Presence of children		-1.1083***		
		-3.50		
Unemployment			-5.4254***	
			-8.58	
Bohemian index				7.2527***
				4.83
SS index	**6.8485*****	**5.6648*****	**7.1613*****	**3.4673****
	4.64	3.83	4.58	2.28
Adjusted R^2	0.5702	0.5746	0.5775	0.5763
SSM index	**11.5412*****	**9.9832*****	**11.9630*****	**6.5891*****
SSF index	**3.3724**	**1.8639**	**3.7369**	**0.4118**

Dependent variable: log (housing value). Housing characteristics and 48 state fixed effects are included but not reported. Standard errors are clustered by 1,581 PUMAs. The numbers below the coefficients are t test statistics. Sample size: 2,431,639. Superscripts "*","**" and "***" indicate significance at the 10, 5, and 1% levels.

The first column in Table 2, also the benchmark model, uses the representative social amenity variables and one of the SS, SSM, and SSF indexes. The coefficient of the SS index in column 1, also the semi-elasticity of housing value to the SS index, is 6.8485. This means that, at a residential PUMA, an increase of one percentage point in the proportion of the population that are same-sex couples is associated with, on average, approximately 6.85% higher housing values at that PUMA; a one standard deviation of the SS index is associated with about 2.81% higher housing value.[15] This association is both economically strong and statistically significant (at the 1% level). When the SS index is replaced by the SSM index, the coefficient is 11.5412, and is still significant at the 1% level. However, when replaced by the SSF index, the coefficient, though still positive, becomes much smaller and insignificant: 3.3724, suggesting that gays and lesbians may have different impacts on urban space and housing markets. This is probably due to the gender difference related to urban space and politics. Castells [13] argued that men seek to dominate space, while women attach more importance to networks and relationships. Adler and Brenner [1] found that lesbians tend not to have access to capital, and are more likely to be the primary caretakers of children. Black et al. [7] argued that many of lesbian couples have children present in their household, and on average, lesbian individuals tend to allocate less lifetime resources to "buying" urban amenities. Tables A-2 and A-3 in the appendix show that male and female same-sex couples tend to be clustered in different locations. The summary statistics in Table A-4 also show that on average lesbians have lower income than gay men, and are more likely to have children in the family. Why the concentration of gay men and lesbians has different impacts on urban space deserves further investigation.

Column 2 in Table 2 adds a new variable–the percentage of households with presence of children under four–to proxy for local school quality; and column 3 adds the percentage of unemployed residents at a PUMA to control for income effects. The overall results are very similar. The most interesting finding is

[15] In the PUMA sample, the mean of the SS, SSM, and SSF indexes are 0.0067, 0.0032, 0.0035; and the standard deviations are 0.0041, 0.0029, and 0.0020, respectively.

the result in column 4 where the variable *Bohemian index*–the percentage of bohemian population in a PUMA–is included. Bohemians are believed to play a vital role in neighborhood gentrification and regional housing markets (Godfrey [29]; Florida and Mellander [22]). With the inclusion of the *Bohemian index*, the coefficients of the SS index and the SSM index are driven down by half, possibly due to the contribution of bohemians to housing values or due to the moderate correlation (the correlation coefficients between *Bohemian index* and the SS, SSM, and SSF indexes are 0.45, 0.46 and 0.27, respectively). However, the coefficients of the SS index and the SSM index are still positive and significant at the 5% and 1% levels, respectively. This demonstrates that the contribution of gay population to housing values is not negligible.

The interpretation of human capital, social capital, and cultural capital variables follows that in Fu [23].[16] Since variations in housing value across locations may be simply due to differences in population size, we also estimate all the models in Table 2 by controlling for PUMA size–the total population or total number of households in a PUMA. The coefficients of the PUMA size variable are all positive but not significant at the 10% level, all other coefficients remain pretty much the same.

We also estimate the benchmark model by the census regions, and the similar pattern holds across regions for the SS, SSM, and SSF indexes except the Midwest. The results are presented in Table 3.

[16] The only difference is the sign of the coefficients of home ownership rate and the five-year households. Fu [23] used the restricted version of the 1990 Massachusetts census data and found that the sign of home ownership rate is positive and the sign of the five-year households is negative. The interpretation, there, is that homeowners have a strong incentive to build social capital, and longer tenure implies the weakness of the strong ties. Here, we found the opposite results. One possible explanation could be the size of the data. Fu [23] studied only the Boston metropolitan area.

Table 3
Benchmark model with the gay index by regions (2000 PUMS)

Variable	Northeast Coefficient	Midwest Coefficient	South Coefficient	West Coefficient
SSM index	26.6100***	7.4791*	23.4044***	4.6972**
t statistic	5.73	1.85	7.28	2.08
Adjusted R^2	0.5697	0.5644	0.5312	0.5793
SSF index	1.3235	1.9765	4.6438	6.6188
t statistic	0.19	0.40	1.19	1.43
Adjusted R^2	0.5655	0.5642	0.5278	0.5791
SS index	15.0418***	4.3724	13.3562***	3.7478**
t statistic	3.39	1.46	6.16	2.14
Adjusted R^2	0.5679	0.5644	0.5299	0.5793
Sample size	512,060	523,752	841,668	554,159

Model specification is the same as column 1 in Table 2, but estimated by census regions. Superscripts "*", "**", and "***" indicate significance at the 10, 5, and 1% levels, respectively.

A few PUMAs have disproportionately high concentration of same-sex couples. We experiment dropping the top three and the top ten PUMAs, in terms of the SS index, and re-estimate all the models in Table 2. The results are very similar and in most of the cases, even better: the coefficients of the SS and SSM indexes become a bit larger. This shows that the general results are not driven by a few PUMAs with very high concentration of same-sex couples.

4.2 City level models

Cities are one of the most frequently studied geographic units. The 2000 census 5% PUMS identifies 150 cities that meet the minimum population threshold of 100,000. All the 150 cities are nested within metropolitan areas. By constructing all the variables at the city level, we estimate hedonic housing models at the city level with metropolitan-area fixed effects. The results are presented in Table 4.

Table 4
Hedonic housing models with the gay index at the city level (2000 PUMS)

	1	2	3	4
Variable	Coefficient	Coefficient	Coefficient	Coefficient
Commuting time	-0.0002	-0.0016	-0.0009	-0.0059
	-0.02	-0.17	-0.10	-0.55
Average education	1.7206***	2.0173***	1.5115***	1.4692***
	7.03	9.11	8.00	4.87
Occupation diversity	11.4375***	12.5912***	7.2031***	9.4687***
	4.38	5.24	3.09	4.52
Ownership rate	-0.6017**	-0.9926***	-0.9435***	-0.2941
	-2.38	-3.77	-4.19	-0.99
Five-year households	2.6613***	3.1392***	2.8561***	2.4545***
	4.75	5.64	5.83	4.18
English proficiency	1.6142***	1.4965***	1.1312***	1.5106***
	4.91	5.10	4.48	5.30
Racial fragmentation	-0.5295**	-0.6657***	-0.1011	-0.4748**
	-2.22	-3.01	-0.48	-2.26
Presence of children		3.0858***		
		3.13		
Unemployment			-7.4999***	
			-3.52	
Bohemian index				5.5814*
				1.84
SS index	**19.9937***	**22.3221***	**23.1566***	**19.575***
	3.42	4.44	4.79	2.81
Adjusted R^2	0.5188	0.5192	0.5196	0.5191
SSM index	**32.4686***	**34.4031***	**40.1240***	**30.2815***
SSF index	**18.6354***	**23.0372****	**18.7768****	**21.4260***
SS index-IV*	30.0353	15.7963	52.0280**	13.5825^{ns}
	1.34	0.91	2.00	0.62

Dependent variable: log (housing value). Housing characteristics and 104 metropolitan area fixed effects are included. Standard errors are clustered by 150 cities.
The numbers below the coefficients are t test statistics. Sample size: 372,949.
Superscripts "*", "**", and "***" indicate significance at the 10, 5, and 1% levels.
*: SS index-IV is the instrumental variables estimator.

Table 4 shows that the coefficient of average commuting time, though not significant, has a negative sign, as the theoretical spatial models predict. The coefficients of the *SS* index are much larger than those at the PUMA level: between 19 and 23, and significant at the 1% or 5% level. This suggests that the city level models with metropolitan-area fixed effects work better, possibly because the city sample includes only cities with at least 100,000 residents, which makes the measurement error problem less serious. The coefficients of *SSM* and *SSF* indexes are all positive and significant at the 1% level and the 10% level, respectively, in all model specifications. We also estimate all the models by controlling for city size. The coefficients of city population size are all positive and significant at the 1% level, and the coefficients of all other variables remain almost the same.[17,18]

The sorting theory argues that gay men sort into high-amenity locations or locations of tolerant social milieu. Such spatial sorting can cause the gay index endogenous in cross-section models: the ordinary least squares (OLS) estimator will not be consistent and the causality cannot be identified. To deal with this issue we try to use instrumental-variables (IV) estimation. One instrumental variable we find for the *SS* index is a dummy variable, *Law*, indicating whether a city has passed a law to prevent discrimination in public employment based on sexual orientation by the year 2000. In the city sample, there are 52 cities that passed at least a law prohibiting discrimination in public employment on the basis of sexual orientation.[19] In Table 4, the row with the variable "*SS* index-*IV*" presents the results of instrumental-variables estimation. The coefficients

[17] Again, we experiment dropping the top five and the top ten cities in terms of the *SS* index and re-estimate the models in Table 4. The results, again, are very similar and in most of the cases, even better: the coefficients of the *SSM* and *SS* indexes become a bit larger. This shows that the general results are not driven by a few cities with a high concentration of same-sex couples.

[18] Government regulation restricts housing supply, thus resulting in higher housing prices, especially in coastal cities (Glaeser, Gyourko, and Saiz [26, 27]). Most of the top 10 cities and PUMAs are located on the eastern and western U.S. coasts. The fact that dropping those cities does not change the model results suggests that even after controlling for housing supply regulation, the results are still robust.

[19] The data are from a table "States, Cities and Counties with Civil Rights Ordinances, Policies or Proclamations Prohibiting Discrimination on the Basis of Sexual Orientation," available at www.thetaskforce.org. The correlation coefficients between *Law* and housing value, the *SSM*, *SSF*, and *SS* indexes are 0.10, 0.39, 0.34 and 0.41, respectively.

are all positive, but not significant, except the one in column 3 (significant at the 10% level). The Hausman test statistics indicate that there is no systematic difference between the coefficients of OLS and IV estimators in column 1, 2, and 3 specification, which means that the OLS estimator is consistent and efficient.[20] We also perform the Davidson-MacKinnon augmented regression test for all the models and found that in all cases the SS index can be considered exogenous.[21] Therefore, we conclude that endogeneity is not a serious issue even the spatial sorting might exist.

We also estimate the PUMA and city level models using the 5% PUMS of the 1990 census. The pattern of the results is similar (the results are not reported here).[22]

5 Further identification issues and suggestive causality

The cross-section models show that the partial correlation between the gay index and housing values is significant and robust. However, possible unobserved characteristics of same-sex couples and omitted location attributes may generate endogeneity problems. We cannot identify the causality between housing values and the spatial location of same-sex couples because it could be the case that same-sex couples sort themselves into a particular city or residential PUMA, based on their location preferences, personal characteristics, or location-specific attributes. The case studies on residential communities gentrified by gay people suggest that it is gay people who have improved their neighborhoods, and have made their communities better, and not the case that gay people choose to move into high-quality communities. However, special cases probably cannot be generalized to justify the general relationship between property values and

[20] Column 4 model fitted on the data fail to meet the asymptotic assumptions of the Hausman test.

[21] We first treat the SS index as endogenous and dependent on housing price, *law*, and other neighborhood attributes, then perform the Davidson-MacKinnon exogeneity test (Davidson-MacKinnon [17], pp.237-242).

[22] In the 1990 census PUMS, there are 727 heads of households who reported more than one unmarried partner in the PUMA sample, 293 in the city sample. Those people are not included in the gay index.

the spatial distribution of gay people.

In the previous section the IV estimator and the exogeneity test suggest that the endogeneity of the gay index is not a serious problem. Here, to control for unobserved city attributes, we use city-fixed effects model. By using the 5% PUMSs of both the 1990 and 2000 census we construct a panel data set. Although PUMA information is available in both the 1990 and 2000 PUMSs, the PUMA boundaries are not comparable. Fortunately, most of the identified cities remain within the same boundaries, which enables us to estimate a panel data model with city fixed effects.[23] Since there is no way to identify whether a housing unit was surveyed in both the 1990 and 2000 census, we can only construct a variable to measure the general housing price level in a city. We have not found a city-level housing price index, so we use median housing value.[24]

Two types of models are specified using the panel data set. The first is the city fixed effects model, using the logarithm of median housing value in each city each year as the dependent variable. The model is specified as follows:

$$\log MP_{tj} = \alpha + \lambda_j + t_{20} + \gamma' X_{tj} + \epsilon_{tj}, \tag{4}$$

where MP_{tj} is the median reported housing value in city j in year t; α is a constant; λ_j is a city fixed effect, controlling for all unobservable, time-independent city-specific attributes, such as land use regulation and adult amenities; X_{tj} is the attributes vector of social amenities in city j in year t, including variables measuring human capital, social capital, cultural capital, and the gay index; γ is the coefficient vector to be estimated; ϵ_{tj} is the disturbance term. The reported housing values are nominal. Since the city-level consumer price index is not available, a time dummy $t20$ (=1 if year=2000) is added to control for year-specific shocks. Table 5 presents the results of the city fixed effects model.

[23] The IPUMS provides detailed city codes which allow the identification of some cities that merged with others in the past, but most users will probably find this extra detail unnecessary.

[24] Median housing value is not a perfect measure since the distribution of housing values may change over time. We also tried the mean housing value in a city, and the results are similar. Florida and Mellander [22] used the MSA median housing values to study the effect of bohemians, artists and gays on housing values.

Table 5
City fixed effects model

	1	2	3	4
Variable	Coefficient	Coefficient	Coefficient	Coefficient
$t20$ (Time dummy)	0.0401**	-0.0151	-0.0070	0.0431**
	1.96	-0.55	-0.34	2.04
Average education	2.4002***	1.9263***	2.0216***	2.3802***
	12.30	9.18	8.44	8.58
Occupation diversity	7.7841***	7.7873***	6.9065***	7.7855***
	5.52	5.21	4.03	5.34
Home ownership rate	0.7230***	1.1525***	0.4683***	0.7077***
	5.02	6.28	2.78	4.91
Five-year households	-0.4580***	-1.0716***	-0.3364**	-0.4770***
	-2.63	-3.94	-2.34	-2.73
English proficiency	3.8531***	4.2422***	3.7256***	3.8428***
	39.41	31.65	34.41	34.83
Racial fragmentation	0.1113	0.2531**	0.0923	0.1025
	1.04	2.21	0.82	0.94
Presence of children		-3.4032***		
		-4.19		
Unemployment			-3.1547***	
			-2.99	
Bohemian index				0.4485
				0.14
SS index	**5.5452***	**1.5306**	**13.0255*****	**5.9222***
	1.71	0.39	3.38	1.74
Wald χ^2	5487	2824	3077	5115
SSM index	**16.8980*****	**10.7860****	**22.6929*****	**17.9222*****
SSF index	**-12.1698**	**-19.0946*****	**-7.4421**	**-13.2971**

Dependent variable: log(median housing value in a city). Generalized least squares estimator, corrected for heteroskedasticity. Numbers below the coefficients are z values. Superscripts "*", "**", and "***" indicate significance at the 10, 5, and 1% levels, respectively. Fixed effects: 92 cities. Sample size: 184.

The models in Table 5 are estimated first by using the linear least squares method. The modified Wald statistics show that there exists group-wise heteroskedasticity. We then use the generalized least squares estimate to correct for within-panel heteroskedasticity. After controlling for unobservable city-specific attributes, the coefficients of the SSM index are still positive and significant at the 1% or 5% level. The coefficients of the SS index are positive and significant at the 5% or 10% level in three of the four models. The coefficients of the SSF index are negative and not significant in three of the four models, which again suggests that gays and lesbians may have different impacts on housing markets. We also re-estimate all the models by controlling for city size. The coefficients of city size are all positive and significant at the 1% level; the coefficients of the gay indexes remain similar but the magnitude attenuates a little bit. Since we have only two time periods (1990 and 2000), there is no way to correct for within-panel autocorrelation and cross-panel correlation at the same time; the results in Table 5 are rather experimental.

The second model tests the Granger causality. The logarithm of the median housing value in a city in 2000 is regressed on the lagged city attributes in 1990. The model is specified as

$$\log MP_{2000,j} = \alpha + \beta \log MP_{1990,j} + \gamma' X_{1990,j} + \epsilon_{2000,j}, \tag{5}$$

where $MP_{1990,j}$ and $MP_{2000,j}$ are the median reported housing values in city j in year 1990 and 2000, respectively; α is a constant; $X_{1990,j}$ is the attributes vector of social amenities in city j in year 1990. Since the data are available for only two time periods, city fixed effects have to be dropped. We add the state fixed effects since the boundaries of some metropolitan areas changed. The results are presented in Table 6.

Table 6
A Granger causality test

Variable	1 Coefficient	2 Coefficient	3 Coefficient	4 Coefficient
Log median housing value	0.7178***	0.7221***	0.7111***	0.7107***
	10.10	9.60	9.59	9.34
Average education	0.9456***	0.9991**	0.9035***	0.8823***
	2.83	2.22	2.59	2.65
Occupation diversity	-1.9215	-2.0681	-2.3126	-2.1184
	-0.83	-0.84	-0.82	-0.87
Home ownership rate	-0.0233	-0.0688	-0.0397	-0.0072
	-0.12	-0.30	-0.21	-0.04
Five-year households	1.0997***	1.1793***	1.1126***	1.0757***
	5.64	2.86	5.60	5.16
English proficiency	0.4905	0.4433	0.4938	0.4875
	1.40	1.04	1.41	1.40
Racial fragmentation	-0.2818*	-0.2946*	-0.2637	-0.2959*
	1.70	-1.68	-1.36	-1.74
Presence of children		0.3747		
		0.25		
Unemployment			-0.5666	
			-0.25	
Bohemian index				1.9622
				0.45
SS index	**9.3631***	**9.9615***	**9.4916***	**9.0266**
	2.62	2.71	2.64	2.38
Adjusted R^2	0.94	0.94	0.94	0.94
SSM index	**10.5746**	**11.0467**	**10.6777**	**10.0714*
SSF index	**27.4304**	**28.0087*	**27.9877**	**27.1163**

Dependent variable: log (median housing value in a city in 2000). Sample size:92.
Superscripts "*", "**", and "***" indicate significance at the 10, 5, and 1% levels.
All independent variables are at the 1990 level. State fixed effects are included.

The results in Table 6 are encouraging: not only past housing values, the quality of human capital stock, the percentage of households with longer tenure, and racial fragmentation, but also the proportion of same-sex couples in a city are good predictors of future median housing values. Since future events cannot be used to predict past events, the results in Table 6 suggest that it is, indeed, the case that same-sex couples "*cause*" property values to increase. For example, after controlling for housing value and other city attributes in 1990, a one percentage point increase in the proportion of the population comprised of same-sex couples in a city in 1990 can generate approximately a 9% increase in median housing value in the year 2000. After controlling for city sizes in 1990, the coefficients of the SS, SSM, and SSF indexes even become slightly larger.

Equation (5) can be easily modified to be the growth convergence model: By subtracting $\log MP_{1990,j}$ from the both sides, the model is reduced to a regression of the growth rate of median housing value on the lagged city attributes. Therefore, the results from Table 6 also imply that the growth of median housing value is faster where the proportion of same-sex couples is higher one decade ago. The results are consistent with the life-cycle theory of gentrification.

6 A passion to preserve?

Sections 4 and 5 provide some suggestive evidence that the spatial sorting bias due to the preference towards adult amenities or tolerant social milieu is not a serious issue, implying that gay people *make*, rather than choose to live in, better neighborhoods. If gay people were randomly distributed across locations, then what could motivate gay people to make better neighborhoods everywhere? Obviously, the data themselves cannot reveal such information.

Castells [13] provided an interesting description of how gay people gentrified the Castro neighborhood. Being discriminated against in the housing market, gay realtors and interior decorators discovered the hard way how to survive in the tough San Francisco housing market: They used their commercial and artistic skills, bought housing units in low-cost areas, repaired and renovated them, and resold them for profit. Fellows [19] argued that gay men are very sensitive to beauty, and have long been impassioned keepers of culture: restoring decrepit buildings, revitalizing blighted neighborhoods, etc. If gay people, indeed, have stronger aesthetic tastes than (or other attributes different from) heterosexuals, then the intrinsic preference hypothesis is consistent with our findings.

However, whether gay people have stronger innate aesthetic or artistic tastes than heterosexual people is not a consensus. Crimmins [16] described, in a non-technical way, how gay people's aesthetic preference in many fields, from fashion to housing, has shaped the mainstream American pop culture. Lewis and Seaman [37] used the 1993 and 1998 General Social Survey data and tested the relationship between sexual orientation and the demand for arts. They found that gay people are much more likely to attend the arts than demographically similar heterosexuals, but do not demonstrate higher innate creativity through greater amateur production of arts.

Table A-4 uses the PUMA sample drawn from the 2000 census 5% PUMS and presents a set of summary statistics for the male and female same-sex couple samples, and the heterosexual sample. The results show that gay people are, indeed, more likely to work as artists, to choose bohemian occupations, and to receive more school education. How to explain gay people's stronger affinity for the arts is worthy of further investigation.

7 Discussion and conclusion

This paper, to the best of our knowledge, is the first to identify the general relationship between housing values and the spatial distribution of same-sex couples across the U.S. The results show that not only is the correlation between the spatial concentration of same-sex couples and housing values significant and robust, but, also, housing values are higher in a city where the proportion of same-sex couples was higher a decade ago. Therefore, we tentatively conclude that same-sex couples *make* better communities. The results are consistent with the intrinsic preference theory that the intrinsic artistic tastes of gay people motivate them contribute to nicer neighborhoods everywhere and all the time. We recognize that our findings are provocative. Given the current state of theory and data in this arena, it is impossible to conduct direct tests of the causal links. Therefore, our results represent only an initial step towards exploring the potentially complex relationship between property values and the spatial location of gay population.

We also recognize the major drawbacks in this study. First, even though the 1990 and 2000 census PUMSs are the most comprehensive and systematic data publicly available for the economic study of sexual orientation, the data quality is still not absolutely guaranteed. There exist undercount, measurement error, and report error. Badgett and Rogers [4] discussed, in detail, the possible ways of causing undercount. The count of same-sex unmarried-partner households increased in the U.S. from 145,130 in 1990 to 594,691 in 2000 (Badgett and Rogers [4]). For a location with an increase in its same-sex couple population in 2000, one possibility could be that more gay people who lived there in 1990 filled out the 2000 census survey as same-sex unmarried partners, while they did not do this in the 1990 census. We are not certain how serious this issue is. Black et al. [6, 9]) discussed the measurement and record error of the 1990 and 2000 census PUMSs. The measurement error, however, is not a problem with regard to our conclusion, as measurement error causes the coefficients of independent variables to be underestimated.

Second, the neighborhood in this study is defined at the macrogeographic levels: the PUMA level and the city level. Since social interactions and neighborhood externalities are very localized, using macrogeographic level data will result in some bias. To better control for unobservable neighborhood attributes, microgeographic level data are highly desired. The ideal data sets are the restricted version of the census data, from which the PUMSs are drawn, containing one-sixth of the households in the U.S., with detailed microgeographic information down to the census block levels. In general, the results from the block level is expected to be qualitatively similar to those from the PUMA or city level. A good example is that human capital externalities are significant at the census block level as well as at the MSA level (Fu [24]). Also, housing studies at macrogeographic levels are common. For example, the housing price index constructed by the Office of the Federal Housing Enterprise Oversight is at the metropolitan area level. In this study, if the gay index is a proxy for tolerant social milieu, then it applies to macrogeographic levels; if the gay index implies the intrinsic preference towards aesthetics, then, it is independent of locations. In both cases, the qualitative effects of gay index on housing values should remain similar regardless of the geographic levels.

Finally, a theory that identifies the causality links would be to model the process of gentrification, based on the intrinsic preference of gentrifiers. One possible way might be to introduce the heterogenous preferences towards housing maintenance into housing filtering models.

Acknowledgement

This study benefits tremendously from conversations with many people. Particularly, the author is deeply indebted to Richard Arnott for his constant encouragement and critical comments. The author also would like to thank Stephen Ross, Vernon Henderson, Bo Zhao, Jed Kolko, Jim Davis, Aimee Wagenen, Judith Pereira, Man Jia, two anonymous referees, and seminar participants at the Southwestern University of Finance and Economics (China), and the China Center for Economic Research at Peking University very much for the very helpful comments and suggestions.

Appendix

Table A-1
Some geographic summary statistics in the 2000 PUMS

Obs.	Variable	Mean	Std. Deviation	Minimum	Maximum
48	Number of PUMAs in a state	33	42	1	224
48	Number of housing units in a state*	82,543	108,441	1,641	586,232
1581	Number of housing units in a PUMA	2,506	737	1,071	6,569
150	Number of PUMAs in a city	2	5	1	55
150	Number of housing units in a city	5,501	12,250	1,422	129,781

*: The total number of housing units is unweighted.

Table A-2
Top ten cities of same-sex couples in the 2000 PUMS

City	SS index	SSM index	SSF index	Housing units
San Francisco, CA	2.985	2.160	0.825	15,578
Seattle, WA	2.439	1.303	1.136	12,176
Minneapolis, MN	1.734	0.821	0.912	5,956
Washington, DC	1.690	1.235	0.455	14,201
Cambridge, MA	1.626	0.707	0.909	2,819
Long Beach, CA	1.595	0.931	0.665	7,899
Boston, MA	1.536	0.871	0.665	13,225
Alexandria, VA	1.420	0.861	0.559	3,086
New Orleans, LA	1.411	0.804	0.607	8,708
Salt Lake City, UT	1.391	0.636	0.755	3,162

The unit of the SS, SSM, and SSF indexes is %.

Table A-3

Top ten PUMAs of same-sex couples in the 2000 PUMS

State	PUMA ID	SS index	SSM index	SSF index	Housing units
CA	02204	8.040	5.643	2.397	2,410
NY	03810	4.261	3.344	0.917	2,771
WA	01804	4.097	2.269	1.828	2,065
NY	03807	3.764	3.340	0.424	2,946
DC	00105	3.342	3.042	0.299	3,094
CA	08004	3.236	2.653	0.582	5,141
GA	01201	3.191	1.799	1.393	2,481
GA	01104	3.019	1.937	1.082	2,626
CA	02403	2.876	1.195	1.681	2,910
TX	02302	2.811	2.222	0.589	2,639

The unit of the SS, SSM, and SSF indexes is %.

Table A-4

Summary statistics from the PUMA sample: Mean or frequency

Variable	Same-sex male	PUMA Sample same-sex female	Heterosexuals
Total personal income	$41,416	$34,386	$30,352
Wage and salary income	$3,4436	29,356	$25,484
White	79.07%	77.57%	74.51%
Age	40.4	39.7	38.5
Home owner	63.34%	64.8%	65.40%
College degree or above	35.53%	35.39%	24.33%
Employed	77.37%	77.23%	69.24%
Lived more than 5 years	41.66%	40.96%	50.51%
Presence of children under 4	5.55%	8.11%	13.63%
Management occupation	16.54%	13.87%	11.04%
Service occupation	12.22%	13.44%	13.92%
Office occupation	11.07%	14.49%	13.89%
Sales occupation	11.07%	9.07%	10.22%
Education occupation	3.95%	6.36%	4.59%
Health occupation	3.94%	6.02%	3.73%
Artistic occupation	4.03%	2.87%	1.76%
Bohemian	3.1%	1.9%	1.13%

The size in the male and female same-sex couple samples, and the heterosexual sample are 22,642, 23,040, and 6,431,343, respectively. Only people of primary age (16-65) are included. The samples are drawn from the 2000 census PUMS.

Table A-5
1990 and 2000 PUMS occupation code

Occupation	1990 code	2000 code
Management, business operation, professional	1-42	1-95
Engineers, architects, surveyors	43-63	130-156
Mathematical, computer scientists	64-68	100-124
Natural scientists	69-83	160-176
Health	84-112	300-354
Teachers, librarians, education	113-165	220-255
Social scientists, legal service	166-173	180-186, 210-215
Social service	174-182	200-206
Writers, artists, entertainers, athletes	183-202	260-292
Technicians	203-242	190-196
Sales	243-302	470-496
Administrative, office	303-402	500-593
Service	403-472	360-465
Agriculture,forestry, fisheries	473-502	600-613
Mechanics, repairers, precision	503-552, 628-702	700-762
Construction	553-612	620-676
Mining, extraction	613-627	680-694
Machine operators, production	703-802	770-896
Transportation, movers	803-863	900-975
Handlers, equipment cleaners, laborers	864-902	
Military	903-908	980-983
Unemployed, others	>908	0, >983
Bohemian	183, 185-194	260, 263, 270, 271 274-276, 285, 291

References

[1] S. Adler, J. Brenner, Gender and space: Lesbians and gay men in the city, International Journal of Urban and Regional Research 16 (1992) 24-34.

[2] A. Alesina, R. Baqir, W. Easterly, Public goods and ethnic divisions, The Quarterly Journal of Economics 114 (1999) 1243-1284.

[3] S. Allegretto, M. Arthur, An empirical analysis of homosexual/heterosexual male earnings differentials: Unmarried and unequal? Industrial and Labor Relations Review 54 (2001) 631-648.

[4] M.V.L. Badgett, M. Rogers, Missing same-sex couples in census 2000, Working Paper, Institute for Gay and Lesbian Strategic Studies, University of Massachusetts-Amherst, 2003.

[5] G. Becker, A treatise on the family, Harvard University Press, Cambridge, 1981.

[6] D. Black, G. Gates, S. Sanders, L. Taylor, Demographics of the gay and lesbian population in the United States: Evidence from available systematic data sources, Demography 37 (2000) 139-154.

[7] D. Black, G. Gates, S. Sanders, L. Taylor, Why do gay men live in San Francisco? Journal of Urban Economics 51 (2002) 54-76.

[8] D. Black, G. Gates, S. Sanders, L. Taylor, The earnings effects of sexual orientation, Industrial and Labor Relations Review 56 (2003) 449-469.

[9] D. Black, G. Gates, S. Sanders, L. Taylor, The measurement of same-sex unmarried couples in the 2000 U.S. Census, unpublished paper, 2006.

[10] J. Blandford, The nexus of sexual orientation and gender in the determination of earnings, Industrial and Labor Relations Review 56 (2003) 622-642.

[11] J. Brueckner, S. Rosenthal, Gentrification and neighborhood housing cycles: Will American's future downtown be rich? Working Paper, University of California, 2005.

[12] C. Carpenter, New evidence on gay and lesbian household incomes, Contemporary Economic Policy 22 (2004) 78-94.

[13] M. Castells, The city and the grassroots, University of California Press, Berkeley and Los Angeles, 1983.

[14] M. Castells, K. A. Murphy, Culture identity and urban structure: The spatial organization of San Francisco's gay community, in: N. Fainstein and S. Fainstein (Eds), Urban policy under capitalism, Sage Publications, Beverly Hills, 1982, pp. 237-259.

[15] J. Coleman, Social capital in the creation of human capital, The American Journal of Sociology 94 (1988), S95-S120.

[16] C. E. Crimmins, How the homosexuals saved civilization: The true and heroic story of how gay men shaped the modern world, Putnam Pub Group, 2004.

[17] R. Davidson, J. MacKinnon, Estimation and inference in econometrics, Oxford University Press, Oxford, 1993.

[18] D. DiPasquale, E. Glaeser, Incentives and social capital: Are homeowners better citizens? Journal of Urban Economics 45 (1999) 354-384.

[19] W. Fellows, A passion to preserve: Gay men as keepers of culture, The University of Wisconsin Press, Madison, 2004.

[20] R. Florida, Bohemian and economic geography, Journal of Economic Geography 2 (2002) 55-71.

[21] R. Florida, Cities and the creative class, Routledge, New York, 2005.

[22] R. Florida, C. Mellander, There goes the neighborhood: How and why bohemians, artists and gays effect regional housing values, Working Paper, George Mason University, 2007.

[23] S. Fu, What has been capitalized into property values: Human capital, social capital, or cultural capital? Working Paper CES-WP-05-25, U.S. Census Bureau Center for Economic Studies, 2005.

[24] S. Fu, Smart café cities: Testing human capital externalities in the Boston metropolitan area, Journal of Urban Economics 61 (2007) 86-111.

[25] G. Gates, J. Ost, The gay & lesbian atlas, The Urban Institute Press, Washington, D.C., 2004.

[26] E. Glaeser, J. Gyourko, A. Saiz, Why is Manhattan so expensive? Regulation and the rise in house prices, NBER Working Paper 10124, 2003.

[27] E. Glaeser, J. Gyourko, A. Saiz, Why have housing prices gone up? NBER Working Paper 11129, 2005.

[28] E. Glaeser, J. Kolko, A. Saiz, Consumer city, Journal of Economic Geography 1 (2001) 27-50.

[29] B. Godfrey, Neighborhoods in transition: The making of San Francisco's ethnic and nonconformist communities, University of California Press, Berkeley, 1988.

[30] J. Kain, J. Quigley, Housing market discrimination, home-ownership, and savings behavior, American Economic Review 62 (1972) 263-277.

[31] T. King, P. Mieszkowski, Racial discrimination, segregation, and the price of housing, Journal of Political Economy 81 (1973) 590-606.

[32] M. Klawitter, Why aren't more economists doing research on sexual orientation? Feminist Economics 4 (1998) 55-59.

[33] L. Knopp, Gentrification and gay community development in a New Orleans neighborhood, Ph.D. Dissertation, Department of Geography, The University of Iowa, 1989.

[34] L. Knopp, Gentrification and gay neighborhood formation in New Orleans: A case study, in: A. Gluckman and B. Reed (Eds), Homo economics, Routledge, New York and London, 1997, pp. 45-63.

[35] J. Kolko, The economics of gentrification, Working Paper, Department of Economics, Harvard University, 2006.

[36] M. Levine, Gay ghetto, Journal of Homosexuality 4 (1979) 363-377.

[37] G. Lewis, B. Seaman, Sexual orientation and demand for the arts, Social Science Quarterly 85 (2004) 523-538.

[38] L. Lochner, Education, work, and crime: A human capital approach, NBER Working Paper 10478, 2004.

[39] J. Moorehouse, M. Smith, The market for residential architecture: 19th century row houses in Boston's South End, Journal of Urban Economics 35 (1994) 267-277.

[40] M. Moss, Reinventing the central city as a place to live and work, Housing Policy Debate 8 (1997) 471-490.

[41] S.O. Murray, American gay, University of Chicago Press, Chicago, 1996.

[42] J. Palen, B. London (Eds), Gentrification, displacement and neighborhood revitalization, State University of New York Press, New York, 1984.

[43] R. Putnam, Bowling alone: The collapse and revival of American community, Simon & Schuster, New York, 2000.

[44] S. Ross, The continuing practice and impact of discrimination, Working Paper 2005-19, Department of Economics, University of Connecticut, 2005.

[45] N. Smith, P. Williams (Eds), Gentrification of the city, Allen & Unwin, Boston, 1986.

[46] C. Stack, 1974, All our kin, Harper & Row, New York, 1974.

[47] J. Yinger, The black-white price differential in housing: Some future evidence, Land Economics 54 (1978) 187-206.

Appendix (Not for publication)

Geographical levels in the decennial census data

There are four different geographic levels in the 1990 and 2000 census PUMS: state, metropolitan area, city, and PUMA (or Super PUMA).

Briefly speaking, metropolitan areas are counties or combinations of counties centering on a large population center (a substantial urban area) that have a high degree of economic and social interaction with that center. Before 1950, the Census Bureau did not define metropolitan areas. Though the concept of a metropolitan area has remained essentially the same over time, there has been slight change. For example, the boundaries of each metropolitan area may have been adjusted; new metropolitan areas regularly emerged; the definition varies slightly.

Metropolitan areas have been referred to by several different names. In 1950, the term was Standard Metropolitan Area (SMA). In 1970 and 1980, it was called Standard Metropolitan Statistic Area (SMSA). In 1990 and 2000, the term was Metropolitan Area (MA), including:

1. Free-standing Metropolitan Statistical Areas (MSAs), which are generally surrounded by non-metropolitan territory, and, therefore, are not integrated with other metropolitan areas, and

2. Primary Metropolitan Statistical Areas (PMSAs), which are the same as MSAs except that they are near, and economically/socially linked to, other PMSAs. Two or more adjacent PMSAs form Consolidated Metropolitan Statistical Areas (CMSAs). Many PMSAs were separate SMSAs or SMAs before 1990.

A metropolitan area may span across state boundaries. For example, Lawrence-Haverhill crosses both Massachusetts and New Hampshire.

The Public Use Microdata Area (PUMA) is a geographic area of at least 100,000 residents, as defined by the Census Bureau. In the 1990 PUMS, a PUMA follows the boundaries of a central city (the largest city within a metropolitan area), a PMSA, or a non-metropolitan place. If the population of one of these entities exceeded 200,000 at that time, then the Census Bureau split the area into as many PUMAs of 100,000+ residents as possible.

A PUMA may be a portion of a central city. In the majority of cases, PUMAs are nested within a metropolitan area. If a PUMA crosses a metropolitan area boundary, then, the metropolitan area households located in that PUMA do not receive the relevant MA code, and that MA is only partially identified. In the 1990 PUMS, PUMAs sometimes cross state lines; but for all the PUMAs that are nested within metropolitan areas, none crosses a state boundary. The PUMAs in the 2000 census do not cross state lines. There are 1,726 and 1,581 PUMAs identified in the metropolitan area in the 5% sample of the 1990 and

2000 census PUMS, respectively. For the 1990 and 2000 5% census PUMS, the PUMA is the lowest geography level.[25]

In 1990, cities are identified when at least 99% of the PUMA residents lived in a given city and no more than 1% of the PUMA residents lived outside the city limits (there are a few exceptions). As in the 5% sample of the 2000 census PUMS, only cities meeting the minimum population threshold of 100,000 are identified. There are 126 and 150 cities in the 5% sample of the 1990 and 2000 census PUMS, respectively. The IPUMS provides detailed city codes which allow the identification of some cities that merged with others in the past, but most users will probably find this extra detail unnecessary.

The restricted version of the census data, also called the long form, contains one-sixth of the households in the U.S. The long form data contain detailed microgeographic information down to the census tract and census block levels.

A census tract is a small, relatively permanent statistical subdivision of a county delineated by a local committee of census data users for the purpose of presenting data. A census tract averages about 4,000 persons. In the 2000 census, the tract population criterion is 1,500 to 8,000 persons. Census tracts are designed to be relatively homogenous units with respect to population characteristics, economic status, and living conditions at the time of establishment.

A census block is a subdivision of a census tract. A block is the smallest geographic unit for which the Census Bureau tabulates 100% data. Census blocks are small areas bounded on all sides by visible features such as streets, roads, streams, and county limits.

A block group is a cluster of blocks having the same first digit of their identifying numbers within a census tract. A block group generally contains between 250 and 550 housing units, with the ideal size being 400 housing units. A block group is the smallest geographic unit for which the Census Bureau tabulates sample data. (Source: IPUMS and U.S. Census Bureau web sites.)

[25] A Super Public Use Microdata Area (Super-PUMA) is a geographic area with 400,000 or more residents within a state. In the 1% sample of the 2000 census PUMS, a Super-PUMA is the smallest geographical division available. Super-PUMAs do not cross state lines.