Estimating Ethnic Preferences Using Ethnic Housing Quotas in Singapore

Maisy Wong*†

November 30, 2012

Abstract

This paper estimates people's taste for living with own-ethnic-group neighbors using variation from a natural experiment in Singapore: ethnic housing quotas. I develop a location choice model that informs the use of policy variation from the quotas to address endogeneity issues well-known in the social interactions literature. I assembled a dataset on neighborhood level ethnic proportions by matching more than 500,000 names in the phonebook to ethnicities. I find that all groups want to live with some own-ethnic-group neighbors but they also exhibit inverted U-shaped preferences so that once a neighborhood has enough own-ethnic-neighbors, they would rather add a new neighbor from other groups. Welfare simulations show that about 30% of the neighborhoods are within one standard deviation of the first best allocation of ethnic groups.

---

*I wish to thank my advisers Esther Duflo, Amy Finkelstein, Panle Jia and Bill Wheaton for guidance and support. I also benefited from comments by two anonymous referees and the editor. I am indebted to Peter Diamond, Glenn Ellison, Fernando Ferreira, Alex Gelber, Michael Greenstone, Joe Gyourko, Whitney Newey, Stephen Ryan and Stephen L. Ross and Todd Sinai for their helpful comments. I have benefited from conversations with Patrick Bayer, Chua Beng Huat, Jessica Cohen, Greg Fischer, Raymond Guiteras, Kam Lee Ching, Robert McMillan, Trang Nguyen and Wong Liang Kit. Lee Hye Jin provided excellent research assistance. I would like to thank the Singapore Housing Development Board for permission to use their data. This study would not have been possible without support from the MIT Shultz Fund and the Wharton Zell/Lurie Real Estate Center. All remaining errors are my own.

†Maisy Wong. Wharton Real Estate Department, University of Pennsylvania, 3620 Locust Walk, 1464 SHDH, Philadelphia PA 19104-6302 (E-mail: maisy@wharton.upenn.edu).

1 Introduction

There are many policies around the world designed to encourage ethnic desegregation in housing markets. In Chicago, the Gautreaux program (the predecessor of the Moving Towards Opportunity program) offered rent subsidies to African American residents of public housing who wanted to move to desegregated areas. Germany, the United Kingdom and Netherlands, impose strict restrictions on where refugee immigrants can settle. Many countries also have "integration maintenance programs" or "neighborhood stabilization programs" to encourage desegregation.[1] These policies are often controversial as they are alleged to favor some ethnic groups at the expense of others. Regardless of the motivation behind these policies, knowing the welfare effects is important because these desegregation policies affect the location choices of many individuals.

This paper estimates the taste for own-ethnic-group neighbors (ethnic preferences) using variation from one such ethnic desegregation policy in Singapore: the ethnic housing quotas. I develop a residential location choice model to study how heterogeneous households sort into neighborhoods as the ethnic proportions in the neighborhood change.[2] My model provides theoretical underpinnings that inform the use of policy variation from the ethnic quotas to address endogeneity issues well-known in the literature on social interactions and residential segregation.[3] I am able to model social interactions at such a local level because I assembled a dataset of ethnic proportions by hand-matching more than 500,000 names to ethnicities using the Singapore residential phonebook. I combined this phonebook data with data I collected on housing transaction prices, neighborhood choices of movers from different ethnic groups (calculated by matching names between two sequential phonebooks) and attributes of neighborhoods, such as school quality and the age of buildings.

The ethnic housing quotas in Singapore is a fascinating natural experiment.[4] It was

[1]See Polikoff (1986) for a review of these policies.

[2]There is a large literature on sorting in housing markets (eg. Tiebout (1956), Benabou (1993) and Epple and Sieg (1997)). My location choice model uses a discrete choice framework that builds on McFadden (1973, 1978), Berry (1994) and Berry et al. (1995). For examples of discrete choice models in the urban economics literature, see Quigley (1985) and Nechyba and Strauss (1998).

[3]See Manski (1993), Brock and Durlauf (2001, 2002), Bayer and Timmins (2005, 2007) for papers on social interactions. The seminal paper on residential segregation is Schelling (1971). Most studies that investigate the causes of residential segregation (Gabriel and Rosenthal (1989); Hårsman and Quigley (1995); Cutler et al. (1999); Bajari and Kahn (2005); Card et al. (2008)) do so at the MSA or PUMA level. See Bayer et al. (2007) for an exception (they use restricted data at the Census block group level). I estimate location choices at the neighborhood level. The average neighborhood in Singapore has 4000 households and an average land area of 1.5 square miles.

[4]There is a growing literature on ethnic desegregation policies (eg. Rosenbaum (1992, 1995), Edin et al. (2003) and Damm (009b)). Moffitt (2001) investigates the use of desegregation policies as natural experiments to study social interactions. These papers do not include welfare analysis of these policies

implemented in public housing estates in 1989 to encourage residential desegregation amongst the three major ethnic groups in Singapore – Chinese (77%), Malays (14%) and Indians (8%) (Singapore Department of Statistics, 2000). The quotas are upper limits on the proportions of Chinese, Malays and Indians at a location. Locations with ethnic proportions that are at or above the quota limits are subjected to restrictions designed to prevent these locations from becoming more segregated. For example, non-Chinese sellers living in Chinese-constrained locations are not allowed to sell to Chinese buyers because this transaction increases the Chinese proportion and makes the location more segregated. Without the quota policy, profit-maximizing sellers would sell to the highest bidder and equilibrium prices for the same location would not differ by the ethnic group of the buyer because any differences would be arbitraged away by sellers.[5] The ethnic-based restrictions of the quota limits arbitrage opportunities. This price discrimination-like mechanism generates equilibrium price dispersion across buyers from different ethnic groups, making it possible, for example, to observe Chinese and non-Chinese buyers paying different prices for the same location, in equilibrium.

I begin my empirical analysis by documenting price dispersion across ethnic groups using a descriptive analysis of price effects, in the spirit of the "regression kink design" (Card et al., 2009).[6] The strategy is to identify kinks in the outcome variable that coincide with kinks in the policy rule. Using transactions data close to the quota limits and controlling for polynomials of ethnic proportions, I estimate statistically significant discontinuities in prices that suggest heterogeneity in ethnic preferences. If Chinese and non-Chinese buyers had the same willingness-to-pay for Chinese neighborhoods, the quotas would have no price effects because there would have been no price differences across ethnic groups to be arbitraged away in the first place. Moreover, when I control for the ethnic weights of the buyers, I find evidence of price dispersion across ethnic groups that is consistent with the price discrimination mechanism of the quotas.

I take these findings of price dispersion from the reduced form analysis to develop and estimate a location choice model. In an improvement over existing research, my

because they do not estimate preferences.

[5]This paper focuses on estimation of the demand side. I do not model supply decisions explicitly. All sellers are assumed to be profit-maximizing so that the ethnicity of the seller does not matter. An important caveat is discrimination in the housing market (eg. Chinese sellers charging Chinese and non-Chinese buyers different prices), something I return to later in the paper.

[6]While the estimation equations are very similar to regression discontinuity (Angrist and Lavy, 1999; Hahn et al., 2001; Lee, 2008), it does not fit within a standard regression discontinuity design (RDD) framework (Imbens and Lemieux, 2008) because the regressor of interest (ethnic proportions) is endogenous. To implement RDD, I would need pre-policy data on ethnic proportions and prices. Therefore, the identification strategy in the descriptive analysis is more similar to Card et al. (2009)'s study on the impact of previous earnings on unemployment insurance benefits.

model allows the taste for unobserved neighborhood amenities to vary across ethnic groups. This is an important improvement in urban models with social interactions because the observed ethnic proportions in a neighborhood (the key explanatory variable) are correlated with ethnic-specific taste for unobserved amenities. Estimating utility over locations without allowing the taste for unobserved neighborhood amenities to vary across ethnic groups would bias the estimates of ethnic preferences (the coefficient on ethnic proportions) upwards.

However, a location choice model with ethnic-specific taste for unobserved amenities is under-identified in most empirical settings. The standard empirical strategy is to correlate neighborhood choices with ethnic proportions. Without the quotas, we would be under-identified because the neighborhood choices of Chinese movers, for example, would be positively correlated with both the taste for observed Chinese neighbors and the taste for unobserved Chinese amenities.

The ethnic quotas generate another source of neighborhood-by-ethnic group variation: prices. In Chinese-constrained neighborhoods, the price paid by Chinese and non-Chinese buyers can be different for the same neighborhood due to the price discrimination effects of the quota. This price dispersion helps to identify the Chinese taste for amenities separately from the non-Chinese taste for amenities: Conditional on the proportion of Chinese living in a neighborhood (stock), Chinese buyers (flow) are willing to pay more than non-Chinese buyers if they have a stronger taste for amenities in that neighborhood than non-Chinese. Therefore, the model with both ethnic-specific taste for neighborhood proportions and ethnic-specific taste for neighborhood amenities is no longer under-identified using neighborhood-by-ethnic group variation in both location choices and prices.

I operationalize the identification approach above by using the quota policy to generate new instruments (and new moments) for ethnic-specific prices, that are conditionally mean independent from amenities. I first follow the literature and use instruments (including historical ethnic settlements in Singapore and attributes from nearby but not adjacent neighborhoods) to isolate plausibly exogenous variation in ethnic proportions, and hence, variation in the probability that the quotas bind. That is, my instruments are quota probabilities (probability that the quota is constrained) that are constructed based on whether the estimated ethnic proportions are above the quota limits.[7]

The step function of the policy rule is important.[8] The identification assump-

[7]Actual quota dummies are not instruments because they are likely to bind when ethnic proportions or the level of ethnic-specific amenities are high.

[8]The rule is a step function because the block or neighborhood is constrained when the ethnic proportions are above the quota limit and unconstrained otherwise.

tion is that the estimated quota probabilities are correlated with ethnic-specific prices through price discrimination, but conditionally mean independent of the taste for unobserved amenities. This assumption fails if other neighborhood attributes also affect prices discontinuously at the policy limits. The effect of the quota probabilities is non-parametrically identified using the step function of the quota policy. This represents a novel use of non-parametric instrumental variables within a structural location choice model.[9] Using the method of simulated moments, I estimate three location choice models simultaneously, one for each ethnic group.

My estimates show that all groups have strong preferences for living with members of their own ethnic group but the shapes of the preferences are very different across the three ethnic groups. All groups have ethnic preferences that are inverted U-shaped but with different turning points. This means that once a neighborhood has enough members of their own ethnic group, households want new neighbors from other ethnic groups. Previous research in the United States has documented evidence of tastes for diversity using data on racial attitudes from the General Social Survey (Aldrich et al., 2005) but most estimation of ethnic preferences have not investigated such non-monotonicity because they have focused on linear models.

These estimates that distinguish the taste for own-ethnic-neighbors from the taste for ethnic-based amenities are important for policy. They are inputs for cost benefit analyses that can inform trade-offs between the use of ethnic-based amenities versus the use of location choice restrictions to encourage residential desegregation. The inverted U-shape of Chinese preferences suggest that the average Chinese prefers neighbors from other ethnic groups on the margin, and yet, 9% of the neighborhoods are Chinese-constrained in my data in an average month. This suggests that these neighborhoods could be constrained because of strong Chinese taste for amenities in these neighborhoods.

I use these estimates of ethnic preferences to perform first best simulations. To my knowledge, this is the first welfare analysis of desegregation policies. Due to externalities (a mover affects the utility of his current and future neighbors by changing the ethnic composition of the neighborhood), the decentralized equilibrium may not achieve the first best spatial allocation of ethnic groups. In cases where both mixed and segregated equilibria exist, policies such as the ethnic quotas could be used as a coordination mechanism to achieve the mixed equilibrium (Schelling, 1971). Under the quota policy, I find that 32%, 35% and 31% of neighborhoods are within 1 standard deviation of the first

[9] In a related location choice model, Bayer et al. (2007) embed a boundary discontinuity design in a structural model. Instead of using instrumental variables, the idea is that differential sorting across school district boundaries generates arguably exogenous variation in ethnic proportions, since the main source of sorting–school quality–is observed and can be controlled for.

best allocation of Chinese, Malays and Indians respectively. The standard deviations are 7% for Chinese and Malays and 3% for Indians.

In the next section, I discuss the background of ethnic quotas in Singapore. Then, I describe the data (Section 3) and provide a descriptive analysis of quota effects on prices (Section 4). I then build a model of individual utility over residential locations that incorporates price discrimination (Section 5), discuss estimation of the model (Section 6) and present the results (Section 7). Finally, I conclude in Section 8.

2 Background

Singapore is a multi-ethnic country with a population of 4.5 million (Singapore Department of Statistics, 2006). The three major ethnic groups are the Chinese (77%), the Malays (14%) and the Indians (8%).[10] The Chinese have the highest median monthly income (S$2335), followed by the Indians (S$2167) and the Malays (S$1790) (Singapore Census, 2000).

Public housing is the most popular choice of housing in Singapore. The units are built and managed by the Housing Development Board (HDB). Its missions include providing affordable housing of quality and value, creating vibrant and sustainable towns and promoting the building of active and cohesive communities. The HDB was created in 1960 to solve Singapore's housing crisis. At that time, many Singaporeans were living in slums and crowded settlements. HDB promptly began building thousands of HDB units and rapidly increased the share of Singaporeans living in public housing from 9% in 1960 to 82% today.[11]

Public housing in Singapore is based on the concept of new towns: self-contained, large scale satellite housing developments that usually includes public housing units, a town center and a range of amenities. HDB dwellings are relatively uniform. There are eight types of HDB units: type 1 are one-room units (studios), types 2 to 5 are two- to five-room units. The remaining three types (Executive, HUDC and Multi-Generation) are essentially larger and more expensive units with different layouts.

There are three ways Singapore residents can live in an HDB unit. They may apply through the primary allocation system for new HDB units,[12] they may purchase existing

[10] All three ethnic groups are citizens or permanent residents, none are immigrants.

[11] For more details, see http://www.hdb.gov.sg/fi10/fi10320p.nsf/w/AboutUsHDBHistory?OpenDocument.

[12] All eligible Singapore citizens can apply to buy new HDB units in the primary market. To be eligible, the applicant must be married, aged 21 and above and have gross income below a ceiling determined by HDB for that year. The primary market comprises mostly new HDB units that are built-to-order in new HDB estates. Most applicants are first time buyers because applicants must not have interest in any other property within 30 months of the application. Applicants can submit a request for one of

HDB units in the resale market or they may rent. The rental market is negligible (98% percent of the HDB units are owner-occupied) because rentals are regulated to ensure that public housing is used for primary residences only (Housing Development Board, 2006).[13] This paper focuses on the resale market only and only includes HDB neighborhoods in the resale market. Relative to the primary market which is heavily regulated, the resale market functions as an open market.

To understand the ethnic quotas, it is important to understand the geography of housing markets in Singapore. The smallest spatial unit is an HDB *unit*. An HDB *block* is a multi-storeyed apartment block with an average of 74 households. HDB *neighborhoods* are clusters of HDB blocks. The average neighborhood in Singapore has 4000 households, 45 HDB blocks and an average land area of 1.5 square miles. Due to the high population density in Singapore, a neighborhood is comparable to a US Census block group by land area but it is comparable to a US Census tract by population size. HDB *towns* are clusters of HDB neighborhoods.

Figure 1 shows a map of an HDB community with HDB blocks and HDB neighborhoods. HDB *blocks* and *neighborhoods* are terms used by HDB to describe clusters of public housing units. Throughout the paper, blocks and neighborhoods refer to HDB blocks and HDB neighborhoods. Each number in the map corresponds to an HDB block. Notice that the block numbers range from 100 to 600. There are four HDB neighborhoods in the map. All blocks that range from 100 to 199 belong to neighborhood 1 and all blocks from 200 to 299 are in neighborhood 2 and neighborhoods 4 and 5 are defined similarly. HDB neighborhoods are clusters of HDB blocks that are spatially contained, and separated from other HDB neighborhoods and other private housing by main roads. All HDB blocks and neighborhoods include public housing units only. There are no private housing units in this map.

The government of Singapore introduced the Ethnic Integration Policy to address the 'problem' of the increase in the 'concentrations of racial groups' in HDB estates (Parliamentary Debates, 1989).[14] The policy was debated in parliament on February 16, 1989 and was implemented two weeks later on March 1, 1989. It is a set of quota limits at the HDB block and HDB neighborhood level. Table 1 lists the quota limits,

the eight HDB types of units and also a preference for one of the three HDB zones (North, North East and West). A computer ballot will determine the applicants' queue position to book a unit. Lottery winners are given three months to select a new unit. They will typically wait two to three years before the unit is completed in the new HDB estate. After five years, the owners are free to sell their units in the resale market.

[13] In my sample period, owners of public housing are only allowed to rent if they can prove that they need to be out of the country for an extended period.

[14] Racial harmony is important in Singapore because of violent racial riots in the 1960s.

in comparison to the 2000 national ethnic proportions. These limits are the same for all locations in Singapore. They were calculated in 1989 according to the population's ethnic mix and the projected rate of household formation. Neighborhood quotas are 2% to 8% above the national ethnic proportions in 2000. Block quotas are 3% above the neighborhood quotas, allowing more flexibility at the block level because blocks can be more segregated than neighborhoods. In practice, the HDB did not want to evict owners in existing units that were in violation of the quotas. To this day, there exist blocks and neighborhoods whose ethnic proportions exceed the quota limits.

Price discrimination mechanism of quotas

The quotas are upper limits on ethnic proportions to prevent HDB communities that are already segregated from becoming more segregated. Any transaction that makes a HDB block or neighborhood that is already above the upper limit more segregated is prohibited.[15] Table 2 lists the types of transactions allowed or not allowed, for each ethnic quota. Transactions involving buyers and sellers from the same ethnicity will always be allowed because this does not increase the ethnic proportion. If a transaction involves units in blocks or neighborhoods that are Chinese-constrained, Chinese buyers cannot buy from non-Chinese sellers because this increases the proportion of Chinese in that neighborhood (similarly for Malay and Indian quotas).

This ethnic-based restriction prevents arbitrage and thus allows prices to differ across ethnic groups in equilibrium. Without the quotas, profit-maximizing sellers will sell to top bidders. With the quotas, even if Chinese buyers are willing to pay more than non-Chinese buyers for units in a Chinese-constrained neighborhood, non-Chinese sellers cannot arbitrage the price differences between Chinese and non-Chinese buyers because non-Chinese sellers cannot sell to Chinese buyers. Therefore, equilibrium prices can differ by ethnic groups. I refer to this as the price discrimination effect of the quota.

3 Data

I use data covering resale transactions in the public housing market in Singapore between April 2005 and March 2006. This dataset encompasses virtually all of Singapore.[16] I

[15] These restrictions are easily enforced because the identity cards of all Singaporeans report their ethnicity. Also, all resale transactions have to be approved by the HDB. One of the approval steps involves checking whether the transaction violates the ethnic housing quotas. During my sample period, an inter-ethnic married couple can choose to use either ethnicities of the spouses.

[16] The analysis only focuses on the public housing market which represents 82% of the citizens and permanent residents in Singapore. To the extent that households with strong ethnic preferences have sorted away from being regulated by the quotas and into the unregulated private housing market, the estimates of ethnic preferences from the resale market would be a lower bound because the public

begin with the universe of all HDB blocks that I obtained from a Census of HDB units. I dropped 88 HDB blocks because I could not match the street address in the HDB Census to HDB neighborhoods (the Census only included block numbers and street names) and I dropped another 152 HDB blocks that were not part of HDB neighborhoods in the Ethnic Integration Policy. In total, I have 8,587 HDB blocks, 170 HDB neighborhoods and 7 markets. A market is a cluster of neighborhoods, categorized according to the Straits Times Real Estate Classifieds (the leading English newspaper in Singapore). The newspaper maps postal sectors to markets (postal sectors are represented by the first two digits of the postal code. Each HDB neighborhood is represented by the postal sector and the fourth digit of the postal code). The number of neighborhoods in each market varies from 12 to 38. Table A1 in Appendix 1 describes the sample sizes at the block, neighborhood and market level in more detail.

Neighborhood attributes

The neighborhood attributes include ethnic proportions (of the stock of public housing residents in a neighborhood), school quality, access to public transportation, the average age of HDB blocks in a neighborhood and the average number of rooms in HDB neighborhoods.

To calculate ethnic proportions, I hand matched more than 500,000 names to ethnicities. The 2005 phonebook was published on April 1st 2005, and includes a total of 795,208 phone listings. I matched names to ethnicities for 589,760 HDB listings (there is a variable in the phonebook that indicates which address is an HDB address). Households can request for phone and address records to be unlisted at a charge of $20 per annum plus a one-time administrative fee of $20. The sample size of the phonebook suggests that a majority of Singapore residents did not opt to be unlisted. The ethnic proportions calculated using the phonebook data are also similar to the national ethnic proportions, suggesting that the choice to list in the phonebook did not differ by ethnic groups.

Of the 589,760 HDB phone listings, I was able to match 588,550 names to ethnicities (a 99.8% match) using differences in the structure of Chinese, Malay and Indian names. For example, most Chinese names only have two or three words; Malay names are primarily Muslim names since 99% of Malays in Singapore are Muslims (Singapore Department of Statistics, 2000); Indian names are matched according to popular first and last names. Nevertheless, 1,210 names remain unmatched. Three listings were firms and 1,207 listings had names that only included initials or first names only or the ethnicity was not identifiable (usually because I could not determine whether the names were

housing market would be a selected sample of households with weaker ethnic preferences.

Indian or Malay names).

The match between names and ethnicity is likely to be most accurate for Chinese names because of distinct last names.[17] On the other hand, Indian and Malay proportions may be more prone to measurement error because many Indian Muslims adopt Arabic names that are very similar to Malay names. Of the matched listings, 493,155 were matched using popular first and last names. Many Chinese names were matched this way. 84% of these 493,155 names were identified as Chinese names, 13% were Malay names and 3% were Indian names. Another 95,395 names were matched individually. 50% of these names were identified as Chinese names, 17% as Malay names and 33% as Indian names. Overall, the ethnic proportions calculated using the phonebook were 79% Chinese, 13% Malay and 8% Indian, very close to the national proportions reported in the 2000 Census (77% Chinese, 14% Malay and 8% Indian).

I collected the remaining neighborhood attributes from online street directories, the HDB website and a non-public dataset purchased from HDB that is a census of all HDB blocks in Singapore. This dataset includes the number of type 1 to type 8 units in each HDB block. See Appendix 1 for definitions of these variables and their sources.

Prices

I collected data on 23,199 transaction prices between April 2005 and March 2006. These prices were uploaded every three months to the HDB website. I averaged these monthly transaction prices to the neighborhood level. Unfortunately, I do not know the ethnicity of the buyer and seller. I address this in Section 3.1.

Location choice data

A choice is a neighborhood. Ideally, I would have choice data of resale buyers and their ethnicity. But, I do not. Instead, I match names in the 2005 phonebook to names in the 2006 phonebook.[18] Each phonebook includes six-digit postal codes that uniquely identify an HDB block. I am not able to identify addresses within an HDB block. The first two digits of the postal code identifies the postal sector. The postal sector together with the fourth digit of the postal code identifies the HDB neighborhood.

I define stayers as households living in the same postal codes in 2005 and 2006. Movers are households who changed their postal code from 2005 to 2006. Entrants are phonebook

[17] Even Chinese Muslims would tend to keep their Chinese last names.

[18] Phonebooks are published every April 1st. So, comparing 2005 and 2006 phonebooks would cover most households who changed phone listings between April 1st, 2005 and March 31st, 2006. But, households have to update their contact information within a month of moving. So, this comparison may include households who moved less than a month before April 1st, 2005 and only updated their phone listing after April 1st, 2005. It may also exclude some households who moved in March, 2006 but did not update their phone listings by March 31st, 2006.

listings that only appeared in the 2006 phonebook.[19] I need to calculate shares for each of the 170 HDB neighborhoods that appear on the Ethnic Integration Policy website. I focus on collecting neighborhood choice data for the sample of 584,483 phone listings residing in these neighborhoods by the publication date of the 2006 phonebook. Using names and postal codes, I was able to identify 524,538 stayers, 20,646 movers and 39,299 entrants.

I use the neighborhood choice data of movers to proxy for the choice data of resale buyers. I was able to identify the ethnicity of 20,572 movers. I did not use the location choice of entrants because the large number of entrants relative to the number of resale transactions suggest that many entrants may not be resale buyers and I have no way to determine which entrant is a buyer. Most probably, these entrants' listings were not identified in the 2005 phonebook because they changed the spelling of their names. The correlation between the number of movers who moved to an HDB neighborhood and the number of entrants in a neighborhood is 0.93 and the correlation between the number of movers in a neighborhood and the number of resale transactions is 0.88. These high correlations suggest that neighborhoods that attract more movers also tend to be neighborhoods that attract more entrants and buyers.[20]

[19] In doing so, I am basically matching names in 2005 to names in 2006, by postal codes. About 8.5% of the data had instances of households sharing the same name residing in the same postal code. Since names are used to define ethnicities, listings with the same names also have the same ethnicities. So, I did not have to worry about the ethnicity, but I had to determine which of the duplicate names were stayers, movers and entrants. As an example, suppose there were two listings (both with the name John Doe) living in postal code 123123 in 2005. In 2006, there were three listings (all with the name John Doe) living in three different HDB postal codes and they appear in the phonebook in the following order: John Doe-123123, John Doe-456456 and John Doe-789789 respectively. There is one stayer in postal code 123123 (there is the possibility that both John Doe's living in 123123 moved to postal codes 456456 and 789789 in 2006 and the John Doe in 123123 in 2006 is an entrant, but I assume that a John Doe moving out and another John Doe moving in is unlikely compared to the possibility that one of the John Doe's was a stayer). Now that I have determined that one of the John Doe's is a stayer, I need to know which of the John Doe's in 456456 or 789789 is the mover and which is the entrant. To do this, I assume that the order in the phonebook is preserved across years. Each phone listing includes an id (essentially a unique number affiliated with each phone listing). If there were multiple John Doe's living in the same postal code in 2005 and 2006, I assume that the order of the names (and the order of their id's) is preserved in 2005 and 2006. I checked this assumption by looking at stayers. There were 153,450 cases of stayers sharing the same name and postal code. I compared the id of these stayers in 2005 and 2006 and checked that the order of their id's are preserved from 2005 to 2006. I sorted these phone listings in increasing order using their id's in 2005, then checked that the id's in 2006 are also in increasing order for a large majority of the cases (this is true for all but 29 listings). Back to the example, since an entrant is a new listing, it would be added after the first two listings. So, John Doe in 456456 is the mover who moved from 123123 and John Doe in 789789 is the new listing (the entrant). This is how I classified stayers, movers and entrants for the 8.5% of data with multiple names in the same postal codes.

[20] The number of movers is different from the number of resale transactions for a variety of reasons that could lead to more or fewer movers, compared to resale buyers in the sample period. First, some of the entrants could be resale buyers but I do not know which entrants are buyers. Second, some buyers

I summarize location choices using **ethnic shares** for each of the 170 HDB neighborhoods. For example, the denominator for the Chinese share for neighborhood j is the number of Chinese residents who lived in HDB neighborhoods, according to addresses in the 2005 phonebook. The numerator counts the number of Chinese movers who moved into neighborhood j, according to their addresses in the 2006 phonebook.

Note that ethnic proportions describe the ethnic distribution of the **stock** of residents while ethnic shares depend on the **flow** of movers. In my analysis, I use **ethnic shares** as a proxy for aggregate location choice probabilities (dependent variable) and **ethnic proportions** as a neighborhood attribute (explanatory variable). In the data, the flow of movers is small enough that the ethnic proportion of the stock of residents is essentially constant within a year.

Quota dummies

I also collected data on whether an HDB block was quota-constrained each month, between March 2005 and April 2006. This data was updated every month on the HDB website. The website only indicates whether an HDB block is constrained but does not specify whether it is because the block or neighborhood quota limit was constrained in that month. If all blocks in a neighborhood are constrained, I know the neighborhood limit is binding.

Early ethnic settlements

I use data on early 19th century ethnic settlements in Singapore to instrument for ethnic proportions. Lieutenant Philip Jackson was appointed to create an urban plan for Singapore, then a British colony. Figure 2 shows the map of early 19th century Singapore according to the Jackson Plan (Crawfurd, 1828). Four separate residential areas were designated for the Chinese, Malays, Indians and Europeans. The Malay and European towns were to the east of the Singapore River while the Chinese and Indian areas were to the west of the river.

Table 3 lists the summary statistics of the full dataset. There are 170 neighborhoods. The ethnic shares are very low (the means for all groups are below 0.5%) indicating

---

in the resale market may have been excluded because they chose to not list in the phonebook or they moved within HDB blocks. Third, some movers may not be resale buyers because they are renters (2% of the HDB market) or they inherited their unit. Finally, there could be a time difference between the recording of a resale transaction on the HDB website and the recording of a phone listing. As mentioned in Footnote 18, comparing the 2005 and 2006 phonebooks would cover most households who changed phone listings between April 1st, 2005 and March 31st, 2006 (including some who moved slightly before April 1st, 2005 and excluding some who moved close to March 31st, 2006). I use resale transactions data for the months of April 2005 to March 2006 (the dates most likely to match the publication of the 2005 and 2006 phonebooks). However, some resale buyers who bought in the later months of the sample period may not have moved yet (they may want to renovate their new unit) and some movers who moved in the beginning of the sample period may have bought their units before April 2005.

that the flow of movers is very low. The Chinese quotas bind for almost one-fifth of the sample, the Malay quotas bind for one-tenth of the sample and the Indian quotas bind for a quarter of the sample. Tables A2 to A3 in Appendix 1 include more summary statistics of the data.

3.1 Estimating ethnic-specific prices

Ideally, if I had information on how much Chinese, Malay and Indian buyers paid for their units respectively, I could test for evidence of price dispersion. In particular, if Chinese buyers paid different prices than non-Chinese buyers for the same neighborhood, this would suggest heterogeneous taste for Chinese neighbors and/or Chinese amenities. For the structural estimation later, I need data on the average Chinese-, Malay- and Indian-price for a neighborhood.

Unfortunately, I do not have data on buyer ethnicity. For an apartment block, I know the prices paid for the sold units from the transactions dataset and I know the ethnicity of the movers from the phonebook data, but I cannot match the prices to the movers within a block. However, at the block level, I can calculate the average price using the transactions data and the likelihood that the buyer is Chinese, Malay and Indian using the phonebook data.

To estimate the average Chinese-, Malay- and Indian-price for a neighborhood, I assume that $P_j^C = P_j(X) + \pi_j^C$, ie. the average Chinese price for neighborhood j is the average price for neighborhood j (as a function of attributes, X) and the Chinese-specific deviation from the average price. I use two steps to estimate these two components.

First, I use transactions level data from the HDB website to estimate the following equation for each neighborhood j.[21]

$$P_{ibjt} = X_{bjt}\gamma_j + C_{bj}\pi_j^C + M_{bj}\pi_j^M + I_{bj}\pi_j^I + \tau_t + \epsilon_{ibjt} \tag{1}$$

where P_{ibjt} is the price for transaction i in block b in neighborhood j in month t, X_{bjt} includes the a constant that is neighborhood-specific, the age of block b, dummies indicating whether the Chinese, Malay and Indian quotas are binding for block b in month t, and dummies that indicate if the transaction involved a type 3, type 4, type 5, type 6 or type 7 unit,[22] C_{bj} is calculated as the number of Chinese movers for block b in

[21] I could also have aggregated the data to the neighborhood level and estimated a conditional expectation function for prices for all neighborhoods, $P(X)$ instead of $P_j(X)$. To allow more flexibility in the price function (so that age could affect prices differently in different neighborhoods, for example) I estimated a separate regression for each neighborhood using variation between blocks in a neighborhood.

[22] Since there are very few type 1 units and very few type 8 units in the transactions data, I combined

neighborhood j, divided by the number of resale transactions for block b in the sample period. Likewise for M_{bj} and I_{bj}. The omitted group counts the difference between the number of transactions and the total number of movers with identified ethnicity. This difference could arise because I could not identify the ethnicity of some movers, or because movers are not resale buyers (see Footnote 20). Including this omitted group allows for the possibility that the ethnic distribution of buyers could be different from the ethnic distribution of movers and acts as a control for blocks where this measurement error is likely to be large.[23]

Once I have estimated the coefficients in (1), I then calculate the average attributes for neighborhood j ($\overline{X}_j$), averaged over all transacted and untransacted units (here, I use the data from the HDB Census). To recover the average Chinese price for neighborhood j, $\hat{P}_j^C = \overline{X}_j \hat{\gamma}_j + \hat{\pi}_j^C + \frac{1}{12} \sum_t \hat{f}_t$, where $\hat{\pi}_j^C$ is the estimated Chinese-specific deviation in price. I repeat the same for Malay and Indian prices. These are the prices I use for the structural estimation. The standard errors will be adjusted to account for the prediction error.

I included all units in the second step because I want to calculate what the average price would be for an average unit in neighborhood j. In the structural model, all observable attributes are assumed to be exogenous, except ethnic proportions and prices. Therefore, the average age, for example, is averaged over all HDB units in a neighborhood, not just transacted units. This is why the price calculation in the second step includes attributes for all units.[24] Using attributes for transacted units in the first step and attributes for all units in the second step allows for the possibility that transacted units are selected and the average transaction price may not reflect the price of the average unit. Suppose older units are less expensive and are less likely to transact in a neighborhood, then, the price for an average unit in that neighborhood should be lower than the average transaction price in that neighborhood because the average unit is older than the average transacted unit.

types 1 and 2 into one category (the omitted category) and I combined types 7 and 8 into one category.

[23]Consider a simple case with two groups only: Chinese and non-Chinese. One possible concern arises if there were more Chinese movers identified in expensive blocks than there were Chinese buyers. Then, measurement error in the Chinese ratio is positively correlated with the dependant variable, leading to an overestimate of $\hat{\pi}_j^C$ (the estimate picks up positive correlation between the Chinese ratio and prices and also between the measurement error and prices).

[24]I wish to thank a referee for making this suggestion.

4 Price Effects close to the Quota Limits

I present a descriptive analysis of the behavior of prices above and below the quota in the spirit of the "regression kink design" (Card et al., 2009). The strategy is to identify kinks in the outcome variable that coincide with kinks in the policy rule.[25]

I test whether there is a jump in the probability that the quota binds at the block quota limit. To do this, I regress the monthly quota status from the HDB website on a dummy that is 1 if ethnic proportions are above the block quota limit (measured using the phonebook data) while controlling for fourth order polynomials of ethnic proportions. Standard errors are clustered at the block level. Figure 3a summarizes the effect of being in a block with 87% or more Chinese on the probability that the Chinese quota binds in a month. Figures 3b and 3c measure the same effect for Malay and Indian block quotas.[26]

There is a statistically significantly positive discontinuity in the probability that the quota binds, right at the policy thresholds associated with the Chinese, Malay and Indian quotas. The magnitudes of the jumps are 23%, 12% and 10% for the Chinese, Malay and Indian quotas (all significant at the 1% level). The probability that the quota binds is greater than 0 below the quota limits and less than 1 above the quota limits due to two reasons. First, there is time series variation because the quota data from the HDB website (vertical axis) is monthly and the phonebook data (horizontal axis) is annual. Conditional on the ethnic proportions from the phonebook data, whether a quota is binding or not can change from month to month. Secondly, there is measurement error in the matching of names to ethnicities, as discussed in the data section. The noise introduced by the measurement error would bias against finding discontinuities unless the measurement error is correlated with the quota status (non-classical measurement error), which seems unlikely.

I proceed in two steps: First, I test whether there is price dispersion using observed prices close to the quota limit. Second, I test whether the quota had price discrimination effects (ie. whether prices differed by buyer ethnicity for constrained neighborhoods).[27]

[25]See Guryan (2003); Nielsen et al. (2009); Simonsen et al. (2009) for examples of non-parametric identification of outcomes using kinked policy rules. See Footnote 6 for comparisons to regression discontinuity design.

[26]Figures 3b and 3c show a kink in the predicted probability around 50% above the Malay quota limit (25%) and the Indian quota limit (13%) because there are too few blocks with more than 75% Malays and more than 63% Indians, given that Malays and Indians are only 14% and 8% of the population.

[27]The focus of this section is to demonstrate the impact of the quota on prices because this is an important assumption for the structural estimation. In a separate paper, I analyze the effect of the quota on outcomes other than price Wong (2012).

4.1 Is there a discontinuity in observed prices?

I begin with a transaction level analysis that compares transactions for units that are constrained and unconstrained:

$$\ln P_{ibjtk} = \alpha + \gamma QC_{bjt-1,k} + f(\text{percent}C_{bjk}) + \varepsilon_{ibjtk} \tag{2}$$

$$\ln P_{ibjtk} = \alpha + \gamma QC_{bjt-1,k} + f(\text{percent}C_{bjk}) + B_{bjk}\beta + \tau_t + \omega_k + \varepsilon_{ibjtk} \tag{3}$$

where $\ln P_{ibjtk}$ is the log of the price of unit i in block b, neighborhood j, town k and month t; $QC_{bjt-1,k}$ is a dummy for whether the Chinese block or neighborhood quotas are binding in the previous month,[28] $f(\text{percent}C_{bjk})$ are polynomials of the percent of Chinese in a block (calculated from the phonebook), centered around the block quota limit; B represents other observable attributes of the block (age of block b and age^2, share of units of block b that is of each of the eight HDB types (type 1 is omitted)); τ_t and ω_k are month and HDB town fixed effects.

The coefficient of interest is γ, which summarizes the price effects. I estimate these equations for units that are 10% above and below the Chinese quota limits and I repeat the analysis for the Malay and Indian quotas. For each ethnic quota, I estimate the regressions controlling for fourth order polynomials of ethnic proportions.[29] In Table 4, columns 1 and 2 report estimates for equations 1 and 2 for the Chinese quota, columns 3 to 4 report estimates for the Malay quota and columns 5 and 6 report estimates for the Indian quota. Standard errors are clustered at the HDB block level.

If sellers could arbitrage perfectly, we should expect no price effects. However, I find robust and statistically significant evidence of discontinuities in prices, ranging from magnitudes of 3% to 8%.[30]

[28] Each resale transaction begins with a resale application submitted to HDB for approval. The application has to show that the buyers are citizens or permanent residents, and the transaction satisfies the quota policy at the time of the application. Approval can take up to 2 months. I have matched transaction prices in a month to quota constraints in the past month, assuming that approval takes 1 month. The results are similar if I use a 2-month lag.

[29] This table reports estimates using polynomials that are the same above and below the limits. The Chinese quota and Malay quota estimates are robust to allowing the polynomials to differ (see Appendix 3). The estimates for the Indian quotas are not significant with separate polynomials, if standard errors are clustered at the block level.

[30] The overall effect depends on the elasticities of demand for buyers of different ethnic groups (to be estimated in the structural model) and also the shares of transactions involving buyers of different ethnic groups. A price discrimination model with two types of buyers, Chinese and non-Chinese, with different location preferences (the Chinese prefer to live in Chinese neighborhoods), will show that prices paid by non-Chinese buyers are likely to be weakly lower and prices paid by Chinese buyers are likely to be weakly higher than unconstrained units. A majority of the transactions will likely involve Chinese buyers, so that the overall price effect is likely to be heavily weighted towards a positive sign. This is because (a) non-Chinese buyers are likely to buy from non-Chinese sellers since Chinese sellers who can

4.2 Do estimated prices differ across ethnic groups?

Next, I test whether the quotas had any price discrimination effects. For example, whether the prices paid by Chinese and non-Chinese buyers differed for Chinese-constrained neighborhoods. This neighborhood-by-ethnic group variation in prices is important for identification of the location choice model in Section 5.

For unconstrained units, profit-maximizing sellers would sell to the top bidder. For constrained units, the price effects depend on how willingness-to-pay (WTP) differs by ethnic groups. If Chinese and non-Chinese buyers had similar WTP, then, there would be no price discrimination effects because non-Chinese sellers who cannot sell Chinese-constrained units to Chinese buyers would just sell the units to non-Chinese buyers who have similar WTP. Similarly, if non-Chinese buyers were willing to pay more than Chinese buyers, there would not be any price discrimination effects because the transaction that the quota bans does not include the buyer with the higher WTP. With or without the quotas, non-Chinese sellers would sell to non-Chinese buyers because they have the higher WTP.

Therefore, we only see price discrimination effects if non-Chinese buyers are willing-to-pay *less* than Chinese buyers for Chinese-constrained units. In this case, non-Chinese sellers would have to lower the price of Chinese-constrained units to attract non-Chinese buyers. Comparing constrained and unconstrained neighborhoods, prices for Chinese constrained neighborhoods should be *lower* as the share of non-Chinese buyers increase. To test the predictions above, I averaged transaction level prices to the neighborhood-month level and interacted the neighborhood quota dummy with ethnic buyer weights that were calculated using the phonebook data. These regressions can be viewed as heterogeneous treatment effect regressions to the average treatment effect regressions in (2) and (3).

$$\ln P_{jt} = \gamma nQC_{jt-1} + \varphi_1 nQC_{jt-1} * wM_j + \varphi_2 nQC_{jt-1} * wI_j + f(percentC_j) + X_j\beta + \tau_t + \varepsilon_{jt} \quad (4)$$

where $\ln P_{jt}$ is the log of average transaction prices for neighborhood j in month t, nQC_{jt-1} is 1 if the Chinese neighborhood quota was binding for neighborhood j in month $t-1$, wM_j and wI_j are the shares of movers for neighborhood j who are Malay and Indian (from the phonebook data), $f(percentC_j)$ is a fourth order polynomial of the Chinese

sell to both Chinese and non-Chinese buyers will prefer to sell to Chinese buyers who are willing to pay more and (b) at most 13% of the sellers are non-Chinese because the Chinese block quota limit is 87% Chinese. A similar argument for Malay and Indian quotas suggest that the overall price effects are heavily weighted towards a negative sign because the Malay and Indian quota limits are only 25% and 13%. These predictions are consistent with the signs of the discontinuities in Table 4.

neighborhood proportion centered around the neighborhood quota limit (84% Chinese), X_j includes observed neighborhood attributes (the average number of rooms, school quality, average age of buildings, average age^2, distance to the closest subway) and τ_t are month fixed effects. Standard errors are clustered at the neighborhood level. If WTP of non-Chinese buyers are lower than WTP of Chinese buyers for Chinese-constrained neighborhoods, we would expect φ_1 and φ_2 to be negative. This would suggest that the treatment effect of the quota results in lower prices for Chinese-constrained units as the share of non-Chinese buyers increases.

There are 2 drawbacks to this price dispersion test that would bias against finding price discrimination effects. First, there is less statistical power because these regressions are at the neighborhood-month level, instead of the transaction level. Second, there is measurement error in the buyer weights. Table 5 reports the results for Chinese, Malay and Indian quotas. The key coefficients are the interaction terms between the quota dummy and the buyer weights.

I find evidence of price discrimination effects that are consistent with the ethnic-based restrictions in Table 2. Column 1 shows that as the Malay buyer weight increases by 1 percentage point, prices for Chinese-constrained neighborhoods fall by 1.53% (p-value: 5%). The coefficient for the share of Indian buyers is also negative but not significant. Both coefficients (φ_1 and φ_2) are jointly significant (p-value: 1%). For Malay quotas, prices fall by 0.662% as the Chinese weight increases by 1 percentage point (p-value: 5%) but prices increase as the Indian weight increases (not significant). For Indian quotas, prices fall as Malay and Chinese buyer weights increase but the coefficients are only jointly significant (p-value: 1%). In summary, 5 out of the 6 φ_i's are estimated to be negative.

These suggestive findings of price discrimination indicate that WTP differs by ethnic groups. Otherwise, there would be no price differences to arbitrage in the first place. These WTP's can differ because taste for observed neighborhood attributes or taste for unobserved amenities differ by ethnicity. In the next section, I build a location choice model that allows for ethnic-based taste for observed and unobserved amenities. In Section 6, I use instrumental variables to identify these preferences separately.

5 Utility Specification

To recover ethnic preferences away from the discontinuity, I begin with a random coefficients utility model of households choosing neighborhoods. There are $m = 1, ..., M$ markets, each with $i^G = 1, ..., I_m^G$ buyers of ethnic group G and $j = 1, ..., J_m$ neighborhoods. The indirect utility of buyer i of group G from choosing neighborhood j in market m is

$$U^G_{ijm} = X^G_{jm}\beta^G_i + \alpha^G_i P^G_{jm} + \xi^G_{jm} + \varepsilon^G_{ijm} \tag{5}$$

for $j = 1, ..., J_m$ $\forall m$, where X^G_{jm} is a K-dimensional (row) vector of observed neighborhood attributes that could include attributes common to all ethnicities as well as ethnicity-specific attributes; P^G_{jm} is the price that a buyer of group G pays for a unit in neighborhood j in market m; ξ^G_{jm} is taste for the unobserved neighborhood amenity that is specific to ethnic group G and ε^G_{ijm} represents mean-zero, idiosyncratic individual preferences for a consumer of group G. The observed neighborhood attributes include a constant, the average age of HDB buildings, school quality, the average number of rooms of HDB units in the neighborhood, distance to the closest subway station, percent own ethnic neighbor, the squared term and ethnic prices.

I make two departures from standard residential location choice models: I allow taste for unobserved neighborhood amenities to vary by ethnic groups, ξ^G_j. The ξ^G_j's aggregate up the impact of all attributes that are unobserved by the econometrician but affect group G's utility for neighborhood j. This specification allows the interpretation of taste for Chinese neighbors, $\overline{\beta}^C_{percentChinese}$, that is separate from taste for unobserved amenities that are preferred by Chinese (I call these Chinese amenities).[31] Examples of ethnic-specific amenities include kindergartens that teach ethnic languages, places of worship, community centers that set aside space for cultural events and activities for different ethnic groups (eg. Tai-chi for Chinese, sepak takraw courts for Malays and cricket fields for Indians).[32] In a qualitative study of ethnic relations, Singaporeans indicated a preference for "special ethnic community places", suggesting that ethnic based taste for amenities could be important (Lai, 1995).

Secondly, notice also that prices vary across ethnic groups, P^G_{jm}, because of the price discrimination mechanism of the quotas. To keep the notation simple, I will drop the subscript for markets from here on. An important assumption that is common in this literature is that utility from neighborhood j only depends on the attributes of that

[31]The identification strategy in this paper relies on the assumption that the taste for unobserved neighborhood amenities is additive and separable, a standard assumption in many random coefficient discrete choice models. Without additivity and separability, I will not be able to identify ethnic preferences. For example, if amenities and ethnic proportions enter the utility function jointly (eg. Chinese derive utility from living with Chinese neighbors and living near Chinese amenities, but they also derive more utility from Chinese amenities if there are more Chinese neighbors), the coefficient on the percent of Chinese in a neighborhood will be over-estimated under the assumption of an additive and separable utility function.

[32]Sepak takraw is a traditional Malay sport.

neighborhood alone.[33] This will be useful for identification, as I discuss below.

One limitation of the specification is the absence of income in the model because I do not observe income. Admittedly, this is unrealistic for housing. An immediate implication is the inability to distinguish between income- and ethnic-segregation. I return to this in the results section. Second, like most papers in the literature on residential segregation, this location choice model is static. A dynamic discrete choice model would require panel data. Third, this model estimates taste for own-ethnic-neighbors (Chinese taste for Chinese versus non-Chinese), and cannot distinguish the Chinese taste for Malays versus the Chinese taste for Indians. I chose this specification because I only have a small number of neighborhoods. In most theoretical models (including Schelling's), taste for own-ethnic-neighbors is thought to be the main driver of residential segregation.

We can write the consumers' taste parameters as a mean and a consumer-specific deviation from the mean

$$\begin{pmatrix} \beta_i^G \\ \alpha_i^G \end{pmatrix} = \begin{pmatrix} \bar{\beta}^G \\ \bar{\alpha}^G \end{pmatrix} + \Sigma v_i^G \tag{6}$$

where $v_{i1}^G, ..., v_{iK}^G$ is individual i's unobserved taste for attribute k, drawn independently (for each individual in each group) from a standard normal distribution and v_{iP}^G is drawn from a log normal distribution because I assume that all individuals do not like to pay high prices. Σ is a (K+1)x(K+1)-dimensional scaling matrix whose diagonal elements are denoted by σ_k and σ_P.[34]

The specification is completed with the introduction of an 'outside good' ($j = 0$) — buyers may choose not to move to another HDB neighborhood. The utility of the outside good is normalized to 0.

$$U_{i0}^G = \xi_0^G + \varepsilon_{i0}^G \tag{7}$$

Substituting (6) into (5) and grouping consumer-specific terms together, we can write the utility specification parsimoniously as $U_{ij}^G = \delta_j^G + \mu_{ij}^G$ which is simply group G's mean utility for neighborhood j

[33]This excludes utility specifications where buyers have higher utility if their neighborhood is better than adjacent neighborhoods.

[34]Note that I assume mean preferences vary by ethnicity but the standard deviation does not (Σ is not indexed by G). This is mostly a limitation of my data. Identification of Σ relies on variation in choice sets across markets (Petrin, 2002; Berry et al., 2004). While the choice sets differ substantially because the neighborhood attributes differ a lot, I only have 7 markets in my data.

$$\delta_j^G = X_j{}^G\bar{\beta}^G + \bar{\alpha}^G P_j{}^G + \xi_j^G \tag{8}$$

and a consumer-specific deviation from that mean

$$\mu_{ij}^G = \sum_k \sigma_k x_{jk}^G v_{ik}^G + \sigma_P P_j{}^G v_{iP}^G + \varepsilon_{ij}^G \tag{9}$$

The parameter, σ, is commonly thought of as a measure of heterogeneity. As σ increases, neighborhoods that are similar in attributes become better substitutes because individuals with strong tastes for attribute k (υ_{ik}) will tend to substitute towards products that are abundant in attribute k (x_{jk}). Assuming unobserved heterogeneity avoids the independence of irrelevant alternatives (IIA) problem.

Market-level aggregates are obtained by aggregating over the distribution of consumer characteristics. Let the group G share for neighborhood j be

$$s_j^G(\delta^G, \theta^G; x^G, P^G, F_{\mu^G}) = \int_{A_j^G(\delta^G,\theta^G;\ x^G, P^G)} F_{\mu^G}(d\mu^G), \tag{10}$$

where A_j^G is the set of consumers of group G who choose neighborhood j, θ^G is the set of taste parameters, $\{\bar{\beta}^G, \bar{\alpha}^G, \sigma\}$, and F_{μ^G} denotes the population distribution function for ethnic group G.

$$A_j^G(\delta^G, \theta^G;\ x^G, P^G) = \left\{\mu^G : U_{ij}^G > U_{ij'}^G \ \forall j' \in J\right\} \tag{11}$$

Following the literature, I assume that the idiosyncratic errors, ε_{ij}^G, have an independently and identically distributed Type I extreme value distribution. Integrating out the ε's yields the Logit form for the model's choice probabilities. Letting s_{ij}^G denote the probability that individual i of group G chooses neighborhood j,

$$s_{ij}^G = \frac{\exp(\delta_j^G + \sum_k \sigma_k x_{jk}^G v_{ik}^G + \sigma_P P_j{}^G v_{iP}^G)}{1 + \sum_{j'} \exp(\delta_{j'}^G + \sum_k \sigma_k x_{j'k}^G v_{ik}^G + \sigma_P P_{j'}^G v_{iP}^G)} \tag{12}$$

6 Empirical framework

6.1 Estimation

Step 1: Contraction mapping ($\delta_j^G = s^{-1}(s_j^{G,sample})$)

I estimate 3 separate discrete choice models, one for each ethnic group. I use a contraction mapping algorithm from Berry (1994) to find the value of δ^G that makes the observed ethnic shares match the shares predicted by the model.[35] The ethnic share is calculated as the number of group G movers who chose to move to neighborhood j divided by the number of group G households in a market.[36] I simulate the integral in (10) by drawing 10,000 v_i^G's independently for each group G. The assumption in these models is that market shares are a non-linear and non-separable function of ξ^G, but mean utility, by assumption, is a linear combination of observed and unobserved neighborhood attributes. Using the contraction mapping algorithm, one can recover mean utility as the dependent variable of a linear equation, $\delta_j^G = X_j^G \bar{\beta}^G + \bar{\alpha}^G P_j^G + \xi_j^G$, that can then be estimated using instrumental variables.

Step 2: Method of simulated moments

Once δ^G has been recovered from step 1, I can construct the moments. I recover the taste parameters, θ^G, by matching aggregate moments predicted from the model to sample moments using the Method of Simulated Moments. The following moment condition is assumed to hold at the true parameter value, $\theta_0 \in R^p$:

$$E[g(\theta_0)] \equiv E[\xi(\theta_0)\, Z] = 0 \tag{13}$$

where $g(\bullet) \in R^l$ with $l \geq p$ is a vector of moment functions that specifies that the structural error, ξ, is uncorrelated with the instruments, denoted by an JxL matrix, Z.

I stack the moments for the estimation of each ethnic group and define $\theta = \{\theta^C, \theta^M, \theta^I\}$. The simulated moments are

$$\sum_{j=1}^{J} \hat{g}_j(\theta) = \sum_{j=1}^{J} Z_j \hat{\xi}_j(\theta) \tag{14}$$

The MSM estimator, $\hat{\theta}$, minimizes a weighted quadratic form in $\sum_j^J g_j(\hat{\theta})$. I use a 2-step estimator where the second step uses estimates from the first step to calculate a consistent weighting matrix. The standard errors are calculated based on Pakes and Pollard (1989) and McFadden (1989). To account for the error from using estimated prices instead of observed prices, I follow Newey (1984). Since the price and taste parameters are estimated sequentially, the first set of moments come from the OLS price regressions in Section 3.1, with parameters π and γ (the first moments are $g^1(\pi, \gamma, \tau)$). Then, using these parameters as inputs, the second moments are the moments in (14),

[35]To find the fixed point, I use a tolerance of $1 \times e^{-14}$.

[36]There are two neighborhoods with no Chinese movers, eleven with no Malay movers, four with no Indian movers. I assign these shares to be the minimum share for each ethnic group, because the estimation involves the inversion of ethnic shares and shares of neighborhoods that are zero are not invertible.

$g^2(\theta, \hat{\pi}, \hat{\gamma}, \hat{\tau})$.

6.2 Identification

In this section, I discuss identification of the mean utility equation, $\delta_j^G = X_j^G \overline{\beta}^G + \overline{\alpha}^G P_j^G + \xi_j^G$. Identification of the unobserved heterogeneity parameters (σ in μ_{ij}) follows Berry et al. (1995). I first explain how I use variation from the quotas to identify a location choice model where taste for neighborhood amenities vary across ethnic groups (ξ_j^G). Then, I describe the instruments used for endogenous variables in the mean utility equation: ethnic proportions and ethnic prices. Table 6 summarizes the identification assumptions behind 6.2.1 to 6.2.3. My examples describe identification for Chinese preferences, assuming only 2 types of households (Chinese and non-Chinese). Identification for Malays and Indians are similar.

6.2.1 Identification of ξ_j^G

The model in this paper improves over existing location choice models by allowing the mean utility for neighborhood amenities to vary by ethnic group. Consider a specification commonly assumed in existing location choice models:[37]

$$u_{ij} = \delta_j + \sum_k x_{jk} d_{ic} \beta_k^{OC} + \epsilon_{ij} \tag{15}$$

$$\delta_j = x_j \overline{\beta} + \overline{\alpha} P_j + \xi_j \tag{16}$$

where δ_j is the mean utility for neighborhood j, averaged over Chinese and non-Chinese, ξ_j is the average household's taste for unobserved amenity, d_{ic} is 1 if household i is Chinese, β_k^{OC} is the Chinese-specific marginal utility for attribute k where β^O indicates observed heterogeneity, and the other variables and parameters are defined in the same way as Section 5.

In this model, the Chinese marginal utility for percent Chinese, is $\overline{\beta}_{percentC} + 2\overline{\beta}_{percentCsq} * percentC + \beta^{OC}_{percentC} + 2\beta^{OC}_{percentCsq} * percentC$, the mean marginal utility plus the Chinese-specific deviation from the mean. The endogeneity problem of unobserved amenities, is circumvented here because ξ_j is "absorbed" in δ_j. Once δ_j is recovered using contraction mapping, the ethnic preferences of the average household can be identified separately from ξ_j using instrumental variables.

[37] Assume no unobserved heterogeneity. The intuition is the same with unobserved heterogeneity.

Importantly, the observed heterogeneity terms, $\beta^{OC}_{percentC}$, are identified by matching the covariances between $percentC$ and Chinese location choices without instrumental variables.[38]

The problem arises if Chinese taste for unobserved amenities is different from non-Chinese taste for unobserved amenities. Intuitively, this model only allows taste for observed neighborhood attributes (eg, $percentC$) to differ by ethnic groups. This will be the only margin in the model that explains variation in the location choices of Chinese movers. If neighborhood j has abundant Chinese amenities, the sample moment would indicate a high probability that Chinese choose neighborhood j. To match the sample moment, $\beta^{OC}_{percentC}$ will have to be higher because in the model, Chinese movers disproportionately choose one neighborhood over another only because observed neighborhood attributes that are preferred by Chinese are abundant in that neighborhood. This results in an upward bias in $\beta^{OC}_{percentC}$. A model that allows taste for unobserved amenities to differ by ethnic groups would address this problem.

However, estimating such a model is challenging because it is under-identified in many empirical settings. The standard empirical strategy in discrete choice models is to correlate neighborhood choices with ethnic proportions. Without the quotas, the model would be under-identified because the neighborhood choices of Chinese movers would be positively correlated with both the Chinese taste for observed Chinese proportions and the Chinese taste for unobserved amenities.

The ethnic quotas are useful in this setting because of the price discrimination mechanism of the policy. Even after accounting for Chinese buyers' taste for Chinese neighbors, if Chinese have a stronger taste for unobserved amenities in that neighborhood, then, they would be willing to pay more than non-Chinese buyers for units in that same neighborhood. Therefore, a model with both Chinese taste for own-ethnic-neighbors and Chinese taste for unobserved amenities is no longer under-identified because there are now two sources of neighborhood-by-ethnic group variation: (i) the location choices of Chinese movers; (ii) the within neighborhood variation between Chinese and non-Chinese prices.

6.2.2 Instruments for ethnic proportions

Note that quota dummies are not instruments for ethnic proportions for three reasons. First, the policy rule is itself a function of ethnic proportions. Second, within the time

[38]Formally, $\beta^{OC}_{percentC}$ is identified by matching the covariance between $percentC$ and the difference between the sample moment, $\frac{\text{Number of Chinese movers choosing neighborhood j}}{\text{Number of movers choosing neighborhood j}}$, and the same moment predicted by the model. This follows equation (8) in Berry et al. (2004).

period in my data (April 2005-March 2006), the policy does not generate exogenous variation in neighborhood ethnic proportions because the flows of movers are too small to change the ethnic make up of residents in a neighborhood. Third, the probability that a quota binds is likely to be correlated with amenities. The Chinese quota is likely to bind in neighborhoods with high Chinese proportions or neighborhoods with amenities that the Chinese prefer.

Historical ethnic settlements

My first instrument for ethnic proportions relies on historical ethnic settlements created in the 19th century by the Jackson Plan (described in Section 3). The instrument is a dummy variable that is 1 for units to the east of the early Malay settlements. The idea is that the Jackson Plan's assignment of Malay settlements to the east of the Singapore River (as shown in Figure 2) increased the likelihood that subsequent Malay neighborhoods would be developed on the east side of the Singapore River. But, conditional on the percent of Malays in the neighborhood, this assignment is assumed to be independent of Malay amenities.

Figure 4a shows the spatial distribution of quota-constrained neighborhoods in 2005. Malay-constrained neighborhoods are primarily in the east while the Chinese and the Indian neighborhoods are not. This suggests that Malay neighborhoods indeed expanded to the east of the river after the Malay settlements were assigned there.

Figure 4b is a plot of the spatial distribution of mosques, an important Malay amenity because 99% of Malays are Muslims. While the exclusion restriction is not testable and it is impossible to collect data for all neighborhood amenities, it is reassuring that the spatial distribution of mosques does not seem to be concentrated to the east of early Malay settlements.

Attributes of nearby neighborhoods

My second set of instruments for ethnic proportions (the sum of exogenous attributes of nearby neighborhoods) follows Bayer and Timmins (2007) and Bayer et al. (2004). According to them, attributes of surrounding neighborhoods could be correlated with ethnic proportions if Chinese, Malays and Indians have different preferences for neighborhood attributes, perhaps due to demographics such as family sizes.[39] The thought experiment involves two similar neighborhoods: A and (little) a. Neighborhood A has big units relative to the surrounding neighborhoods and neighborhood (little) a has small units relative to the surrounding neighborhoods. Malays would tend to sort into neighborhood

[39]Forty-three percent of Malay households have five or more family members while only 24% and 26% of Chinese and Indian households have such large families (Housing Development Board, 2000).

A since Malays prefer big units to the surrounding small units. In this manner, unit sizes of surrounding neighborhoods are correlated with Malay neighborhood proportions.

The exclusion restriction for this instrument is satisfied because of the separability assumed in the functional form of the mean utility equation: $\delta_j^G = X_j^G \bar{\beta}^G + \bar{\alpha}^G P_j^G + \xi_j^G$ implies that $X_{j \neq j}$ are not correlated with ξ_j^G by assumption. This exclusion restriction is commonly assumed in similar demand estimation models for products such as cars or cereal (Berry et al. (1995, 2004); Petrin (2002) and Nevo (2001)). To adapt it to location choice models, where the taste for adjacent neighborhoods could be non-zero because of spatial correlation, I follow Bayer et al. (2004) and use exogenous attributes of neighborhoods (everything except price and ethnic proportions) that are nearby, but not adjacent (ie. within 1-3km rings, 3-5km rings and 5-7km rings).[40]

6.2.3 Instruments for ethnic prices

Predicted quota probabilities

I use the quota policy to create new instruments for ethnic-specific prices that are conditionally mean independent from amenities. The ideal experiment is to randomly assign quota probabilities (the probability of whether a quota binds or not) across blocks and neighborhoods so that the quotas bind for exogenous reasons. The price discrimination mechanism of the policy will then generate plausibly exogenous variation in ethnic-specific prices.

In practice, quotas bind for both endogenous and exogenous reasons because ethnic proportions could be highly correlated with ethnic tastes for neighborhood amenities. I approximate the ideal experiment by using the instruments for ethnic proportions above and the step function of the policy rule to isolate plausibly exogenous variation in ethnic proportions, and hence, variation in quota probabilities. These quota probabilities are estimated using three steps.

First, I estimate the block and neighborhood level ethnic proportions using the instruments for ethnic proportions described above:

$$percentG_{bj} = Z_{bj}\gamma + u_{bj} \tag{17}$$

$$percentG_j = Z_j\rho + v_j \tag{18}$$

where G=(C)hinese, (M)alays and (I)ndians, b indexes blocks and j indexes neighbor-

[40]I chose 1km as the cutoff because the neighborhoods would be far enough. I chose 2km widths so that all neighborhoods would have at least one nearby neighborhood within the ring. One neighborhood, Changi Village, is located at the Eastern tip of Singapore. There are no neighborhoods within 1-3km of Changi Village. I assign values of the instruments to be zero for Changi Village.

hoods. The variable, $percentG$ is the percent of residents from group G, the instruments are $Z = [X^{ex}\ East\ X^{ex\square}]$, where X^{ex} is the set of exogenous attributes for neighborhood j (everything in X except price and ethnic proportions), $East$ is a dummy for whether the neighborhood is to the east of the early Malay settlements, $X^{ex\square}$ is the sum of exogenous attributes of nearby neighborhoods. For equation (17), Z_{bj} includes $East$, $X^{ex\square}$, the average number of rooms in block b in neighborhood j, the average age of block b, and other exogenous attributes in X^{ex} at the neighborhood level (school quality and the distance to public transportation).

In the second step, I use the predicted ethnic proportions in step one and the block and neighborhood quota limits in Table 1 to determine whether a block and a neighborhood are quota-constrained. That is, I define a block quota dummy that is 1 if the predicted block level ethnic proportion is above the quota limit for blocks, $Q\hat{G}_{bj} = 1(perc\hat{e}ntG_{bj} \square$ block quota for group G), and similarly for the predicted neighborhood quota dummies $(Q\hat{G}_j)$.

Finally, I calculate the probability that the quota is binding in a neighborhood using the following cases: (i) the predicted probability that a quota is binding in neighborhood j is 1 if the predicted neighborhood quota dummy $(Q\hat{G}_j)$ is 1; (ii) it is 0 if the predicted neighborhood quota dummy is 0 and the predicted block quota dummies $(Q\hat{G}_{bj})$ for all blocks in neighborhood j are 0; (iii) otherwise, the predicted quota probability is the average of the predicted block quota dummies, averaged using all blocks in the neighborhood.

These three steps are designed to isolate plausibly exogenous variation in the probability that a quota binds because the ethnic proportions in step 1 were predicted using exogenous variables only and the quota limits used in step 2 are not correlated with amenity levels in a neighborhood (the quota limits were set in 1989 to be the same for all blocks and neighborhoods). Therefore, the exclusion restriction is satisfied.

What remains to be shown is that estimated quota probabilities are correlated with ethnic prices (the analog of a "first stage"). The concern is that these probabilities were estimated using other instruments, Z. Conditional on Z, the predicted quota probabilities may not be correlated with prices if there is no variation in the predicted probabilities, after controlling for Z (ie. there may not be a "first stage"). Here, the step function of the policy rule introduces non-linearities that are useful. Essentially, the first step of the prediction exercise estimates ethnic proportions as a linear combination of the instruments, Z. However, the second step constructs the neighborhood and block level quota dummies using the policy rule. Thus, the effect of the estimated quota probabilities is non-parametrically identified using the step function of the quota policy.

As long as the instruments do not affect prices in the same way as the step function, the estimated quota probabilities should have an effect on prices.

To test this, I estimate the following equations

$$\Pr(QG_j) = \varphi_1 \widehat{\Pr}(QG)_j + \upsilon_j \tag{19}$$

$$\Pr(QG_j) = Z_j \gamma + \varphi_2 \widehat{\Pr}(QG_j) + \omega_j \tag{20}$$

That is, I regress the actual quota probabilities for Chinese, Malay and Indian quotas, $\Pr(QG_j)$, on the estimated quota probabilities $\widehat{\Pr}(QG_j)$. For example, $\Pr(QC_j)$ is the percent of blocks in neighborhood j where the Chinese quota is binding. Then, I repeat the same regression, controlling for the instruments used to estimate the quota probabilities. If the estimated quota probabilities have power above and beyond the exogenous attributes used to estimate them, then, the coefficient, φ_2, should be significant. This regression is akin to the first stage of an instrumental variables regression except that the dependent variable is not ethnic-specific prices (what the quotas instrument for), because I do not observe ethnic-specific prices.

Panel A of Table 7 reports the results for equations (19) and (20). The first 3 columns show that the estimated quota probabilities are all positive and significantly correlated with actual quota probabilities even though the actual quota probabilities were not used to estimate the quota probabilities. Importantly, the next 3 columns show that after controlling for the instruments used to estimate these quota probabilities, the estimated Chinese and Malay quota probabilities still have a significant effect, indicating that the step function is powerful. But not so for the Indian quotas perhaps because of a lack of statistical power (see Footnote 29).

To test that the effect is non-parametrically identified from discontinuities due to the quota limits, Panels B and C repeat the same regressions using placebo limits that are expected to have no effect on prices. I use placebo limits that are 3% above and below the actual limits. Column 4 in Panels B and C show that the coefficient on the predicted Chinese quota probabilities is no longer significant, after controlling for the instruments for ethnic proportions. This shows that the estimates for Chinese quotas in Panel A are identified from the policy rule.

However, the placebo test suffers from several drawbacks in the context of Malay and Indian quotas. First, placebo limits that are too far above the actual limits are not useful. For example, few neighborhoods were estimated to be Malay-constrained and all Indian quota probabilities were estimated to be zero when the placebo limit was 3% above the actual limit. Accordingly, Panel B shows that predicted Malay quota probabilities are negative and significant using placebo limits that are 3% above the actual limits. The corresponding coefficient for Indian quotas was dropped because there was no variation in the predicted Indian quota probabilities. But, the coefficient on Indian quota probabilities, predicted using placebo limits 3% below actual limits, is not significant after controlling for instruments (Panel C, column 6).

Secondly, placebo limits that are slightly below the actual limit could result in estimated quota probabilities that are still highly correlated with actual quota probabilities because of measurement error. This issue is more severe for Indians and Malays because their names are harder to match, as discussed in Section 3. This could be why the Malay quota probabilities constructed using placebo limits that are 3% below actual limits are still positive and significant (Panel C, column 5). For these reasons, these placebo tests are most valid for the Chinese quotas.

Predicted prices

My second instrument for prices is the same as the price instruments used in Bayer et al. (2007). The first step uses the price instruments (Z and the predicted quota probabilities) to estimate a demand model with exogenous variation only (no social interactions). This location choice model without social interactions has a unique equilibrium price. The second step uses contraction mapping to solve for the unique vector of equilibrium prices that clears the market. This price vector is a weighted average of the price instruments where the method relies on equilibrium conditions of the model to approximate the optimal weights for the instruments.

7 Results

Table 8 reports estimates for the taste parameters.[41] The top panel reports results on the mean of the taste parameters, $\bar{\beta}$ and $\bar{\alpha}$, and the bottom panel reports results on the heterogeneity term, σ. The first column refers to estimates that are restricted to be common across groups and the next three columns are preference parameters for Chinese, Malays and Indians. Because of the small number of neighborhoods, I constrained all coefficients other than ethnic preferences to be common across ethnic groups.[42]

[41] See Appendix 3 for the Logit model estimated with OLS and IV where the dependent variables are the log of the ethnic shares, $\ln(s^G_{jm})$ subtracted by the log of the ethnic share for the outside good, $\ln\left(s^G_{0m}\right)$ and G indexes for the (C)hinese, the (M)alays and the (I)ndians. I estimated all three Logit models simultaneously. The Logit model is nice because it is computationally simple and transparent. I use the Logit model to estimate models with linear, quadratic and cubic ethnic preferences. The results show non-linearities in ethnic preferences (columns 6-8 and 10-12).

[42] For example, this restricts Chinese, Malays and Indians to share the same tastes for school quality. Suppose tastes for school quality differed by ethnicity. In this case, the reported coefficient in Table 8

In general, the signs of most of the coefficients are as expected: Higher school quality, shorter distances to the subway, newer HDB blocks and lower prices are all associated with higher mean utility. The magnitudes of the mean utility parameters ($\bar{\beta}$) also appear to be consistent with each other. For example, living 1 km further away from the subway station is as bad as living in a neighborhood where the average building is 7.3 years older (both are worth 0.3 utils less). The average distance to the subway station is 0.8km, the average age of a building in a neighborhood is 19 years. The coefficient on the number of rooms is negative but insignificant. The coefficient on price is the right sign but insignificant. The heterogeneity terms are insignificant.

Moving on to ethnic preferences, the estimates show that all groups want to live with at least some members of their own ethnic group. The sign on the linear term is always positive for all ethnic groups. This is consistent with the findings in the literature on residential segregation. The sign on the quadratic term, however, is negative and significant for Chinese and Indians and negative and insignificant for Malays. Table 9 calculates marginal rates of substitution (MRS) between percent own ethnic group, and other attributes using the marginal utility estimates from Table 8. I quantify ethnic preferences in terms of the average distance to the subway station (in kilometers) instead of dollars because the coefficient on price is insignificant. Because of the quadratic term on ethnic proportions, the MRS depends on ethnic proportions in. I report MRS for ethnic proportions at the 1st percentile, the mean and the 99th percentile in the neighborhood data.

All groups have inverted U-shaped preferences, so that once a neighborhood has enough own-ethnic-neighbors, the average household prefers a new neighbor from another ethnic group. Table 9 shows that a Malay living in a neighborhood with 13% Malays (the mean) is willing to move to a neighborhood where the average distance to the subway station is 0.33km (1km = 0.62 miles) farther. Indians have very steep indifference curves. An Indian living in a neighborhood with 2% Indians (the 1st percentile) is willing to move 1.43km farther from the subway station to live in a neighborhood with 1% more Indians. By contrast, an Indian living in a neighborhood with 21% Indians (the 99th percentile) needs to move 1.24km closer.

Malays have the highest turning points relative to the mean Malay proportion. The turning point for Malays (28%) is 2.2 standard deviations above the average proportion of Malays (13%). The turning point for Indians (12%) is 1.4 standard deviations above the mean (8%) and the turning point for the Chinese (52%) is 3.9 standard deviations

is the average of the Chinese, Malay and Indian marginal utility for school quality.

below the mean (79%).[43]

The quadratic functional form allows us to test the location of the turning points where the sign of the marginal utilities change. Using the Delta Method, I tested whether the first order conditions with respect to ethnic proportions are significantly different from zero at these turning points. The t-statistics were less than 0.5 for all three groups, indicating that the inverted-U shape is significant but the location of the turning point is not precisely estimated. I am able to reject with 95% confidence that the turning point is greater than 59% for the Chinese and greater than 17% for Indians.[44] For the Malays, even at 98% Malays, I cannot reject that the first order condition is significantly different from 0 (the t-statistic is 1.33). This difference in the shape of ethnic preferences between Malays and non-Malays is consistent with findings from qualitative surveys in Singapore. In a survey of 191 movers, Sin (2002) finds that "Malays are different from the other ethnic groups in terms of their tendency to move (to locations) where they are overrepresented" (p. 1360).

Finding tastes for diversity and differences in the shapes of ethnic preferences is consistent with previous research using data on racial attitudes from the General Social Survey (Aldrich et al., 2005). In a survey of ethnic relations in Singapore, Lai (1995) found that households did express a preference for "multi-ethnic living". But most estimates of ethnic preferences have not demonstrated such non-monotinicity in ethnic preferences because they have focused on linear models.

These findings also provide evidence against ethnic discrimination, a common confounder in the literature on residential segregation. If non-Malays discriminated against Malays and forced them into Malay enclaves, then Malay enclaves could still arise even if Malays did not have preferences for other Malay neighbors. Therefore, ethnic discrimination would upward bias the estimates on Malay preferences. However, ethnic discrimination is not consistent with the inverted-U shaped preferences for Chinese and Indians. If the Chinese and Indians discriminated against Malays, then, the marginal utilities for own-ethnic-neighbors would be unlikely to be negative.

Another common confounder is income segregation. In Singapore, income varies across ethnic groups with the Chinese having the highest median income and the Malays having the lowest median income. Even if there was no taste for own-ethnic-neighbors, neighborhoods that are segregated by ethnic groups could still arise due to income seg-

[43] Although 52% represents the Chinese proportion that yields the maximum utility for the average Chinese, all neighborhoods do not converge towards 52% because the Chinese make up 77% of the population and there is variation in other neighborhood attributes that are desired by the Chinese.

[44] At proportions above 59% Chinese and 17% Indians respectively, the first order conditions are statistically significantly different from 0.

regation and the correlation between income and ethnicities. Not controlling for income could lead to an upward bias of ethnic preferences. However, the results suggest that income segregation cannot be the only mechanism. In a model with income segregation only, everyone would prefer to live with the Chinese. In Indian- and Malay-constrained neighborhoods, the Malay and Indian prices should not be higher than the Chinese prices because all groups would want to live with the Chinese instead of their own ethnic groups (see Table 5).

The findings for the Chinese are striking because the turning point (52%) is far below the Chinese neighborhood limit (84% Chinese) and yet the Chinese neighborhood quotas are binding for 9% of the neighborhoods in an average month.[45] This suggests that the Chinese neighborhood quotas are likely to be binding because the Chinese have strong preferences for amenities in those neighborhoods. This is also consistent with stylized findings borne out in the data. First, the predicted Chinese taste for neighborhood amenities ($\hat{\xi}^C$) is positively correlated with the Chinese mover shares (0.6). In addition, in Section 4.2, we saw that price differences for Chinese-constrained units suggested that the Chinese price was greater than the non-Chinese price. These price differences can either arise due to higher WTP for Chinese proportions or stronger Chinese taste for unobserved amenities. But, the inverted U-shaped Chinese preferences suggest that the average Chinese buyer prefers a lower Chinese proportion on the margin (see Table 9). This suggest that the price differences are due to an abundance of Chinese amenities in these neighborhoods.

These findings have implications for policy. In an interview in January 1989 (just before the policy was implemented), the Minister of National Development acknowledged the possibility of using amenities to attract particular ethnic groups. But, the ethnic quotas were preferred because of the concern that using such measures to achieve desegregation was deemed to be "risky" and "slow" (The Straits Times, 1989). The preference estimates in this paper are important inputs for cost benefit analysis that could inform such policy considerations. One drawback of using amenities to encourage desegregation is fixed costs (it is costly to build a Chinese temple in a neighborhood with very few Chinese). However, this is less of an issue because the Chinese are the majority in Singapore.

[45]For Indians, the turning point (12%) is 2% above the Indian neighborhood limit (10%).

7.1 First Best Simulations

In this section, I use the preference estimates in Table 8 to find the first best allocation of ethnic groups.[46] The social planner's problem is to find the allocation of ethnic groups to neighborhoods that will maximize a social welfare function. In a decentralized equilibrium, the first best allocation may not be achieved because of externalities. An individual chooses the neighborhood that maximizes his own utility, without internalizing the effect of his choice on the ethnic proportions in the neighborhood.

To find the first best allocation, I first assume a utilitarian social welfare function. Then, I make use of the preference estimates from Table 8 and also the average group G household's taste for unobserved neighborhood amenities ($\hat{\xi}^G$). While the ethnic proportions for neighborhood j change, I assume that the average group G household's taste for unobserved amenities in that neighborhood stay fixed throughout the welfare simulation. Therefore, the exercise is to search for the allocation of ethnic groups that would maximize a utilitarian social welfare function, holding fixed the average household's taste for amenities at current levels. This is the benchmark I use to compare against the allocation of ethnic groups in my sample. See Appendix 2 for the simulation details.

Figure 5 plots the density of percent Chinese, percent Malay and percent Indian now (dashed line) and under first best (solid line). The first best has more neighborhoods with low Chinese and Indian proportions but there are also some neighborhoods that are above the neighborhood quota limits (given the distribution of taste for unobserved amenities). I find that 32%, 35% and 31% of the neighborhoods in my sample have Chinese, Malay and Indian proportions that are within one standard deviation of the first best allocation. The standard deviations are 7% for Chinese and Malay proportions and 3% for Indian proportions.

One would like to know whether the equilibrium after the quota was implemented was closer to the first best than the equilibrium allocation before the quota. Unfortunately, I do not have data on pre-quota ethnic proportions. Lum and Tien (2003) report the ethnic proportions for three towns in Singapore, in 1988 (before the policy) and 1998 (after the policy).[47] Table 10 compares these ethnic proportions to the first best allocation.

Before the policy, Redhill was known as a Chinese town, Bedok was a Malay town and Yishun was an Indian town. In 1988, the Malay and Indian proportions in Bedok

[46] In theory, we could use preference estimates to estimate compensating and equivalent variation for the policy. However, this exercise is outside the scope of this paper because models with social interactions inherently have multiple equilibria and there is no consensus on how to select the counterfactual equilibrium. See Bajari et al. (2007) for an example of how to estimate structural models with multiple equilibria.

[47] Towns are bigger than neighborhoods. An average town has 22,000 households.

and Yishun were almost 3.5 times the first best levels. Ten years after the introduction of the quotas, they were within two and four percentage points of the first best Malay and Indian proportions. The magnitude of this "improvement" towards the first best allocation is likely to be an upper bound because these towns were very segregated to begin with. The ethnic proportions of these neighborhoods would likely have reverted towards the mean (and hence closer towards the first best), even without the quotas.

For Redhill, the Chinese proportion did not change much. This could be because of fixed amenities in Redhill that are preferred by the Chinese discourage Chinese from moving out (Redhill has a high density of Chinese temples and it is also close to Chinatown). Also, the Chinese are such a majority (77%) that it is hard to lower Chinese neighborhood proportions by much despite the Chinese having inverted U-shaped preferences.

There are four caveats to the discussion above. First, the taste estimation assumes that all ethnic groups share a common taste for attributes such as school quality. Second, this welfare exercise assumes that other attributes, such as school quality, do not change in response to the change in ethnic proportions. If Chinese proportions are positively correlated with school quality (perhaps due to the higher income and higher education of Chinese parents), creating more diverse neighborhoods by lowering Chinese proportions improves welfare because the Chinese like diversity but this could lower the school quality in those neighborhoods. Hence, the welfare effects would be ambiguous if neighborhood attributes and amenities were allowed to evolve as ethnic proportions changed. Third, it is possible for multiple allocations to maximize social welfare. This would depend on how differentiated the neighborhoods are. In the extreme, if two neighborhoods, A and B, are identical, a welfare-maximizing allocation with 70% Chinese in neighborhood A and 30% Chinese in neighborhood B implies that an allocation with 70% Chinese in neighborhood B and 30% Chinese in neighborhood A would also be welfare-maximizing. It is less likely in this case because most of the neighborhood attributes are continuous, making it unlikely that two neighborhoods are identical (for example, the $\hat{\xi}_j$'s are unique for each neighborhood). Fourth, the welfare simulation assumes that preferences are invariant across time and markets. As ethnic proportions evolve, ethnic preferences could also change. If the quotas caused preferences to change, the first best would also change.[48]

[48] If preferences were more skewed towards segregation pre-policy, it is possible that the first best allocation before the policy would also have more segregated neighborhoods.

8 Conclusion and Future Research

This paper estimates preferences for own-ethnic-group neighbors and uses the preference estimates to benchmark the welfare consequences of the ethnic quota policy in Singapore relative to the first best. To my knowledge, this is the first set of welfare results on desegregation policies even though these policies affect the location choices of many households around the world.

This paper develops and estimates a discrete choice model of residential location choices where the taste for unobserved neighborhood amenities is allowed to vary by ethnic groups, an improvement over existing location choice models. I first document evidence of price dispersion due to the quota policy and use the model to inform how this policy variation can be used to identify ethnic preferences. I operationalize the reduced form identification approach in the structural model using the step function of the policy rule. Importantly, I show that the step function of the policy non-parametrically identifies ethnic-specific prices, an important source of variation in a location choice model with ethnic-specific taste for amenities.

I find that all groups have strong preferences to live with at least some other members of their ethnic group. Preferences are inverted U-shaped so that after a neighborhood has enough own-ethnic-neighbors, households would rather add a new neighbor from other groups. To my knowledge, this represents the first estimate of non-monotonic ethnic preferences, made possible due to the rich variation from the phonebook data. Welfare simulations show that a Malay and an Indian town became desegregated after the quota was imposed, with the post-quota ethnic proportions within four percentage points of the first best allocation. The lack of change of Chinese proportions for the Chinese town could be due to Chinese taste for fixed amenities.

In ongoing work, I explore the use of hedonic versus discrete choice models to estimate preferences (?). Future work will also include the use of ethnic preference estimates to simulate counterfactuals. The challenge in performing such welfare calculations is that sorting models with social interactions typically have multiple equilibria. These findings will be important complements to the first best simulations in this paper.

References

Aldrich, E., Arcidiacono, P., and Vigdor, J. (2005). Do People Value Racial Diversity? Evidence from Nielsen Ratings. Topics in Economic Analysis & Policy, 5(1):1396–1396.

Angrist, J. and Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. Quarterly Journal of Economics, 114(2):533–575.

Bajari, P., Hong, H., and Ryan, S. (2007). Identification and Estimation of Discrete Games of Complete Information. Working Paper.

Bajari, P. and Kahn, M. E. (2005). Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities. Journal of Business and Economic Statistics, 23(1):20–33.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. Journal of Political Economy, 115(4):588–638.

Bayer, P., McMillan, R., and Rueben, K. (2004). An Equilibrium Model of Sorting in an Urban Housing Market. NBER Working Paper No. 10865.

Bayer, P. and Timmins, C. (2005). On the Equilibrium Properties of Locational Sorting Models. Journal of Urban Economics, 57(3):462–477.

Bayer, P. and Timmins, C. (2007). Estimating Equilibrium Models of Sorting Across Locations. The Economic Journal, 117(518):353–374.

Benabou, R. (1993). Workings of a City: Location, Education, and Production. The Quarterly Journal of Economics, 108(3):619–652.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. The Rand Journal of Economics, 25(2):242–262.

Berry, S. T., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. Econometrica, 63(4):841–890.

Berry, S. T., Levinsohn, J., and Pakes, A. (2004). Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. Journal of Political Economy, 112(1):68–105.

Brock, W. A. and Durlauf, S. N. (2001). Discrete Choice with Social Interactions. *Review of Economic Studies*, 68(2):235–260.

Brock, W. A. and Durlauf, S. N. (2002). A Multinomial-Choice Model of Neighborhood Effects. *American Economic Review*, 92(2):298–303.

Card, D., Lee, D., and Pei, Z. (2009). Quasi-Experimental Identification and Estimation in the Regression Kink Design. Working Paper No. 553, Princeton University Industrial Relations Section.

Card, D., Mas, A., and Rothstein, J. (2008). Tipping and the Dynamics of Segregation. *Quarterly Journal of Economics*, 123(1):177–218.

Crawfurd, J. (1828). Journal of an Embassy from the Governor-General of India to the Courts of Siam and Cochin-China : Exhibiting a View of the Actual State of those Kingdoms. Cornell University, South Asia Visions. http://dlxs.library.cornell.edu/s/sea/index.php.

Cutler, D. M., Glaeser, E. L., and Vigdor, J. L. (1999). The Rise and Decline of the American Ghetto. *Journal of Political Economy*, 107(3):455–506.

Damm, A. P. (2009b). Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence. *Journal of Labor Economics*, 27(2):281–314.

Edin, P.-A., Fredriksson, P., and Åslund, O. (2003). Ethnic Enclaves and the Economic Success of Immigrants-Evidence from a Natural Experiment. *Quarterly Journal of Economics*, 118(1):329–357.

Epple, D. and Sieg, H. (1997). Estimating Equilibrium Models of Local Jurisdictions. *Journal of Political Economy*, 107(4):645–681.

Gabriel, S. and Rosenthal, S. (1989). Household Location and Race: Estimates of a Multinomial Logit Model. *The Review of Economics and Statistics*, 71(2):240–249.

Guryan, J. (2003). Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts. NBER Working Paper No. 8269.

Hahn, J., Todd, P., and van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1):201–209.

Hårsman, B. and Quigley, J. (1995). The Spatial Segregation of Ethnic and Demographic Groups: Comparative Evidence from Stockholm and San Francisco. *Journal of Urban Economics*, 37(1):1–16.

Housing Development Board (1998). *HDB Profile of Residents in HDB Flats*. Singapore: HDB.

Housing Development Board (2000). *Profile of Residents Living in HDB Flats*. Singapore: HDB.

Housing Development Board (2006). *Annual Report*. Singapore: HDB.

Imbens, G. and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Economics*, 142(2):615–635.

Lai, A. E. (1995). *Meanings of Multiethnicity: A Case Study of Ethnicity and Ethnic Relations in Singapore*. Oxford University Press.

Lee, D. (2008). Randomized Experiments from Non-Random Selection in U.S. House Elections. *Journal of Econometrics*, 142(2):675–697.

Lum, S. K. and Tien, T. M. (2003). Integrating Differentials: A Study on the HDB Ethnic Integration Policy. *Crestimes, Singapore Center for Real Estate (CRES)*, 3(1):5.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3):531–542.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In Zarembka, P., editor, *Frontiers in Econometrics*. Academy Press, New York.

McFadden, D. (1978). Modelling the Choice of Residential Location. In Karlqvist, A., editor, *Spatial Interaction Theory and Planning Models*. North-Holland, New York.

McFadden, D. (1989). A Method of Simulated moments for Estimation of Discrete Response Models without Numerical Integration. *Econometrica*, 57(5):995–1026.

Moffitt, R. (2001). Policy Interventions, Low Level Equilibria, and Social Interactions. In Durlauf, S. N. and Young, H. P., editors, *Social Dynamics*. Brookings Institution Press, Washington, D.C.

Nechyba, T. J. and Strauss, R. P. (1998). Community Choice and Local Public Services: A Discrete Choice Approach. *Regional Science and Urban Economics*, 28(1):51–73.

Nevo, A. (2001). Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, 69(2):307–342.

Newey, W. (1984). A Method of Moments Interpretation of Sequential Estimators. *Economic Letters*, 14:201–206.

Nielsen, H., Sørensen, T., and Taber, C. (2009). Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform. NBER Working Paper No. 14535.

Pakes, A. and Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, 57(5):1027–1057.

Parliamentary Debates (1989). Official Report, 16 February. Singapore.

Petrin, A. (2002). Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy*, 110(4):705–729.

Polikoff, A. (1986). *Sustainable Intergration or Inevitable Resegregation*, pages 41–71. University of North Carolina Press.

Quigley, J. M. (1985). Consumer Choice of Dwelling, Neighborhood, and Public Services. *Regional Science and Urban Economics*, 15(1):41–63.

Rosenbaum, J. (1992). Black Pioneers-Do Their Moves to the Suburbs Increase Economic Opportunity for Mothers and Children? *Housing Policy Debate*, 2(4):1179–1213.

Rosenbaum, J. (1995). Changing the Geography of Opportunity by Expanding Residential choice: Lessons from the Gautreaux Program. *Housing Policy Debate*, 6(1):231–269.

Schelling, T. C. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1(3):143–186. (Abbreviated version appeared as "Models of Segregation" in *American Economic Review* 59(2), May 1969).

Simonsen, M., Skipper, L., and Skipper, N. (2009). Price Sensitivity of Demand for Prescription Drugs: Exploiting a Regression Kink Design. Working Papers 10-1, University of Aarhus, Aarhus School of Business, Department of Economics.

Sin, C. H. (2002). The Quest for a Balanced Ethnic Mix: Singapore's Ethnic Quota Policy Examined. *Urban Studies*, 39:1347–1374.

Singapore Department of Statistics (2000). Singapore 2000 Census. Key Indicators of the Resident Population.

Straits Times (7 January 1989).

The Straits Times (January 31, 1989). Singaporeans Accept Need for Racially Cohesive Society. *The Straits Times*, page 20.

Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64(5):416–424.

Wong, M. (2012). Estimating the Impact of the Ethnic Housing Quotas in Singapore. University of Pennsylvania, Real Estate Department Working Paper.

TABLE 1: NEIGHBORHOOD AND BLOCK LEVEL ETHNIC QUOTAS

Ethinicity	Neighborhood Quotas	Block Quotas	National Population (2000)
Chinese	84%	87%	77%
Malay	22%	25%	14%
Indian	10%	13%	8%

SOURCE. 2000 Census (Singstat), Lum and Tan (2003)

TABLE 2: RELATIONSHIP BETWEEN QUOTAS AND THE ETHNICITY OF BUYERS AND SELLERS

Binding Quota	Buyer Ethnicity	Seller Ethnicity	Status
Chinese	Chinese	Chinese	Allowed
	Non-Chinese	Non-Chinese	Allowed
	Non-Chinese	Chinese	Allowed
	Chinese	Non-Chinese	Not Allowed
Malay	Malay	Malay	Allowed
	Non-Malay	Non-Malay	Allowed
	Non-Malay	Malay	Allowed
	Malay	Non-Malay	Not Allowed
Indian	Indian	Indian	Allowed
	Non-Indian	Non-Indian	Allowed
	Non-Indian	Indian	Allowed
	Indian	Non-Indian	Not Allowed

TABLE 3: SUMMARY STATISTICS*fi*

Variables	N	Mean	Std. Dev.	Description
Chinese Share	170	0.12%	0.14%	Number of Chinese movers who chose a neighborhood/Number of Chinese residents in a market
Malay Share	170	0.14%	0.15%	Number of Malay movers who chose a neighborhood/Number of Malay residents in a market
Indian Share	170	0.38%	0.38%	Number of Indian movers who chose a neighborhood/Number of Indian residents in a market
Price	170	238,901	50,755	Average transaction price in a neighborhood (Singapore dollars)
Chinese Neighborhood Quota	170	0.09	0.27	Percent of months Chinese neighborhood quota binds
Malay Neighborhood Quota	170	0.05	0.17	Percent of months Malay neighborhood quota binds
Indian Neighborhood Quota	170	0.17	0.32	Percent of months Indian neighborhood quota binds
Chinese Block Quota	170	0.10	0.17	Percent of months and blocks Chinese block quota binds
Malay Block Quota	170	0.05	0.11	Percent of months and blocks Malay block quota binds
Indian Block Quota	170	0.08	0.13	Percent of months and blocks Indian block quota binds
Percent Chinese	170	79%	7%	Percent of Chinese in a neighborhood
Percent Malay	170	13%	7%	Percent of Malays in a neighborhood
Percent Indian	170	8%	3%	Percent of Indians in a neighborhood
School Quality	170	3.15	4.21	Total number of awards received by schools in a neighborhood
Subway	170	0.80	0.55	Distance to the closest subway station
Rooms	170	4.12	0.63	Number of rooms in a unit in the neighborhood
Age	170	19.22	7.11	Average age of HDB blocks in the neighborhood

NOTE. School quality is measured as the total number of awards given to primary, secondary schools and tertiary institutions by the Singapore Ministry of Education.

TABLE 4: EFFECTS ON OBSERVED PRICES CLOSE TO THE QUOTA LIMITS

	DEPENDENT VARIABLES					
	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price
	(1)	(2)	(3)	(4)	(5)	(6)
Chinese Quota	0.0819***	0.0507***				
	(0.0165)	(0.0060)				
Malay Quota			-0.0387***	-0.0267***		
			(0.0116)	(0.0042)		
Indian Quota					-0.0283***	-0.0330***
					(0.0108)	(0.0039)
Polynomial	Quartic	Quartic	Quartic	Quartic	Quartic	Quartic
Controls	N	Y	N	Y	N	Y
Month	N	Y	N	Y	N	Y
Town	N	Y	N	Y	N	Y
Obs	19314	19314	14862	14862	32114	32114
R-squared	0.0102	0.7977	0.0041	0.7472	0.0092	0.7751

NOTE. The regression equation is $\ln P_{ibjtk} =$ ***fi*** + ***fl***$QC_{bjt-1,k} + f(percentC_{bjk}) +$ ⁴$_{ibjtk}$ where $\ln P_{ibjtk}$ is the log of the price of unit i in block b, neighborhood j, month t and town k; $QC_{bjt-1,k}$ is a dummy that is 1 when the Chinese (C) quotas are binding, $f(percentC_{bjk})$ is a flexible polynomial, centered at the block quota limit. The controls are other observable characteristics of the block (age of block b and its squared, share of units of block b that is of each of the 7 HDB types (type 1 is omitted)), month and town fixed effects. I repeat the exercise for the Malay quotas (columns 3-4) and Indian quotas (columns 5-6); Standard errors are in parentheses and are clustered at the block level.

* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$.

TABLE 5: EFFECTS ON OBSERVED PRICES, CONTROLLING FOR ETHNIC BUYER WEIGHTS

	DEPENDENT VARIABLES		
	LnPrice	LnPrice	LnPrice
	(1)	(2)	(3)
Chinese Neighborhood Quota Status	0.1849***		
	(0.0572)		
Chinese Neighborhood Quota Status x Malay Weights	-1.5308**		
	(0.5909)		
Chinese Neighborhood Quota Status x Indian Weights	-0.12		
	(0.3591)		
Malay Neighborhood Quota Status		0.26	
		(0.2093)	
Malay Neighborhood Quota Status x Chinese Weights		-0.6624**	
		(0.2752)	
Malay Neighborhood Quota Status x Indian Weights		0.11	
		(0.4098)	
Indian Neighborhood Quota Status			0.1373*
			(0.0767)
Indian Neighborhood Quota Status x Malay Weights			-0.8740***
			(0.2300)
Indian Neighborhood Quota Status x Chinese Weights			-0.03
			(0.1260)
Controls	Y	Y	Y
Month	Y	Y	Y
Observations	2043	1393	2741
R-squared	0.3931	0.5266	0.3768

NOTE. □ Each column is a regression restricted to 10% above and below the Chinese, Malay and Indian neighborhood quota limits. The regression equation is $\ln P_{jt} = fl\, nQC_{jt-1} + f_1\, nQC_{jt-1} * wM_j + f_2\, nQC_{jt-1} * wI_j + f(percentC_j) + X_j b + t_t + 4_{jt}$ where $\ln P_{jt}$ is the log of the average price of neighborhood j in month t; nQC_{jt-1} is a dummy that is 1 when the neighborhood Chinese quota is binding in month t-1, wM_j and wI_j are the share of movers in neighborhood j who are Malay and Indian, $f(percentC_j)$ is a flexible polynomial, centered at the neighborhood quota limit and t_t is month fixed effects. Controls (X_j) include a constant, the average age of HDB blocks in a neighborhood, its squared, average number of rooms, school quality and distance to the closest subway station. Standard errors are in parentheses, clustered at the neighborhood level.

* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$.

TABLE 6: SUMMARY OF IDENTIFICATION STRATEGY

Identification Strategy	Assumptions and Thought Experiments
1. Omitted variables that are ethnic-specific ($□_j^G$ instead of $□_j$)	
□ Model neighborhood-by-ethnic-group fixed effects instead of neighborhood fixed effects ($□_j^G$ instead of $□_j$) using the price discrimination mechanism (the quotas generate price variation across ethnic groups by preventing arbitrage).	□ Consider 2 observationally identical neighborhoods, A and B. Prices depend on observed and unobserved neighborhood attributes. If the Chinese buyers paid a higher price than the non-Chinese buyers for a unit in neighborhood A compared to neighborhood B, since price is positively correlated with quality, this observed variation in ethnic-specific prices has information on unobserved ethnic-specific neighborhood quality.
2. Prices and ethnic proportions are correlated with unobserved ethnic-specific neighborhood quality	
2a. Instruments for ethnic proportions:	
□ Characteristics of nearby neighborhoods (1-3km, 3-5km and 5-7km rings).	□ Consider 2 observationally identical neighborhoods, A and (little) a. Neighborhood A has large units relative to the surrounding neighborhoods; Neighborhood (little) a has small units relative to the surrounding neighborhoods. Malays prefer bigger units because they have bigger families. So, they will sort differentially into neighborhood A. Therefore, unit size of surrounding neighborhoods is correlated with ethnic proportions.
□ Historical ethnic settlements.	□ Figure 4 shows that the location of existing Malay neighborhoods is correlated with the location of early-19th century Malay settlements. This assumes that Malay settlements were exogenously assigned to the east of the Singapore River in the early 19th century.
2b. Instruments for ethnic-specific prices, P_j^G:	
□ Characteristics of nearby neighborhoods (1-3km, 3-5km, 5-7km rings).	□ First, use characteristics of nearby neighborhoods to predict demand using only exogenous variation in the data. Then, use the model to solve for a vector of prices that clears the market (Bayer, Ferreira and McMillan, 2007).
□ Estimated quota dummy.	□ Quotas are correlated with ethnic prices through price discrimination. Use the instruments in 2a to predict ethnic proportions. Let $\hat{Q}$ equal 1 if the predicted ethnic proportion is above the quota limit. This assumes that the price effect of all other neighborhood attributes is smooth at the quota limits and only the effect of the quota is discontinuous at the limits.

TABLE 7: REGRESSION OF ACTUAL QUOTA PROBABILITIES ON ESTIMATED QUOTA PROBABILITIES

	DEPENDENT VARIABLES					
	Actual Chinese Quota Probability	Actual Malay Quota Probability	Actual Indian Quota Probability+	Actual Chinese Quota Probability	Actual Malay Quota Probability	Actual Indian Quota Probability+
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Placebo = Actual Limits					
Estimated Chinese Quota Probability	0.4891***			0.1341*		
	(0.0564)			(0.0771)		
Estimated Malay Quota Probability		0.6685***			0.3929***	
		(0.0952)			(0.0965)	
Estimated Indian Quota Probability			0.3473***			-0.13
			(0.0891)			(0.1073)
Controls	N	N	N	Y	Y	Y
Fstat	75.09	49.29	15.20	8.00	5.47	2.41
Obs	170	170	170	170	170	170
R-squared	0.31	0.23	0.08	0.49	0.39	0.22
	Panel B: Placebo = Actual Limits + 3%					
Estimated Chinese Quota Probability	0.3817***			-0.11		
	(0.0994)			(0.0863)		
Estimated Malay Quota Probability		0.15			-0.4090*	
		(0.2417)			(0.2248)	
Estimated Indian Quota Probability			dropped			Dropped
Controls	N	N	N	Y	Y	Y
Fstat	14.74	0.39	.	7.84	4.37	2.45
Obs	170	170	170	170	170	170
R-squared	0.08	0.00	.	0.48	0.34	0.22
	Panel C: Placebo = Actual Limits - 3%					
Estimated Chinese Quota Probability	0.3867***			0.089		
	(0.0356)			(0.0746)		
Estimated Malay Quota Probability		0.4410***			0.2773***	
		(0.0458)			(0.0583)	
Estimated Indian Quota Probability			0.2856***			0.11
			(0.0281)			(0.0869)
Controls	N	N	N	Y	Y	Y
Fstat	117.99	92.54	103.31	7.83	5.97	2.41
Obs	170	170	170	170	170	170
R-squared	0.41	0.35	0.38	0.48	0.42	0.22

NOTE. Panel A uses the actual limits to estimate quota probabilities. Panels B and C use placebo limits that are 3% above and below the actual limits. Each panel has 6 columns. The first three do not control for instruments used to estimate the quota probabilities; the following three columns do. The instruments are the sum of school awards, the distance to the closest subway station, the average age of buildings, the average number of rooms for nearby neighborhoods within 1-3km, 3-5km and within 5-7km, as well as a dummy for being in the east of the early Malay settlements; Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[+] There is no F-stat and R-squared for Panel B because the variable, Estimated Indian Quota Probability, was zero for all neighborhoods.

TABLE 8: RANDOM COEFFICIENTS LOGIT WITH QUADRATIC ETHNIC PREFERENCES

Variables	Units	Common Taste Parameters (1)	Chinese Taste Parameters (2)	Malay Taste Parameters (3)	Indian Taste Parameters (4)
Means:					
Constant		-6.81***			
		(2.29)			
School Quality	.1 awards	1.18***			
		(0.12)			
Distance to Subway	1 km	-0.36***			
		(0.11)			
Average No. of Rooms	0.1 rooms	-4.16			
		(4.58)			
Average Age of Buildings	0.01 years	-4.94***			
		(1.01)			
Percent Own Ethninc Group			10.26***	22.27***	6.17***
			(2.04)	(6.95)	(0.94)
(Percent Own Ethnic Group)^2			-9.95***	-3.99	-2.53***
			(2.16)	(2.50)	(0.60)
Price	S$million	-0.70			
		(14.16)			
Heterogeneity:					
Constant		-0.95			
		(2.59)			
Average No. of Rooms	0.1 rooms	-0.57			
		(10.52)			
Price	S$million	0.08			
		(7.35)			

NOTE. Variables are scaled so that the mean is between 0.1 and 1. The units are in the table. For example, the coefficient on School Quality in the quadratic model implies that an increase in 10 awards is associated with an increase of 1.18 utils. For the ethnic proportions, percent Chinese, percent Chinese^2 and percent Malay are not scaled; percent Malay^2 and percent Indian are multiplied by 10; percent Indian^2 is multiplied by 100. The instruments are the sum of school awards, the distance to the closest subway station, the average age of buildings, the average number of rooms for nearby neighborhoods within 1-3km, 3-5km and within 5-7km, a dummy for being in the east of the early Malay settlements, the estimated quota dummies and the vector of Chinese, Malay and Indian price vectors summarized by the demand model. Standard errors are in parentheses.
* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$.

TABLE 9: MRS EVALUATED AT VARIOUS ETHNIC PROPORTIONS IN THE SAMPLE

	ETHNICITY		
	Chinese	Malays	Indians
Relevant statistics for ethnic proportions:			
Mean of Percent Own Ethnic Group	79%	13%	8%
1st percentile of Percent Own Ethnic Group	63%	1%	2%
99th percentile of Percent Own Ethnic Group	98%	29%	21%
Standard Deviation of Percent Own Ethnic Group	7%	7%	3%
MRS relative to distance to subway (km), per 1% increase in ethnic proportion:			
MRS at mean of Percent Own Ethnic Group	-0.15	0.33	0.59
MRS at 1st percentile of Percent Own Ethnic Group	-0.06	0.60	1.43
MRS at 99th percentile of Percent Own Ethnic Group	-0.26	-0.02	-1.24

NOTE. This table shows calculations of the MRS for ethnic preferences evaluated at different ethnic proportions. The top panel reports the mean, 1st percentile, 99th percentile and standard deviation for percent Chinese, percent Malay and percent Indian. The bottom panels report the MRS's. The MRS's are reported in terms of the number of kilometers to the closest subway station per 1 percentage point increase in the percent of own ethnic group neighbors (1km=0.62 miles). The MRS's are calculated as the marginal utility for a 1 percentage point increase in own ethnic group neighbors divided by the negative of the marginal utility for distance to the closest subway station (2nd panel). A positive MRS reflects preferences for a new neighbor from the own ethnic group. A negative MRS reflects preferences for a new neighbor from other ethnic groups. Since ethnic preferences are quadratic, the MRS changes with ethnic proportions. I calculate the MRS at the mean, the 1st percentile and the 99th percentile. The average distance to the closest subway station is 0.8km.

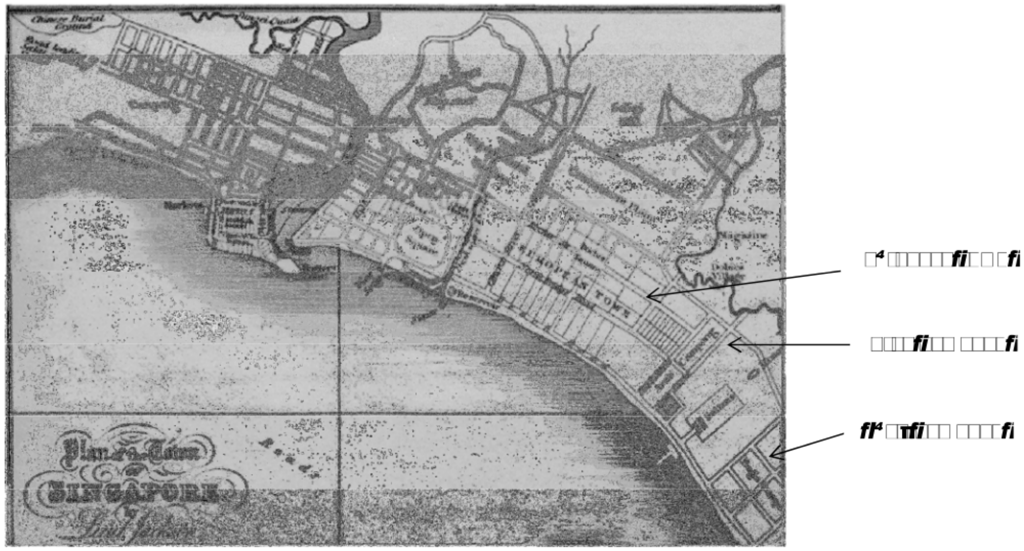
TABLE 10: ETHNIC PROPORTIONS OF THREE TOWNS, BEFORE, AFTER THE QUOTA AND FIRST BEST*fi*

	Before (1988)	After (1998)	First Best
Percent Chinese in Redhill	87%	84%	74%
Percent Malay in Bedok	59%	19%	17%
Percent Indian in Yishun	24%	11%	7%

SOURCE. Straits Times 7 January 1989, HDB profile of residents in HDB flats, 1998.
NOTE. This table reports ethnic proportions for a traditionally Chinese town (Redhill), Malay town (Bedok) and Indian town (Yishun). The three columns report ethnic proportions pre- and post-quota as well as the first best obtained from simulations. A town is a cluster of neighborhoods, with an average of 22,000 households.

Source: Google Maps

FIG. 1. Map of HDB blocks and HDB neighborhoods. Each number in the map corresponds to an HDB block. There are 4 HDB neighborhoods in this map. Neighborhood 1 comprises all HDB blocks between 100 and 199, neighborhood 2 comprises all HDB blocks between 200 and 299, neighborhoods 4 and 5 are defined in a similar manner.



SOURCE: Crawford, 1828

FIG. 2. Map of ethnic settlements in early 19th century. The Malay settlements ("Arab Campong" and "Bugis Campong") are in the south east corner of the map, just east of the European Town. The Chinese and Indian areas are to the west of the Singapore River

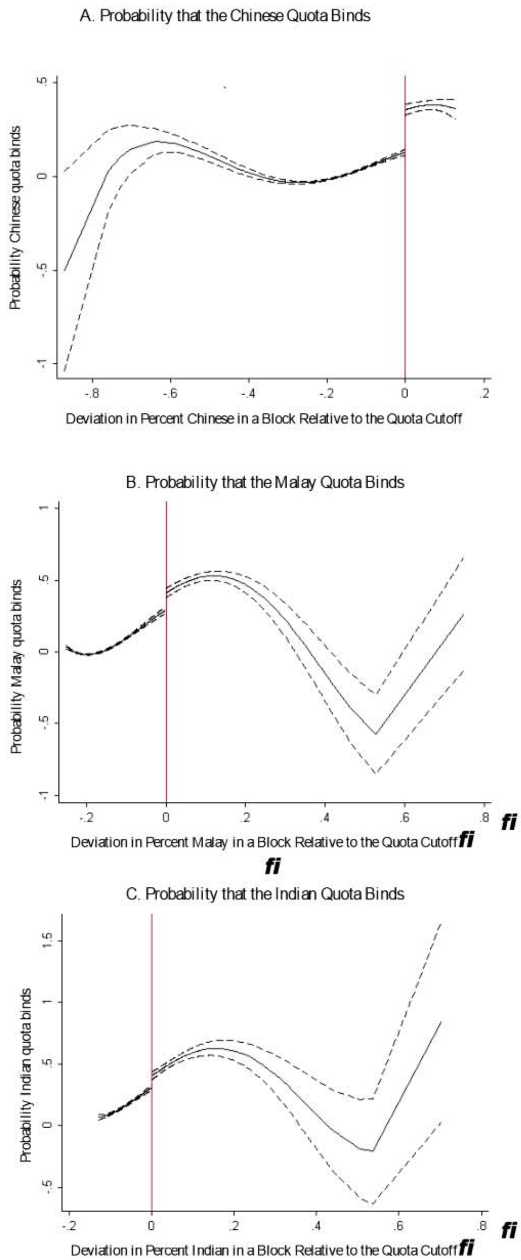


FIG. 3. Testing for discontinuity in the probability that the quota binds. Each panel in this figure is constructed by regressing Q_{bt} (a dummy for whether the quota is binding for block b in month t) on a dummy that is 1 when the ethnic proportions are above the block quota cutoff and 4th order polynomials of ethnic proportions, centered around the quota cutoffs, then plotting the predicted probabilities. Repeat the exercise for the Malay quotas and Indian quotas. The dashed lines represent 95% confidence intervals. Standard errors clustered at the block level. The coefficient estimates are 23%, 12% and 10% for Chinese, Malay and Indian quotas.