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

November 14, 2006

Abstract

This study is the first to identify 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 strong 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. However, more rigorous study requires the restricted version of the U.S. census data.

JEL Classifications: A14; C21; C23; J15; R31

Keywords: Same-sex couple; Hedonic model; Gentrification; Gay index

* Tel.: 86-028-87099206; fax: 86-028-87099193.

e-mail address: fush@swufe.edu.cn.

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, 1997). 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 the 1990 and 2000 census 5% Public Use Microdata Samples and introduces the gay index into the social-amenity-based hedonic housing models. The tentative conclusion is that same-sex couples *make* better communities.

1 Introduction

Since 1960s, gay people started moving into decaying neighborhoods in inner cities as less advantaged citizens. Those residential neighborhoods with concentration of gay people were called "gay ghetto" by sociologists (Levine, 1979).[1] Case studies demonstrate that since then gay people did a lot of rehabilitations and restoration to develop their communities. Property values in those neighborhoods have been increasing much faster than average. For example,

[1] Levine's definition of 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.

Castells and Murphy (1982) and Castells (1983) studied how gay men as gentrifiers developed the gay community in San Francisco. Knopp (1997) studied the Marigny neighborhood in New Orleans. Marigny was experiencing disinvestment and slum landlordism in 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. The similar story happened in Boston's South End. South End was a run-down neighborhood a few decades ago. Since a few gutsy 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 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 Statistic Area (MSA).

Table 1
Median housing values in the selected tracts in 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

Note. 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 to high-quality neighborhoods.

1. 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 their 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. (2002) used this economic approach to explain why gay men live in San Francisco. Another different sorting theory argues that gay people choose to reside where the social milieu and political environment are tolerant and friendly to gays (Murray, 1996). The gay index, the proportion of 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, 2002; 2005).

The sorting theory can not 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, homogenous heterosexual neighborhoods instead of gay communities.

Sorting also raises the endogeneity problem in cross-section models: the ordinary least squares (OLS) estimator will not be consistent and the causality can not be identified. For example, housing rents in cities rich in consumer amenities can grow fast (Glaeser, Kolko, Saiz, 2001), this might suggest that the correlation between the spatial distribution of gay people and property values may be spurious.

2. Gay politics. In the case study of San Francisco's gay community, Castells (1983) argued that gay men struggled for survival. They formed space clusters to vote, to gain social recognition and political power. Knopp (1997) documented a neighborhood-based political action in Marigny: Gay people there founded the Faubourg Marigny Improvement Association 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 can not be extended to communities where gay population is not dominant.

3. Gentrification theory. Gentrification models (Palen and London, 1984; Smith and Williams, 1986) 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 (1983) and Knopp (1997) indeed tell us that it is gay people that improved their neighborhoods, 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 less advantaged group, probably not as real estate investors or speculators. The gentrification theory can describe the dynamic process of gay community devel-

[2] Recent studies on gentrification employed this concept (Brueckner and Rosenthal, 2005; Kolko, 2006).

opment; however, it is not clear what the incentives of gay people as gentrifiers are.

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? Anecdotal evidence shows that gay people have strong aesthetic tastes that help them identify the charm and profitability of run-down housing units. Godfrey (1988) described a three-stage life-cycle theory of gentrification: bohemian influx, middle class transition, and bourgeois consolidation, which suggests that future property values are correlated with the past gay population.[3]

4. Housing market discrimination to gay people. In 1960s some studies on housing market discrimination to black people show that black households paid more than white households for identical bundles of residential services (King and Mieszkowski, 1973; Yinger, 1978). The discrimination markups are mainly due to the supply restrictions that less housing units are available outside ghettos to black households when they are discriminated. 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 white-owned housing units (Kain and Quigley, 1972).

[3] 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 a 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.

[4] Another form of discrimination against racial minority is steering–black homebuyers are shown houses in systematically different neighborhoods than those shown to comparable white homebuyers (Turner and Mikelsons, 1992). Does the housing market discrimination to blacks also apply to sexual orientation? Legal studies do provide evidence of housing market discrimination to 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. Further studies on housing market discrimination based on sexual orientation will mainly depend on the availability of new 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? Fellows (2004) documented 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, saving artifacts of historical significance, etc. It is the stronger aesthetic tastes that gay people have that have made their neighborhoods nicer and better. Freelance Crimmins (2004) even declared that gay people play a very important role in shaping the modern American pop culture. Unfortunately, the intrinsic

[4] The studies from 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 (2005) 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.

preference theory is hard to put into empirical test, as it is very difficult to construct data to measure artistic tastes of people, artistic characteristics of buildings, and household expenditures on housing decoration and arts. However, as we shall see latter, the empirical results of this paper provide indirect evidence to support this theory.

If gay people indeed can make better neighborhoods compared with heterosexuals, then, after controlling for housing characteristics and other neighborhood attributes, housing prices in the neighborhood with proportionately more gay people should be higher than those with proportionately less gay people; and housing prices in the 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 people. To be more surprisingly, even though economists have done extensive studies on the racial and gender minorities, such as discrimination against African-Americans and women, few have been interested in sexual minorities, although Nobel laureate Gary Becker started in early 1980s (Becker, 1981). Klawitter (1998) explained why there are so few economists interested in research on issues of sexual orientation despite the cultural, political, and economic importance. The causes include discrimination against sexual minorities, the absence of support for this research, and the scarcity of appropriate models and data. In all the issues of *Journal of Urban Economics*, only one paper (Black et al., 2002) studies why gay men live in San Francisco by using the 1990 Public Use Microdata Sample

(PUMS). Florida (2002) constructed a gay index to proxy for the openness and diversity of urban social milieu to study the spatial distribution of creative class across metropolitan areas. A few labor economists studied the wage gap between homosexual and heterosexual people (Allergetto and Arthur, 2001; Carpenter, 2004; Black, et al., 2003; Blandford, 2003), and all found that gay men earn less. Gates and Ost (2004) used the 2000 census PUMS and provided comprehensive summary statistics of gay and lesbian population in the U.S. There are two very important case studies on spatial organization and development of gay communities in San Francisco (Castells and Murphy, 1982; Castells, 1983) and in New Orleans (Knopp, 1989; 1997), which have paved the way for gay community research.

This paper aims to test the general relationship between property values and the spatial distribution of gay people, rather than focus on a particular gay community. The primary research goal is to use both the 1990 and 2000 census 5% PUMS to test whether gay people contribute to better communities. A gay index and a set of neighborhood attributes that proxy for local social amenities are constructed. Hedonic housing models, including individual housing characteristics, gay index, and a set of other neighborhood attributes, are then estimated at the Public Use Microdata Area (PUMA) level and the city level. The results of cross-section models show that the correlation between gay index and property values are very strong and robust. The results of the panel data models at the city level further demonstrate that property values grow faster in cities with proportionately higher gay people one decade ago.

Therefore, this study provides 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 premium), but also all the time (which generate 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. This study provides empirical evidence 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 gay people on the real estate markets.[6] This study can also pave the way for further economic research on sexual orientation and gay communities, since up to now only very limited research has been done in these fields.

The rest of the paper is organized as follows: section 2 reviews the concepts of three types of social amenities and gay index, and section 3 introduces the data sets. Section 4 presents the cross-section models and the results, and section 5 the panel data models and results. Section 6 further discusses the interpretation of causality, and section 7 outlines future research agenda and concludes.

[6] Moss (1997) discussed the vital role that gays and lesbians play in renewing central-city neighborhoods.

2 Social amenities, gay Index, and neighborhood quality

Different from the five theories developed in the literature: the sorting theory, gay politics, gentrification, housing market discrimination, and intrinsic preference, this paper mainly draws the framework from the theory of urban social amenities developed by Fu (2005), and extends it to incorporate the gay index.

It has long been known that urban amenities can be capitalized into land rents or property values, such as transportation accessibility (Alonso, 1964), climate (Roback, 1982), the quality of views (Pollard, 1982), pollution (Ridker and Henning, 1967), school quality (Haurin, 1996; Brasington, 1999), crime (Roback, 1982), and racial concentration or segregation (Laurenti, 1960; Smith, 1982). Different from natural or physical 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 (Zimmer and Toma, 2000), criminal behavior (Glaeser, Sacerdote, and Scheinkman, 1996), and labor market outcomes in cities (O'Regan and Quigley, 1991). Such location-specific environment of social interactions, where urban residents interact with each other, is referred to as urban social amenities (Fu, 2005). By constructing indexes to proxy for different types of social interactions, we can directly study the impact of social amenities on property values.

Following the pioneer work by Becker (1964), Coleman (1988), and Bourdieu (1986), urban social amenities can further be classified into three categories:

human capital, social capital, and cultural capital.

2.1 Human capital

Human capital is the knowledge and skills embodied in individuals. Extensive studies (Moretti, 2004) show that social interactions between skilled workers can generate a significant amount of knowledge spillover effects. Those uncompensated knowledge spillovers are considered the reason why cities exist (Henderson, 1974; Lucas, 2001) and why cities keep growing (Black and Henderson, 1998).

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, 2004), and makes residential neighborhoods safer. The third mechanism is more interesting: well-educated neighbors, themselves, are attractive consumption amenities (Glaeser and Saiz, 2004).

Fu (2006) tested four types of human capital externalities in the workplace: the quality, specialization, diversity, and density of human capital. This paper constructs two variables, the quality and diversity of local human capital, and tests their effects on residential property values. Detailed definitions of the variables are left in section 4.

2.2 Social capital

Social capital refers to the relations between people that can be used to reach other resources or facilitate certain actions of actors (Coleman, 1988). Coleman found that social capital within family and beyond family in a neighborhood affects the creation of human capital. Putnam (2000) 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, also has important effects on labor market outcomes (Montgomery, 1991). Friendship and familial relationships are examples of strong ties in terms of the time and emotions invested in a relationship (Granovetter, 1973). People in a strong tie network are familiar with and trust each other, but they have relatively homogenous information and may impose pressures for social conformity. Therefore, strong ties are less important in spreading information or resources. In poor urban 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, 1974). In contrast, people in a weak tie network can provide new and disparate information and impose less conformity.

DiPasquale and Glaeser (1999) argued 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, including home ownership rate, the percentage of household that moved into a location certain years ago, and the percentage of households with kids under four years old.

2.3 Cultural capital

Cultural capital refers to the values, norms, customs, and cultural traditions that serve to identify and bind together a given group of people. It is expressed in people's behavior, through shared language, working attitudes, and belief systems. 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, identification with one's own group can be a powerful motivational force. This may imply that heterogeneity in terms of cultural background may decrease trustworthiness in social groups.

Different ethnic groups have different preferences over local public goods. Alesina, Baquir, and Easterly (1999) showed that after controlling for other socioeconomic and demographic factors, the share of spending on productive public goods–public education, roads, sewers and trash pickup–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 ethnic fragmentation in terms of racial diversity.

2.4 Gay index

The gay index is defined simply 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 same-sex unmarried partners, or same-sex couples we call, including both male same-sex and female same-sex unmarried partners. Lesbians have been considered to have different social and economic behavior from gay men: for example, they 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 gay index could be multi-fold. First, the gay index can proxy for the degree of openness and tolerance of local social environment, which is crucial to attract high-level human capital (Florida, 2002; 2005). Second, the inclusion of gay index to the hedonic housing model can test how sexual orientation of residents is related to neighborhood quality. Third, since we do not have data on the intrinsic characteristics and social behaviors of gay people, the gay index could also catch all the effects of unobservable characteristics of gay people, such as their aesthetic preference. Last but not least important, the gay index may be endogenous because of the sorting problem. We will discuss how to use instrumental variables and panel data to deal with this issue.

3 Data

The data sets used in this study are the 5% samples of the 1990 and 2000 census PUMS, 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 sample this paper uses selects workers of age 16-65 and their housing units in identified 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".[7] We identify same-sex unmarried partners as gay people, or same-sex couples we call, 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 has been employed in all the census-data-based gay studies. 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 form as same-sex unmarried partners were out or in 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 worth detailed explanation because geographic levels affect the estimation and interpretation of the same

[7] The questionaire 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 nonrelative.

econometric model specification. There are four different 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. A PUMA may be a portion of a central city. In the majority cases, PUMAs are nested within a metropolitan area. A metropolitan area may cross state boundary. In the 1990 PUMS, PUMAs sometimes cross state lines; but for all the PUMAs that are nested within metropolitan areas, none crosses 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 (there are a few exceptions). In the 2000 PUMS only cities meeting the minimum population threshold of 100,000 population are identified. For example, only Boston, Cambridge, Lowell, Springfield, and Worcester are identified in Massachusetts.[8]

The relationship between different geographic levels are very important to the specification and estimation of hedonic housing models, as housing markets are very localized, so are neighborhood externalities. The lower geographic level data can better control for localized neighborhood externalities and local amenities. For example, a metropolitan area fixed effect can control for the impact of regional or macroeconomic conditions on local housing markets, while a census tract fixed effect can even control for local zoning regulation, local

[8] An unpublished appendix with detailed explanation on census geographic levels is available upon request.

school quality, and other tract specific amenities. Since PUMA is the smallest geographic level in the 2000 census 5% PUMS, and PUMAs are nested within a state, state fixed effects are used when neighborhood is defined at the PUMA level.[9]

There are 126 cities and 1,582 PUMAs in the 5% sample of the 1990 census PUMS, 150 cities and 1,726 PUMAs in the 2000 census PUMS, identified in the metropolitan areas. The PUMA (city) sample used in this paper selects housing units and workers of age 16-65 in identified PUMAs (cities). Table A-1 and A-2 in the appendix list the top 10 cities and PUMAs in 2000 in terms of the same-sex couple index.

4 Cross-section models and results

This section uses the 5% sample of the 2000 census PUMS to estimate 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 but may cross MSA boundaries, to control for the differences in natural amenities and housing production efficiency at macrogeographic levels larger than PUMAs, state fixed effects are included. Another advantage of using state fixed effects is that the differences in the legal

[9] The restricted version of the census data, also called the long form, from which the PUMS are drawn, contain one-sixth households in the U.S., with detailed microgeographic information down to the census block level.

environment in terms of anti-discrimination against sexual orientation across states are also controlled for.

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

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

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 externalities at the PUMA level:

Average education: Percentage of residents with college or higher degree 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. Let S_{oj} denote the ratio of residents of occupation o at PUMA j

to the total residents at PUMA j, then

$$Occupation\ diversity = 1 - \sum_{o} S_{oj}^2. \tag{2}$$

The classification of occupations is listed in Table A-3 in the appendix.

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

$Home\ ownership\ rate$: Percentage of households who are 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.

$Ethnic\ fragmentation$: Diversity index in terms of races. It equals one minus the Herfindahl index of races. 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

$$Ethnic\ fragmentation = 1 - \sum_{r} S_{rj}^2. \tag{3}$$

The races are classified as White, Black, Hispanic, Asian and Pacific islander, and others.

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 capitalized into residential land rents. Therefore, a variable "*Average commuting time*", the average commuting time to workplace in a residential PUMA, is also included. It is measured by minutes.

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, and with language isolation; the percentage of unemployed, white, or bohemian residents.[10]

We use the Huber/White estimate of variance clustered by PUMAs to produce consistent standard errors.

The models are estimated using the 5% sample of the 2000 census PUMS. Table 2 presents the pure hedonic housing model (without neighborhood attributes) with and without the gay index.

[10] The bohemian index is constructed using the definition by Florida (2002). 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.

Table 2
Hedonic housing models with/without gay index

Variable	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t
Constant	9.6541	263.37	9.5731	248.84	9.5862	245.82	9.5338	243.32
No. of bedrooms	0.2231	66.94	0.2246	67.89	0.2235	67.04	0.2245	67.58
No. of other rooms	0.1523	77.06	0.1500	79.05	0.1515	77.26	0.1500	78.59
Building age	-0.0067	-27.18	-0.0075	-28.14	-0.0070	-29.39	-0.0075	-30.74
Dummy mobile	-0.1238	-3.60	-0.1141	-3.29	-0.1205	-3.55	-0.1141	-3.35
Dummy detach	1.1879	33.03	1.1951	33.00	1.1897	33.47	1.1945	33.47
Dummy attach	1.1201	29.37	1.1000	28.8	1.1157	29.58	1.1025	29.23
Dummy 2 apt.	1.3603	31.65	1.3381	31.23	1.3535	31.83	1.3389	31.53
Dummy 3-4 apt.	1.1329	32.15	1.3061	32.06	1.3377	32.50	1.3135	32.61
Dummy 5-9 apt.	1.1874	29.38	1.1430	29.19	1.1821	29.67	1.1528	29.88
Dummy 10-19 apt.	1.1792	28.61	1.1253	28.06	1.1736	28.88	1.1381	28.97
Dummy 20-49 apt.	1.3164	23.50	1.2327	22.68	1.3061	23.31	1.2509	23.18
Dummy >50 apt.	1.5218	21.23	1.3977	20.97	1.5094	20.92	1.4275	21.22
Dummy lot size>10	0.3534	22.61	0.3714	23.49	0.3597	23.08	0.3715	23.78
Dummy office use	1.1768	32.68	1.1827	32.64	1.1779	33.13	1.1818	33.11
SSM index			**33.5494**	5.72				
SSF index					**22.5533**	5.05		
SS index							**22.1831**	7.50
Adjusted R^2	0.4690		0.4796		0.4713		0.4783	

Note. Dependent variable is log (housing value). 48 state fixed effects are included.
Standard errors are clustered by 1,581 PUMAs. Sample size: 2,431,639.
All coefficients are significant at the 1% level.

In Table 2 all the coefficients are significant at the 1% level. The coefficients of housing characteristic variables have the expected signs. The coefficient of the SSM index, also the semi-elasticity of housing value to the SSM index, is 33.5494, meaning that a 1% increase in the male same-sex couple population at a residential PUMA is associated with on average approximately 33.55% increase in the housing values at that PUMA. The association is very strong, both statistically and economically. The same analysis applies to the SSF index and SS index.

Table 3 presents the results of the models including other social amenity variables. Since housing characteristics are not of particular interest in this study, they are omitted in Table 3.

Table 3
Hedonic housing models with social amenities and gay index at the PUMA Level

	1	2	3	4	5	6
Variable	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
Commuting time	0.0377	0.0385	0.0378	0.0393	0.0327	0.0358
	15.85	16.49	15.99	17.11	13.86	15.56
Average education	1.9079	1.9713	1.9176	1.8562	1.6769	1.5981
	31.94	34.56	32.48	29.60	27.24	20.53
Occupation diversity	1.2581	0.9969^{ns}	1.0733^{ns}	0.8719^{ns}	-1.4042	0.8577^{ns}
	1.83	1.45	1.56	1.27	-1.98	1.27
Ownership rate	-0.7988	-0.8852	-0.8125	-0.6675	-1.0578	-0.6216
	-9.70	-11.34	-10.13	-7.10	-13.44	-7.53
Five-year household	0.7131	0.7060	0.7112	0.5354	0.8644	0.6085
	5.56	5.49	5.57	3.76	7.18	4.87
English proficiency	0.8945	0.8822	0.9057	0.9905	0.8158	0.8718
	11.69	11.52	11.74	12.32	11.77	11.48
Ethnic fragmentation	-0.3493	-0.3490	-0.3594	-0.3345	-0.2533	-0.2825
	-7.52	-7.39	-7.68	-7.17	-5.64	-6.15
SSM index	**13.3627**					
	5.08					
SSF index		**4.8244**				
		1.76				
SS index			**8.0747**	**6.8070**	**8.3496**	**3.8225**
			5.14	4.33	5.03	2.33
Presence of kids				-1.1861		
				-3.63		
Unemployment					-5.7328	
					-8.69	
Bohemian index						8.5972
						5.03
Adjusted R^2	0.5706	0.5693	0.5702	0.5708	0.5740	0.5733

Note. Dependent variable is log (housing value). 48 state fixed effects are included.
Standard errors are clustered by 1,581 PUMAs. Sample size: 2,431,639.
The numbers below the coefficients are t test statistics.
Superscript ns indicates insignificance at the 10% level.

The first three columns in Table 3 use the representative social amenity variables and SSM , SSF, and SS index respectively. Column 4, 5, and 6 include additional PUMA attributes. In all the model specifications, the coefficient of SS index is positive and significant at the 1% level. The variable *Bohemian* in column 6 drives the coefficient of SS index much lower; this is because *Bohemian* and SS index are moderately correlated (the correlation coefficient is 0.46).

The interpretation of human capital, social capital, and cultural capital variables can follow that in Fu (2005). The only difference is the sign of the coefficients of home ownership rate and the five-year households. Fu (2005) used the long form 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 home owners have 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 level of geography. Fu (2005) constructed the variables at the census tract and block levels. Here, unfortunately, the public data enable us to construct variables only at the PUMA level.

A few PUMAs have disproportionately high concentration of same-sex couples. We experiment to drop the top three, top ten PUMAs in terms of the SS index, and re-estimate the models in Table 2 and Table 3, respectively. The results are very similar and in most of the cases, even better: the coefficients of SSM index and SS index become 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 5% sample of the 2000 census 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. Table 4 presents the results of the city-level models.

Table 4
Hedonic housing models with social amenities and gay index at the city level

	1	2	3	4	5	6
Variable	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
Commuting time	-0.0054^{ns}	-0.0010^{ns}	-0.0042^{ns}	-0.0099^{ns}	-0.0081^{ns}	-0.0073^{ns}
	-0.57	-0.10	-0.495	-1.08	-0.97	-0.81
Average education	1.5754	1.5953	1.6162	1.8654	1.4082	1.3979
	7.35	7.00	7.71	9.23	8.03	5.35
Occupation diversity	10.7337	12.5686	11.0409	11.7614	7.2436	9.0113
	4.35	4.30	4.10	4.77	2.98	4.13
Ownership rate	-0.6995	-0.8911	0.7164	-0.9433	-0.9615	-0.4169^{ns}
	-2.84	-3.64	-2.98	-4.36	-4.58	-1.49
Five-year household	2.3305	2.3883	2.5421	2.8365	2.6324	2.4114
	4.77	4.07	5.08	5.61	5.46	4.61
English proficiency	1.5040	1.5184	1.6049	1.4831	1.1203	1.5244
	4.42	4.41	4.69	4.75	3.98	5.19
Ethnic fragmentation	-0.6063	-0.4748	-0.5312	-0.6041	-0.1351^{ns}	-0.4720
	-2.16	-1.78	-2.01	-2.49	-0.55	-2.02
SSM index	**32.3147**					
	4.18					
SSF index		**16.2602**				
		1.63				
SS index			**19.0651**	**22.3089**	**22.7023**	**18.4484**
			3.49	4.75	4.82	2.82
Presence of kids				2.5779		
				2.77		
Unemployment					-6.9150	
					-3.18	
Bohemian index						5.7205
						1.92
Adjusted R^2	0.5190	0.5183	0.5188	0.5192	0.5196	0.5192
SS index-IV^*			29.0420^{ns}	20.5790^{ns}	52.0294	11.2566^{ns}
			1.26	1.10	1.83	0.50

Note. The dependent variable is log (housing value). Sample size: 372,949.
104 metropolitan area fixed effects are included. Standard errors are clustered by150 cities.
The numbers below the coefficients are t test statistics.
$^*SS$ index-IV is the instrumental variables estimator.
Superscript ns indicates insignificance at the 10% level.

Table 4 shows that the coefficient of average commuting time, though not significant, has negative sign, as the theoretical spatial models predict. The coefficients of the SS index in different models are larger than those at the PUMA level: between 18 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 a city is more homogenous than a PUMA in terms of economic and social environment, and the city sample includes only cities with at least 100,000 population, which makes the measurement error problem less serious.

Again, we experiment to drop the top five, top ten cities in terms of 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 index and SS index become larger. This shows that the general results are not driven by a few cities with high concentration of same-sex couples.

Since the SS index may be endogenous, we also try to use instrumental variables estimation. One instrument variable for SS index we find is a dummy variable, *Law*, indicating whether a city has past the law to prevent discrimination in public employment based on sexual orientation by year 2000. In the city sample there are 52 cities that past at least the law prohibiting discrimination in public employment on the basis of sexual orientation.[11] The row with variable "SS index-IV" in Table 4 presents the instrumental variables estimators. The coefficients are all positive, but not significant, except column 5 is significant at

[11] The data is from a table "States, Cities and Counties with Civil Rights Ordinances, Policies or Proclamations Prohibiting Discriminaion on the Basis of Sexual Orientation", available at www.thetaskforce.org.

the 10% level. The Hausman test statistics indicates that there is no systematic difference between the coefficients of OLS and IV estimators in column 3 specification, which means the OLS estimator is consistent and efficient.[12] Therefore, we conclude that column 3 is our preferred model specification.

We also estimate the PUMA and city level models using the 5% sample of the 1990 census PUMS. The pattern of the results is pretty similar (the results are not reported here).[13]

5 Sorting or causality? A panel data model test

The cross-section models show that the correlation between gay index and housing values are strong and robust. However, possible unobserved and omitted residents' characteristics may generate endogeneity problem. We can not 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 preference and personal characteristics.

The case studies on residential communities gentrified by gay people suggest that it is gay people that have improved their neighborhoods, and have made their communities better, not the case that gay people choose to move into high-quality communities. However, special cases probably can not be generalized to justify the general relationship between property values and the spatial

[12] Other models fitted on the data fail to meet the asymptotic assumptions of the Hausman test.

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

distribution of gay people.

To disentangle the causality problem empirically, we use the 5% samples of both the 1990 and 2000 census PUMS and construct a panel data set. Although PUMA information is available in both the 1990 and 2000 PUMS, the PUMA boundaries are not comparable. Fortunately, most of the identified cities remain the same boundaries, which enable us to estimate a panel data model with city fixed effects.

Three type 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.[14] 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 effects; 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 with the representative city attributes.

[14] Median housing value is not a perfect measure of average housing price level in a city. The distribution of housing values in a city may change over time. We also tried the mean housing value in a city, and the results are similar. We have not found city level housing price index.

Table 5
Panel data fixed effects model

Variable	Coeff.	z	Coeff.	z	Coeff.	z
$t20$ (Time dummy)	-0.0082^{ns}	-0.34	0.0874	2.70	-0.0009^{ns}	-0.03
Average education	2.4657	12.63	2.8277	14.26	2.6378	12.99
Occupation diversity	5.6543	3.86	6.1200	3.62	5.6110	3.61
Home ownership rate	0.7056	4.80	0.5022	3.53	0.6619	4.48
Five-year households	-0.4528	-2.42	-0.3974	-2.12	-0.4273	-2.24
English proficiency	3.7412	38.19	3.7791	38.70	3.7789	37.32
Ethnic fragmentation	0.2838	2.46	0.3436	3.21	0.3196	2.77
SSM index	**17.1981**	3.71				
SSF index			**-22.1193**	-2.43		
SS index					**5.5924**^{ns}	1.40
Wald χ^2	3967		2107		2558	

Note. Dependent variable: log(median housing value in a city).
Generalized least squares estimator, corrected for heteroskedasticity.
Fixed effects: 92 cities. Number of observations: 184.
Superscript ns indicates insignificance at the 10% level.

The models in Table 5 are estimated first using linear least squares method. The modified Wald statistics show that there exists groupwise heteroskedasticity. We then use the generalized least squares estimate to correct for within panel heteroskedasticity. The coefficients of the time dummy are not significant in two of the three models, which suggests that using nominal housing values is not too problematic. 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 only two time periods data are available, city fixed effects have to be dropped. The results are presented in Table 6.

Table 6
A Granger causality test

Variable	Coeff.	t	Coeff.	t	Coeff.	t
Log median housing value	0.6288	8.83	0.6458	8.66	0.6413	8.94
Average education	1.0307	2.91	1.1945	3.26	0.9984	2.74
Occupation diversity	-3.8394^{ns}	-1.45	-4.1715^{ns}	-1.53	-3.9907^{ns}	-1.50
Home ownership rate	0.2482^{ns}	1.25	0.2148^{ns}	1.05	0.2572^{ns}	1.29
Five-year households	0.2777^{ns}	1.03	0.3492^{ns}	1.27	0.3013^{ns}	1.12
English proficiency	0.5078^{ns}	1.25	0.4894^{ns}	1.16	0.4710	1.15
Ethnic fragmentation	-0.2770	-2.56	-0.2252^{ns}	-1.34	-0.2757	-1.66
SSM index	**21.7868**	2.46				
SSF index			**22.1912**^{ns}	1.26		
SS index					**15.3518**	2.32
Adjusted R^2	0.8699		0.8631		0.8689	

Note. Dependent variable: log median housing value in a city in 2000.
Number of observations: 92. ns indicates insignificance at the 10% level.
All independent variables are constructed using the 1990 data.

The results in Table 6 are encouraging: not only past housing values, human capital stock, and ethnic fragmentation, but also the proportion of same-sex couples in a city are good predictors of future median housing values. Since future events can not be used to predict the past events, the results in Table 6 suggest that it is indeed that the same-sex couples "*cause*" property values to increase: for example, after controlling for the housing value and other city attributes in 1990, a 1% increase in the same-sex couple population in a city in 1990 can generate a 15.35% increase in median housing value in 2000.

The third model is adapted from the growth convergence model (Barro and Sala-i-Martin, 1992). The growth rate of median housing value is regressed on the lagged city attributes. The model is specified as

$$GP_{2000,j} = \alpha + \gamma' X_{1990,j} + \epsilon_{2000,j}, \tag{6}$$

where $GP_{2000,j}$ is the nominal growth rate of median reported housing value in city j over 1990-2000; α is a constant; $X_{1990,j}$ is the attributes vector of social amenities in city j in 1990. Table 7 presents the growth model results.

Table 7

Median housing value growth model

Variable	Coeff.	t	Coeff.	t	Coeff.	t
Average education	-0.0968	-0.30^{ns}	0.01566	0.05^{ns}	-0.1496	-0.47^{ns}
Occupation diversity	-6.6523	-2.24	-6.8534	-2.29	-6.6846	-2.26
Home ownership rate	0.0217	0.10^{ns}	0.0415	0.18^{ns}	0.0549	0.25^{ns}
Five-year households	0.4593	1.50^{ns}	0.5216	1.70	0.4691	1.54
English proficiency	-1.2953	-5.32	-1.2206	-5.01	-1.2754	-5.29
Race diversity	-0.3617	-1.92	-0.3278	-1.76	-0.3711	-1.99
SSM index	**22.2741**	2.19				
SSF index			**38.9327**	2.01		
SS index					**18.2057**	2.43
Adjusted R^2	0.4148		0.4098		0.4220	

Note. Dependent variable: nominal growth rate of median housing value during 1990-2000. Independent variables are based on 1990 data.

ns indicates insignificance at the 10% level. Sample size: 92.

Table 7 shows that the nominal housing value grow faster in a city with higher proportion of same-sex couples a decade ago. All the coefficients of the three gay indexes are significant at the 5% level, even though the coefficients of human capital stock and other social amenities variables are not significant. The results are consistent with the life-cycle theory of gentrification.

6 A passion to preserve?

The last section provides evidence that gay people *make*, rather than choose to live in, better neighborhoods. What could motivate gay people make better neighborhoods? Obviously, the data themselves can not reveal such information, but sociology literature can shed light on this question.

Castells (1983) provided an interesting description of how gay people gentrified Castro neighborhood. Being discriminated against in the housing market, gay realtors and interior decorators discovered a hard way to survive 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 profits. Fellows (2004) argued that gay men are very sensitive to beauty, and have long been impassioned pioneers as 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 model specification and conclusion.

However, whether gay people have stronger innate aesthetic or artistic tastes than heterosexual people is not a consensus. Crimmins (2004) described, in a non-academic way, how gay people's aesthetic preference in many fields, from fashion to housing, has shaped the mainstream American pop culture. Lewis and Seaman (2004) 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 art.

Table 8 uses the PUMA sample of the 2000 census 5% PUMS and presents a set of summary statistics for the same-sex couple sample and the heterosexual sample. The results shows that gay people indeed more likely to work as artists, to take bohemian occupations, and to receive more school education. How to explain gay people's stronger affinity for arts is worth further investigation.

Table 8
Summary statistics: Mean or frequency

Variable	Same-sex couple sample	Heterosexual people sample
Total personal income	$38,069	$30,352
Wage and salary income	$32,193	$25,482
White	78.38%	74.51%
Age	39.7	38.5
Home owner	63.77%	65.40%
College degree	28.54%	22.34%
Graduate degree	14.33%	8.46%
Employed	77.89%	69.24%
Lived more than 5 years	79.95%	69.86%
Presence of children	18.63%	42.29%
Management occupation	15.32%	11.05%
Service occupation	12.85%	13.94%
Office occupation	12.82%	13.92%
Sales occupation	10.12%	10.24%
Education occupation	5.21%	4.6%
Health occupation	4.99%	3.74%
Artistic occupation	3.46%	1.77%
Bohemian	2.5%	1.13%

Note. Number of observations in the same-sex couple sample: 44,758.
Number of observations in the heterosexual people sample: 6,432,267.

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 the correlation between the spatial concentration of same-sex couples and housing values is strong and robust, but also are housing values 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 nice neighborhoods everywhere and all the time.

This study has two major drawbacks. First, even though the 1990 and 2000 census PUMS are the most comprehensive and systematic data publicly available for economic study of sexual orientation, the data quality is still not absolutely guaranteed. There exist undercount, measurement error, and report error (Black et al., 2000; Badgett and Rogers, 2003). Badgett and Rogers (2003) 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, 2003). For a location with an increase in 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 1990 census. We are not clear how serious this issue is. Black et al. (2000) discussed the measurement and record error of the 1990 census PUMS. The measurement error, however, is not a serious problem to our conclusion, as measurement error causes the coefficients of independent variables to be underestimated. In brief, a better quality assessment of the census PUMS requires the access to the confidential short form (100% sample) and long form (one in six sample) census data.

Second, the neighborhood in this study is defined at the macrogeographic levels: the PUMA level and the city level. Since social interaction and neighborhood externalities are very localized, 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. The restricted version contains detailed location information down to the census block level, which is the best to control for location specific effects.

The coefficient of SSF index in all the models is different from the SSM index, sometimes even reverse the sign (in Table 5). This is probably due to the gender difference related to urban space and politics. Castells (1983) argued that men seek to dominate space while women attach more importance to networks and relationships. Adler and Brenner (1992) found that lesbians tend not to have access to capital, more likely to be primary caretakers of children, etc. Tables A-1 and A-2 show that male and female same-sex couples tend to be clustered in different locations. Why the concentration of gay men and lesbians have different effects on urban space deserves further investigation.

This paper did not construct any theoretical models. A more rigorous analysis would be to model the process of gentrification based on the intrinsic preference of gentrifiers. A possible way might be to introduce the heterogenous preferences to 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 Bo Zhao, Jed Kolko, Aimee Wagenen, Judith Pereira, Man Jia, and seminar participants at the Southwestern University of Finance and Economics (China), and the China Center for Economic Studies at Peking University very much for the very helpful comments and suggestions. The SAS and STATA programs developed by the author are available upon request.

Appendix

Table A-1

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

Note. The unit of the SS, SSM, SSF index is %.

Table A-2

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

Note. The unit of the SS, SSM, SSF index is %.

Table A-3
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

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

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

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

Alonso, W., 1964, Location and land use, Harvard University Press, Cambridge.

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

Barro, R. and X. Sala-i-Martin, 1992, Convergence, Journal of Political Economy 100, 223-251.

Becker, G., 1964, Human capital, The University of Chicago Press, Chicago.

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

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

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

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

Black, D. and V. Henderson, 1999, A theory of urban growth, Journal of Political Economy 107, 252-284.

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

Bourdieu, P., 1986, Forms of capital, in: J. Richardson, ed., Handbook of theory and research for the sociology of education, Greenwood, New York.

Brasington, D., 1999, Which measures of school quality does the housing market value? Journal of Real Estate Research 18, 395-413.

Brueckner, J. and S. Rosenthal, 2005, Gentrification and neighborhood housing cycles: Will American's future downtown be rich? Mimeo, University of California, Irvine.

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

Castells, M. and K.A. Murphy, 1982, 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, 237-259.

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

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

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

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

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

Florida, R., 2002, Bohemian and economic geography, Journal of Economic Geography 2, 55-71.

Florida, R., 2005, Cities and the creative class, Routledge, New York and London.

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

Fu, S., 2006, Smart cafe cities: Testing human capital externalities in the Boston metropolitan area, Journal of Urban Economics, in press.

Gates, G. and J. Ost, 2004, The gay&lesbian atlas, The Urban Institute Press.

Glaeser, E., B. Sacerdote and J. Scheinkman, 1996, Crime and social interactions, Quarterly Journal of Economics 111, 507-548.

Glaeser, E., J. Kolko and A. Saiz, 2001, Consumer city, Journal of Economic Geography 1, 27-50.

Glaeser, E. and A. Saiz, 2004, The rise of the skilled city, Brookings-Wharton Papers on Urban Affairs, 47-94.

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

Granovetter, M., 1973, The strength of weak ties, American Journal of Sociology 78, 1360-1380.

Haurin, D., 1996, School quality and real house prices: Inter- and intrametropolitan effects, Journal of Housing Economics 5, 351-368.

Henderson, V., 1974, The sizes and types of cities, American Economic Review 64, 640-656.

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

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

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

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

Knopp, L., 1997, 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, 45-63.

Kolko, J., 2006, The economics of gentrification, Mimeo, Harvard University.

Laurenti, L., 1960, Property values and race, University of California Press, Berkeley.

Levine, M., 1979, Gay ghetto, Journal of Homosexuality 4, 363-377.

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

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

Lucas, R.E., Jr., 2001, Externalities and cities, Review of Economic Dynamics 4, 245-274.

Montgomery, J., 1991, Social networks and labor-market outcomes: Toward an economic analysis, American Economic Review 81, 1408-1418.

Moretti, E., 2004, Human capital externalities in cities, in: V. Henderson and J.F. Thisse, eds., Handbook of regional and urban economics, Vol. 4, Elsevier, North-Holland, Amsterdam, 2243-2291.

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

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

O'Regan, K. and J. Quigley, 1991, Labor market access and labor Market outcomes for urban youth, Regional Science and Urban Economics 21, 277-293.

Palen, J. and B. London, eds., 1984, Gentrification, displacement and neighborhood revitalization, State University of New York Press.

Pollard, R.,1982, View amenities, building heights, and housing supply, in: D. Diamond and G. Tolley, eds., The Economics of Urban Amenities, Academic Press, 105-124.

Putnam, R., 2000, Bowling alone: The collapse and revival of American community, Simon & Schuster.

Ridker, R. and J. Henning, 1967, The determinants of residential property values with special reference to air pollution, Review of Economics and Statistics 49, 246-257.

Roback, J., 1982, Wages, rents, and the quality of life, Journal of Political Economy 90, 1257-1278.

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

Smith, B., 1982, Racial composition as a neighborhood amenity, in: D. Diamond and G. Tolley, eds., The economics of urban amenities, Academic Press, 165-194.

Smith, N. and P. Williams, eds., 1986, Gentrification of the city, Allen&Unwin, Boston.

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

Tuner, M. and M. Mikelsons, 1992, Patterns of racial steering in four metropolitan areas, Journal of Housing Economics 2, 199-234.

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

Zimmer, R. and E. Toma, 2000, Peer effects in private and public schools across countries, Journal of Policy Analysis and Management 19, 75-92.

Appendix A Geographical levels in the decennial census data (Not for publication)

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 metropolitan area has remained essentially the same over time, there is slight change. For example, the boundaries of each metropolitan area may have been adjusted; new metropolitan areas regularly emerged; the definition slightly varies.

Metropolitan areas have been referred to by several different names. In 1950, the term was Standard Metropolitan Area (SMA). In 1970 and 1980, it is called Standard Metropolitan Statistic Area (SMSA). In 1990 and 2000, the terms 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 cross state boundary. 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, 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, 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 majorities cases, PUMAs are nested within a metropolitan area. If a PUMA cross 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 state boundary. The PUMAs in 2000 census do not cross state lines. There are 1,726 and 1,582 PUMAs identified in the metropolitan area in the 5% sample of 1990 and 2000 census PUMS, respectively. For the 1990 and 2000 5% census PUMS, the PUMA is the lowest level of geography.[15]

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 2000 census PUMS, only cities meeting the minimum population threshold of 100,000 are

[15] 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 2000 census PUMS, Super-PUMA is the smallest geographical division available. Super-PUMAs do not cross state lines.

identified. There are 126 and 150 cities in the 5% sample of 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, contain one of the six 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 4000 persons. In the 2000 census tract population criteria is 1500 to 8000 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 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 contain 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.)