Estimating Ethnic Preferences Using Ethnic Housing Quotas in Singapore

Maisy Wong[*†]

November 19, 2010

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 589,000 names in the phonebook to ethnicities. I find that all groups like own-ethnic-group neighbors. Interestingly, the Chinese majority exhibit inverted U-shaped preferences so that once a neighborhood has enough Chinese neighbors, they would rather add a new neighbor from other groups.

[*]I wish to thank my advisors Esther Duflo, Amy Finkelstein, Panle Jia and Bill Wheaton for guidance and support. I am also deeply 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. In Sweden, the government introduced a refugee settlement program in response to the formation of ethnic enclaves in metropolitan areas. Germany, the United Kingdom and Netherlands, too, impose strict restrictions on where refugee immigrants can settle. 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.[1] 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.[2,3] I am able to model social interactions at such a local level because I assembled a dataset of ethnic proportions by hand-matching 589,000 names to ethnicities using the Singapore residential phonebook.[4] I combined this phonebook data with data I collected on housing transaction prices, neighborhood choices of movers from different ethnic

[1]There is a large literature on sorting in housing markets that began with Tiebout (1956), followed by important papers on location choices by Benabou (1993) and Epple and Sieg (1997). My location choice model uses a discrete choice framework that builds on work by McFadden (1973), McFadden (1978), Berry (1994), Berry, Levinsohn, and Pakes (1995) For examples of discrete choice models in the urban economics literature, see Quigley (1985) and Nechyba and Strauss (1998). A related empirical framework often employed in housing markets is the hedonic model (Rosen, 1974; Epple, 1987; Bartik, 1987; Ekeland, Heckman, and Nesheim, 2004; Bajari and Benkard, 2005).

[2]See Manski (1993), Brock and Durlauf (2001), Brock and Durlauf (2002), Bayer and Timmins (2005, 2007) for papers on social interactions. Moffitt (2001) investigates the use of desegregation policies as natural experiments to study social interactions.

[3]The seminal paper on residential segregation (an example of social interactions in neighborhoods) is Schelling (1971). Empirical papers that investigate the causes of residential segregation include Gabriel and Rosenthal (1989); Hårsman and Quigley (1995); Cutler, Glaeser, and Vigdor (1999); Bajari and Kahn (2005); Bayer, Ferreira, and McMillan (2007); Card, Mas, and Rothstein (2008).

[4]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. The average neighborhood in Singapore has 4000 households and an average land area of 1.5 square miles. Cutler, Glaeser, and Vigdor (1999) and Bajari and Kahn (2005) use ethnic data at the MSA and PUMA level. Bayer, Ferreira, and McMillan (2007) use restricted Census data and are able to observe ethnic proportions at the Census block group level.

groups (calculated by matching names between 2 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.[5] It was 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, Singapore 2000 Census, 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 such differences would be arbitraged away by sellers.[6] The effect of the quotas is to impose restrictions that are ethnic-based, preventing some sellers from arbitraging away price differences across the ethnic group of the buyers, thereby making it possible to observe Chinese and non-Chinese buyers paying different prices for the same location, in equilibrium. By limiting arbitrage opportunities, as does a price discrimination regime, the quotas generate equilibrium price dispersion across buyers from different ethnic groups.

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, Lee, and Pei, 2009).[7] 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

[5]There is a growing literature on ethnic desegregation policies. Rosenbaum (1992, 1995) studies the impact of the Gautreaux program on labor market and schooling outcomes. Edin, Fredriksson, and Åslund (2003) and Damm (2009b) investigate the impact of ethnic enclaves on labor market outcomes by using plausibly exogenous variation in ethnic segregation generated by refugee settlement policies in Sweden and Denmark, respectively. Boisjoly, Duncan, Kremer, Levy, and Eccles (2006) find that randomly assigning African-American roommates to white college students affects their endorsement of affirmative action policies. Kling, Liebman, and Katz (2007) investigate the impact of Moving to Opportunity, a large randomized trial involving the relocation of many public housing residents in five cities in the United States.

[6]This paper is focused on estimation of the demand side and 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.

[7]While the estimation equations are very similar to regression discontinuity (Angrist and Lavy, 1999; Hahn, Todd, and van der Klaauw, 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, Lee, and Pei (2009)'s study on the impact of previous earnings on unemployment insurance benefits.

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 neighbors, there would be no price dispersion across ethnic groups at the quota limits. Hence, the quotas would have no price discrimination 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 ethnicity of the buyer, I find suggestive evidence of price dispersion across ethnic groups in that Chinese and Malay buyers pay more (relative to non-Chinese and non-Malay buyers) for Chinese- and Malay-constrained units, respectively. I do not find this for Indian-constrained units.

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 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 unobserved *ethnic-specific* 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. Addressing this identification problem is also policy-relevant because many residential desegregation policies (including the ethnic quotas) are part of public housing programs. In addition to providing dwellings in public housing estates, they are responsible to ensure the provision of public goods that cater to specific ethnic groups. Distinguishing Chinese buyers' taste for living with other Chinese neighbors versus their taste for unobserved Chinese amenities will be useful for cost benefit analyses.

However, a location choice model with unobserved ethnic-specific 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 through the price discrimination mechanism. When the Chinese quota binds, 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 taste for Chinese amenities separately from the taste for non-Chinese amenities: Conditional on the proportion of Chinese living in a neighborhood (stock), Chinese buyers (flow) are willing to pay more than non-Chinese buyers when they have a stronger taste for amenities in that

neighborhood than non-Chinese.[8] Therefore, the model with both ethnic-specific taste for neighborhood proportions and ethnic-specific taste for neighborhood amenities is no longer under-identified.[9]

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 for ethnic proportions (including historical ethnic settlements in Singapore and attributes from nearby (but not adjacent) neighborhoods[10]) to isolate plausibly exogenous variation in ethnic proportions, and hence, variation in the probability that the quotas bind. That is, I construct a quota dummy that is 1, if the ethnic proportions estimated using the instruments are above the quota limits.[11]

The step function of the policy rule is important.[12] The identification assumption is that the estimated quota dummies (whether the quota is binding or not) are correlated with ethnic-specific prices through price discrimination, but the taste for unobserved amenities is mean-independent of the instruments. This assumption fails if, conditional on the instruments (including the predicted quota dummy), the taste for amenities is discontinuous at the quota limits. Thus, the effect of the estimated quota dummies 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. 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

[8]The underlying assumption, borne out in the data, is that the flows are small so that their effect on the proportion of Chinese living in a neighborhood is negligible.

[9]I follow standard utility specifications where taste for ethnic proportions and the taste for unobserved neighborhood amenity are modeled as additive and separable. Without additivity and separability, I will need more moments 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 restaurants, but they also derive more utility from eating in Chinese restaurants with 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. Identification without additivity and separability of the unobserved neighborhood amenity is non-trivial (Imbens and Newey, 2009).

[10]See Bayer and Timmins (2007) and Bayer, McMillan, and Rueben (2004) for an example of using surrounding neighborhood attributes to instrument for ethnic proportion. Indeed, instruments for neighborhood level ethnic proportions are hard to come by. Many papers circumvent this issue by instrumenting for ethnic proportions at the city level (Cutler, Glaeser, and Vigdor (1999) and Edin, Fredriksson, and Åslund (2003)).

[11]Actual quota dummies are not instruments because they are likely to bind when ethnic proportions or the levelr of ethnic-specific amenities are high. Therefore, I use the step function of the policy rule and other instruments to isolate plausibly exogenous variation in whether a quota binds or not.

[12]The quota dummy is a step function because it is 1 (constrained) when the ethnic proportions are above the quota limit, 0 otherwise.

groups. I find strong evidence of non-linearities and heterogeneity in ethnic preferences. In particular, the Chinese and the Indians have ethnic preferences that are inverted U-shaped but the Malays do not. This means that once a neighborhood has enough members of their own ethnic group, Indians and Chinese want new neighbors from other ethnic groups. Previous research in the United States have documented evidence of tastes for diversity using data on racial attitudes from the General Social Survey (Aldrich, Arcidiacono, and Vigdor, 2005) but most empirical estimates of ethnic preferences have not been able to demonstrate such non-monotinicity in ethnic preferences because they have focused on linear models.[13] I investigate different functional forms, including linear, quadratic and cubic specifications of ethnic preferences.

I use these estimates of ethnic preferences to perform first best simulations that could have implications that extend beyond Singapore. 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 the case 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). Ten years after the quota policy has been introduced, I find that 71% of the neighborhoods have Chinese proportions that are within 1 standard deviation (7%) of the first best allocation and 18% are within half a standard deviation. For the Malay and Indian proportions , roughly half are within 1 standard deviation and 21% and 13% respectively, are within half a standard deviation. The standard deviation for Malay and Indian proportions are 7% and 3%.

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 housing 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%).[14] The Chinese have the highest median monthly income (S$2335),

[13] See for example (Bayer, Ferreira, and McMillan, 2007; Bajari and Kahn, 2005; Cutler, Glaeser, and Vigdor, 1999).

[14] All three ethnic groups are citizens, none are immigrants.

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 with 82% of the resident population living in public housing (Housing Development Board, 2006). The units are built and managed by the Housing Development Board (HDB). There are three ways Singapore residents can live in an HDB unit. They may apply through the primary allocation system for new HDB units, they may purchase existing HDB units in the resale market or they may rent. The rental market is negligible: 98% percent of the HDB units are owner-occupied (Housing Development Board, 2006). This paper focuses on the resale market which is where the ethnic quotas apply. 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*. A group of HDB units constitute an HDB *block*. A group of HDB blocks make up a *neighborhood.* 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. The average neighborhood in Singapore has 4000 households and an average land area of 1.5 square miles. Throughout my analysis, I define a *market* as a cluster of neighborhoods.

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).[15] The policy was announced in a parliamentary debate on February 16, 1989 and was implemented starting March 1, 1989. It is a set of quota limits at the block and neighborhood level. Table 1 lists the quotas, in comparison to the 2000 national ethnic proportions. 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.

The quotas are upper limits on ethnic proportions to prevent HDB communities that are already segregated from becoming more segregated. Once a community hits the upper limit, transactions that make the community more segregated will not be allowed.[16] However, transactions involving buyers and sellers from the same ethnicity will always be allowed because this does not increase the ethnic proportion. For example, Table 1 shows that the

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

[16] 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. An inter-ethnic married couple can choose to use either ethnicities of the spouses.

Chinese neighborhood quota is set at 84%. Once the Chinese make up more than 84% of the neighborhood population, Chinese buyers can no longer buy from non-Chinese sellers because this increases the proportion of Chinese in that neighborhood. Table 2 lists the types of transactions allowed or not allowed, for each ethnic quota. The important thing to note is that once a Chinese quota binds, the Chinese buyers can no longer buy from non-Chinese sellers (similarly for Malay and Indian quotas). This ethnic-specific restriction prevents arbitrage and thus allows prices to differ across ethnic groups in equilibrium.

3 Data

I use data covering 170 neighborhoods and 7 markets, comprising all resale transactions in the public housing market in Singapore between April 2005 and March 2006. This dataset encompasses virtually all of Singapore.[17] A market is a cluster of neighborhoods, categorized according to the Straits Times Real Estate Classifieds (the leading English newspaper in Singapore). The number of neighborhoods in each market varies from 12 to 38.

Neighborhood attributes

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

To calculate ethnic proportions, I hand matched more than 589,000 names to ethnicities using differences in the structure of Chinese, Malay and Indian names.[18] For example, most Chinese names only have 2 or 3 words; Malay names are primarily Muslim names since 99% of Malays in Singapore are Muslims (Singapore Department of Statistics, Singapore 2000 Census, 2000); Indian names are matched according to popular first and last names. The match between names and ethnicity is likely to be most accurate for Chinese names because of distinct last names.[19] 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 589,000 names in the phonebook, 470,000 were matched using popular first and last names and the remaining were matched individually.

[17] 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 housing market would be a selected sample of people with weaker ethnic preferences.

[18] The 2005 phonebook was published on April 1st 2005, and includes a total of 789,048 households. I only included households living in HDB blocks. Movers have to update their contact information within a month of moving. 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.

[19] Even Chinese Muslims would tend to keep their last names when they switch to Muslim names.

Of these, 40% were eventually matched as Chinese names, 28% and 32% were matched as Malays and Indians, respectively.

I collected the remaining neighborhood attributes from online street directories, the HDB website and a non-public dataset purchased from HDB. See Appendix 1 for definitions of these variables and their sources.

Location choice data

A choice is a neighborhood. I collected choice data by matching names from the 2005 and 2006 Singapore residential phonebooks. I define *movers* as individuals whose HDB postal code in 2005 did not match with their HDB postal code in 2006.[20] A postal code uniquely identifies an HDB block and the 1st digit of the block number is used as a neighborhood identifier. There are 16,092 movers.[21] I summarize location choices using *ethnic shares*. For example, the Chinese share for neighborhood j is calculated as the percent of Chinese residents in a market who moved into neighborhood j. 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). The assumption is that the flow of movers is so small that the ethnic proportion of the stock of residents is essentially constant within a year.

Prices

I collected data on 25,182 transaction prices between March 2005 and April 2006.[22] These prices were updated every 3 months on the HDB website. I average these monthly transaction prices to the neighborhood level. Unfortunately, I do not know the ethnicity of the buyer and seller. I address this in the following section.

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.

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 1 shows the map of early 19th century Singapore according to

[20] Of the 589,000 households living in HDB blocks in 2005, 89% remained in the same HDB block from 2005 to 2006, 7% exited the HDB market and 3% were movers.

[21] To restrict the choice data to a more homogeneous group of households, I dropped the location choices of "entries" into the HDB market (ie. households who did not have HDB postal codes in 2005) because they could be entering through the lottery (there is no data available for this primary market) or they could be entering from the private (non-HDB) market.

[22] The actual number of transactions (25,182) is higher than the number of movers (16,092). Part of the difference could be attributed to entries into the HDB market.

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 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.

4 Price Effects at the Quota

A key assumption in the structural estimation is that buyers of different ethnicities will pay different prices for observationally identical units when the quotas are constrained because ethnic quotas prevent some sellers from arbitraging away price differences across ethnic groups. It is important to demonstrate price dispersion exists because I will argue later that observing Chinese buyers paying a different price than non-Chinese buyers for observationally identical units in the same neighborhood will allow me to estimate a location choice model where the taste for Chinese amenities is different from the taste for non-Chinese amenities. The focus of this section is on prices. In a separate paper, I analyze the effect of the quota on outcomes other than price (Wong, 2010a).

I present a descriptive analysis of the behavior of prices above and below the quota in the spirit of the "regression kink design" (Card, Lee, and Pei, 2009).[23] The strategy is to identify kinks in the outcome variable that coincide with kinks in the policy rule.[24]

Figure 1 shows that within 10% of to the quota limits, the probability that the quota binds in month t is discontinuously higher at the quota limit. 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 (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 previous section. The noise introduced by the measurement

[23] While the estimation equations are very similar to regression discontinuity (Angrist and Lavy, 1999; Hahn, Todd, and van der Klaauw, 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.

[24] See Guryan (2003); Nielsen, Sørensen, and Taber (2009); Simonsen, Skipper, and Skipper (2009) for examples of non-parametric identification of outcomes using kinked policy rules.

error would bias against finding discontinuities unless the measurement error is correlated with the quota dummy (whether a quota binds or not). A priori, there is no reason to expect that names that are harder to match to ethnicities would be disproportionately more likely to be associated with quota-constrained units.

I proceed in two steps: First, I test whether there is a discontinuity in *observed* prices at the quota. Second, I demonstrate the price discrimination effect of the quota by testing whether *estimated ethnic-specific* prices differ by buyer ethnicity when the quotas bind.

4.1 Is there a discontinuity in observed prices?

I estimate the following equations:

$$\ln P_{bjit} = \gamma QC_{bjit} + f\left(percentC_{bji}\right) + \varepsilon_{bjit} \quad (1)$$

$$\ln P_{bjit} = \gamma QC_{bjit} + f\left(percentC_{bji}\right) + B_{bji}\beta + \tau_t + \omega_i + \varepsilon_{bjit} \quad (2)$$

where $\ln P_{bjit}$ is the log of the price of units in block b, neighborhood j, town i and month t; QC_{bjit} is a dummy for whether the Chinese (C) quotas are binding, $f\left(percentC_{bji}\right)$ are polynomials of the percent of Chinese, centered around the quota limit (I allow separate polynomials above and below the quota limits); B represents other observable attributes of the block (age of building, number of 1-room units, 2-room units etc.); τ_t and ω_i are month and town fixed effects. I estimate these equations for units that are 10% above and below the Chinese quota limits. Similarly, I repeat the analysis for the Malay and Indian quotas. The coefficient of interest is γ, which summarizes the price effects at the quota limits.

In Table 4, columns 1-5 correspond to the regression close to the Chinese quota, columns 6-10 correspond to the Malay quota regression and columns 11-15 correspond to the Indian quota regression. For each ethnic quota, I estimate the regression controlling for polynomials of the ethnic proportion, up to the 4th order (first 4 columns) and controlling for observed building attributes (such as age, number of 1-room flats, number of 2-room flats etc.) and month and town fixed effects (5th column).

If sellers could arbitrage perfectly, we should expect prices to be smooth across the quota limits. However, I find robust and statistically significant evidence of discontinuities in prices, ranging from magnitudes of 1% to 9%.[25] Figure 2 summarizes these findings.

One take-away from the descriptive analysis is that these discontinuities in prices suggest heterogeneity in ethnic preferences. When the Chinese quota binds, non-Chinese sellers cannot sell to Chinese buyers, so prices have to fall to attract non-Chinese buyers who would

[25] Without knowing the ethnicities of the buyers and sellers, there is no prediction on the sign of the discontinuity in prices at each quota.

have been outbid by Chinese buyers without the quota policy. If willingness-to-pay as a function of Chinese proportions were the same for Chinese and non-Chinese buyers, there would be no price dispersion to arbitrage away and hence, the quotas would have no price discrimination effects. This is not conclusive evidence because the discontinuities could be due to units right above the quota being unobservably different than units right below the quota. To test the price discrimination effects directly, I need data on how much buyers of each ethnicity paid.

4.2 Do estimated prices differ across ethnic groups?

I first use the phonebook data on the ethnicity of the movers and the price data at the block level to estimate what prices would be if buyers were Chinese, Malay or Indian. The idea behind the estimation is that buyers in each apartment block must be from the three ethnic groups. Conditional on block level observables and neighborhood fixed effects, I assume that the variation in prices is due to the variation in the ethnic share of the buyers in each block. That is, I estimate the differential impact on the average price in a block when the share of Chinese, Malay and Indian movers for that block change. Appendix II provides details on the estimation equation.

The price discrimination test is a natural extension of equations (1) and (2) to include an interaction between the quota dummies and buyer ethnicity.[26] Without the quota, profit-maximizing sellers would sell to the top bidder. If the Chinese quota binds, non-Chinese sellers cannot sell to Chinese buyers, the price would have to drop to attract non-Chinese buyers who were previously outbid. Therefore, the quota has a clear prediction on the sign of the interaction term: non-Chinese buyers should pay less than Chinese buyers in Chinese-constrained units.

$$\begin{aligned} \ln \hat{P}_{bjit} &= \alpha + \rho_1 QC_{bjit} + \rho_2 QC_{bjit} * QI_{bjit} + \eta_1 buyNC_{bji} && (3) \\ &\quad + \gamma_1 QC_{bjit} * buyNC_{bji} + \gamma_2 QC_{bjit} * QI_{bjit} * buyM_{bji} && (4) \\ &\quad + g\left(percentC_{bji}\right) + B_{bji}\beta + \varepsilon_{bjit} && (5) \end{aligned}$$

where again, C, M, and I denote Chinese, Malay and Indian respectively. Now, QC_{bjit} is a dummy that is 1 when only the Chinese quota is binding; $QC_{bjit} * QI_{bjit}$ is a dummy when both the Chinese and Indian quotas are binding, $buyNC_{bji}$ is a dummy variable that is 1 when the buyer is non-Chinese and $buyM_{bji}$ is 1 when the buyer is Malay; $g\left(percentC_{bji}\right)$ is

[26] Once I have estimated the price paid by buyers from each ethnic group, I can construct a dummy for the ethnicity of the buyer.

a 4th order polynomial of the Chinese proportion at the block level.[27] Now, the effect of the quotas on prices paid by the Chinese buyers (the omitted group) is summarized by ρ_1 (when only the Chinese quota is binding) and ρ_2 (when both the Chinese and Indian quotas are binding). This equation is estimated for units that are 10% above and below the Chinese block quota. I also estimate a similar equation for the Malay and Indian quotas, each time interacting the Malay and Indian quota dummies, with a dummy for non-Malay and non-Indian buyers respectively.

The key coefficients of interest here are the γ's and the ρ's. The idea is to test if Chinese buyers paid a higher price for Chinese-constrained blocks (ρ's > 0) and non-Chinese buyers paid a lower price for Chinese-constrained blocks (γ's ≤ 0). This tests whether prices differ across ethnic groups when the quota binds. Table 5 shows the results from the estimation. The 3 columns correspond to the regression close to the Chinese, Malay and Indian quotas. The standard errors are corrected for using dependent variables that are estimated (Lewis and Linzer, 2005).

I find suggestive evidence that prices differ across ethnic groups for Chinese and Malay quotas but not the Indian quotas. Column 1 shows that Chinese buyers paid 6% more when only the Chinese quotas are binding (ρ_1) but not non-Chinese buyers (γ's). The Malay quota has a similar effect except the standard errors are larger. I also find that contrary to expectations, blocks where both the Malay and Indian quotas bind, the Malay buyers paid a significantly lower price. The results for the estimation close to the Indian quota (column 3) do not follow the same pattern. This could be because Indians are such a minority that almost 95% of the neighborhoods fall within 10% of the Indian quota. This could make it hard to identify the effect of the Indian quota using regression discontinuity because the effect could be confounded by blocks whose ethnic proportions are too far from the quota limits.

5 Utility Specification

To recover ethnic preferences of households 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

$$U_{ijm}^G = X_{jm}^G \beta_i^G + \alpha_i^G P_{jm}^G + \xi_{jm}^G + \varepsilon_{ijm}^G \tag{6}$$

[27] There is no dummy for when both the Malay and Chinese quotas bind because it is impossible to have a block with 87% Chinese and 25% Malays.

for $j = 1, ..., J_m$ $\forall m$, where X^G_{jm} is a K-dimensional (row) vector of observed neighborhood 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.

I make two departures from standard residential location choice models: I allow taste for neighborhood amenities to vary by ethnic groups, ξ^G_j. This allows the interpretation of taste for Chinese neighbors, $\bar{\beta}^C_{percentChinese}$, that is separable from taste for Chinese amenities. 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.[28]

An important assumption is that utility from neighborhood j only depends on the attributes of that neighborhood alone.[29] 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.

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

$$\begin{pmatrix} \beta^G_i \\ \alpha^G_i \end{pmatrix} = \begin{pmatrix} \bar{\beta}^G \\ \bar{\alpha}^G \end{pmatrix} + \Sigma \nu^G_i \tag{7}$$

where $\nu^G_{i1}, ..., \nu^G_{iK}$ is individual i's unobserved taste for attribute k, drawn independently (for each individual in each group) from a standard normal distribution and ν^G_{iP} is drawn from a log normal distribution because I assume that all individuals do not like to pay high prices. $\sum$ is a (K+1)x(K+1)-dimensional scaling matrix whose diagonal elements are denoted by σ_k and σ_P.[30]

[28]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. Identification without additivity and separability of the unobserved neighborhood amenity is non-trivial (Imbens and Newey, 2009).

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

[30]Note that I assume mean preferences vary by group but the standard deviation does not ($\sum$ 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, Levinsohn, and Pakes, 2004). I only have 7 markets in my data.

To estimate ethnic preferences, the neighborhood attribute of interest is *Percent Own Ethnic Group*, the percent of residents in neighborhood j who belong to the same ethnic group, and its squared.

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

$$U_{i0}^G = \xi_0^G + \varepsilon_{i0}^G \tag{8}$$

Substituting (7) into (6) 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 the mean utility for neighborhood j

$$\delta_j^G = X_j^G \bar{\beta}^G + \bar{\alpha}^G P_j^G + \xi_j^G \tag{9}$$

and a consumer-specific deviation from that mean

$$\mu_{ij}^G = \sum_k \sigma_k x_{jk}^G \nu_{ik}^G + \sigma_P P_j^G \nu_{iP}^G + \varepsilon_{ij}^G \tag{10}$$

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 high tastes for attribute k (υ_{ik}) will tend to substitute towards products that are abundant in attribute k (x_{jk}).

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

$$s_j^G(\delta^G, \theta^G; x^G, P^G, F_{\mu^G}) = \int_{A_j^G(\delta^G, \theta^G;\ x^G, P^G)} F_{\mu^G}(d\mu^G), \tag{11}$$

where A_j^G is the set of consumers of group G who choose neighborhood j and θ^G is the set of taste parameters, $\{\bar{\beta}^G, \bar{\alpha}^G, \sigma\}$

$$A_j^G(\delta^G, \theta^G;\ x^G, P^G) = \left\{\mu^G : U_{ij}^G > U_{ij'}^G\ \forall\ j' \in J\right\} \tag{12}$$

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 $\varepsilon' s$ 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,

$$s_{ij}^G = \frac{exp(\delta_j^G + \sum_k \sigma_k x_{jk}^G \nu_{ik}^G + \sigma_P P_j^G \nu_{iP}^G)}{1 + \sum_{j'} exp(\delta_{j'}^G + \sum_k \sigma_k x_{j'k}^G \nu_{ik}^G + \sigma_P P_{j'}^G \nu_{iP}^G)} \tag{13}$$

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 equal to the shares predicted by the model, where the ethnic share is the percent of group G households in a market who chose neighborhood j.[31] I simulate the integral in (11) by drawing 10,000 ν_i^G's independently for each group G.

Step 2: Method of simulated 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, $\theta_0 \in R^p$:

$$E[g(\theta_0)] \equiv E[\xi(\theta_0)|Z] = 0 \tag{14}$$

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

$$\sum_{j=1}^{J} \hat{g}_j(\theta) = \sum_{j=1}^{J} Z_j' \hat{\xi}_j(\theta) \tag{15}$$

The MSM estimator, $\widehat{\theta}$, 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 the discussion in Newey (1984). Since the price and taste parameters are

[31] There are 13 neighborhoods with no movers in my sample period at all, 2 neighborhoods with no Chinese movers, 4 with no Malay movers, 6 with no Indian movers and 1 with no Malay nor Indian movers. For these neighborhoods, I assign their shares to be the minimum share for each ethnic group. Because the estimation involves the inversion of ethnic shares, shares of neighborhoods that are zero are not invertible. As an approximation, I assign minimum shares to these neighborhoods.

estimated sequentially, the first set of moments come from the OLS price regression with parameters π and γ (the first moments are $g^1(\pi, \gamma)$). Then, using these parameters as inputs, the second moments are the moments in (15), $g^2(\theta, \hat{\pi}, \hat{\gamma})$.

6.2 Identification

Why are quotas useful?

The model in this paper improves over existing location choice models by allowing the mean utility for neighborhood amenities, ξ^G, to vary by ethnic group. This is an important improvement because a model that identifies only the average taste for neighborhood amenities will bias the coefficient on ethnic proportions upwards. For example, let the true model for mean utility be $\delta_j^C = X_j\bar{\beta} + \bar{\alpha}P_j + \xi_j^C$, and decompose the Chinese taste for neighborhood amenities to a mean and a Chinese-specific deviation from the mean, $\xi_j^C = \xi_j + \Delta\xi_j^C$. Then, existing location choice models, $\delta_j = X_j\bar{\beta} + \bar{\alpha}P_j + \xi_j$, will have residual variation in Chinese amenities, $\Delta\xi_j^C$, that will be positively correlated with Chinese neighborhood proportions.

However, a location choice model with ethnic-specific tastes for amenities, ξ^G, 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 through the price discrimination mechanism. This price dispersion helps to identify the taste for Chinese amenities separately from the taste for non-Chinese amenities: Conditional on the proportion of Chinese residents in a neighborhood, Chinese buyers are willing to pay more than non-Chinese buyers for units in the same neighborhood 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, $\beta^G_{percent\,own\,ethnic\,group}$, and ethnic-specific taste for neighborhood amenities, ξ^G, is no longer under-identified.

The discussion above and the descriptive analysis in Section 4 shows that there exists price dispersion in the data that could be correlated with both ethnic proportions and amenities. To identify the taste for ethnic proportions, *separately* from the taste for amenities, we will need instruments for ethnic proportions and prices because they are both correlated with unobserved amenities.[32] Table 6 summarizes the identifying assumptions behind each instrument.

[32] I follow the discrete choice literature and assume that all other observed neighborhood attributes are exogenous.

Instruments for ethnic proportions

First, note that quota dummies are not instruments for ethnic proportions for 3 reasons.[33] First and most importantly, the policy rule is itself a function ethnic propotions. Second, within the time period in my data (April 2005-March 2006), the policy generates variation in the flows of movers from different ethnic groups, but the flows are not enough to change the ethnic make up of residents in a neighborhood. Third, the quota status is endogenous. For example, the Chinese quota can bind both because the Chinese proportion is high or the neighborhood has better Chinese amenities.

My first instrument for ethnic proportions relies on historical ethnic settlements. 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 (Figure 1) 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 2 shows the 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 the Malay neighborhoods expanded to the east of the river after the Malay settlements were assigned there. I define a dummy variable that is 1 when the neighborhood is in the east of early Malay settlements, and 0 otherwise. While this exclusion restriction is not testable and it is impossible to collect good data for all neighborhood amenities, it is comforting that the spatial distribution of mosques, an important amenity for Malays, does not seem to be concentrated to the east of early Malay settlements.[34]

My second set of instruments for ethnic proportions follows Bayer and Timmins (2007) and Bayer, McMillan, and Rueben (2004). I use the sum of nearby neighborhood attributes.[35] The equation for mean utility, $\delta_j^G = X_j^G \bar{\beta}^G + \bar{\alpha}^G P_j^G + \xi_j^G$, implies that using attributes of nearby neighborhoods as instruments will satisfy the exclusion restriction because attributes of $X_{j' \neq j}$ 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,

[33]Indeed, good instruments for ethnic proportions at the neighborhood level are hard to come by because of endogenous sorting. Cutler, Glaeser, and Vigdor (1999) and Edin, Fredriksson, and Åslund (2003) circumvent the problem by instrumenting for ethnic proportions at the city level. Bayer, Ferreira, and McMillan (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.

[34]Mosques are important for Malays because 98% of Malays are Muslims.

[35]In theory, the non-linear nature of the discrete choice model allows me to use polynomials of the surrounding neighborhood attributes and the products of different neighborhood attributes. In practice, I only use the sum because there are only 170 neighborhoods and some of the attributes are collinear.

Levinsohn, and Pakes (1995, 2004); Petrin (2002) and Nevo (2001)).[36] To adapt it to location choice models, where the taste for adjacent neighborhoods could be positive and to account for spatial corelation, my instrument uses 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).[37]

According to Bayer and Timmins (2007) and Bayer, McMillan, and Rueben (2004), 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.[38] The thought experiment involves 2 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 A since Malays prefer big units to the surrounding small units. In this manner, unit sizes of surrounding neighborhoods are correlated with Malay proportions.

Instruments for ethnic-specific prices

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

To approximate the ideal experiment, I use the instruments for ethnic proportions above to isolate plausibly exogenous variation in ethnic proportions, and hence, variation in whether the quota binds or not. First, I estimate the block and neighborhood level ethnic proportions using the instruments for ethnic proportions described above. The estimation equations for

[36]The assumption in the framework of Berry et al. (1995, 2004)Berry et al. (1995, 2004) is that while market shares are a non-linear function of ξ^G, 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 be estimated using instrumental variables.

[37]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.

[38]Forty-three percent of Malay households have 5 or more family members while only 24% and 26% of Chinese and Indian households have such large families (Housing Development Board, 2000).

block and neighborhood proportions are

$$percentG_{bj} = X_{bj}^{ex}\gamma_1 + \gamma_2 East_j + Z_j\gamma_3 + u_{bj} \quad (16)$$

$$percentG_j = X_j^{ex}\rho_1 + \rho_2 East_j + Z_j\rho_6 + v_j \quad (17)$$

where G=(C)hinese, (M)alays and (I)ndians, b indexes blocks and j indexes neighborhoods. The variable, $percentG$ is the percent of residents from group G, X^{ex} is the set of exogenous observed attributes (everything except price and ethnic proportions), $East$ is a dummy for whether the neighborhood is to the east of the early Malay settlements, Z is the set of exogenous attributes of nearby neighborhoods. Using these equations to predict ethnic proportions, I assign the estimated block and neighborhood quota dummies to be 1 if the estimated ethnic proportions are above the quota limits, and 0 otherwise.

Since actual quota status varies for endogenous and exogenous reasons, this method isolates plausibly exogenous variation in the predicted quota dummies. That is, conditional on the instruments for ethnic proportion, the predicted quota dummy is assumed to be correlated with amenities only through ethnic-specific prices.[39]

The step function of the policy rule is important.[40] The identification assumption is that the estimated quota dummies (whether the quota is binding or not) are correlated with ethnic-specific prices through price discrimination, but other neighborhood attributes do not affect prices discontinuously at the quota limits. Thus, the effect of the estimated quota dummies is non-parametrically identified using the step function of the quota policy. That is, even though the quota dummies were estimated by projecting ethnic proportions onto the space of the existing instruments (X^{ex}, Z and $East$), the estimated quota dummies should still have power to predict ethnic prices as long as the effect of X^{ex}, Z and $East$ on prices is not discontinuous like the step function. To check this, I estimate the following equations

$$QG_j = \chi\widehat{QG}_j + v_j \quad (18)$$

$$QG_j = \phi_0 + X_j^{ex}\phi_1 + \phi_2 East_j + Z_j\phi_3 + \phi_4\widehat{QG}_j + \omega_j \quad (19)$$

That is, I regress the actual quota status for Chinese, Malay and Indian quotas (QG_j) on the estimated quota status $\left(\widehat{QG}_j\right)$. For example, QC_j is the percent of blocks in neighborhood j where the Chinese quota is binding.[41] Also, I regress actual quota status on the

[39]The identification assumption is that group G's taste for unobserved amenities is mean-independent of the instruments, $E\left[\xi^G|Z\right] = 0$. This assumption fails if, conditional on the instruments (including the predicted quota dummy), the taste for amenities is discontinuous at the quota limits.

[40]The quota dummy is a step function because it is 1 (constrained) when the ethnic proportions are above the quota limit, 0 otherwise.

[41]This is a percentage instead of a dummy because there are block and neighborhood quotas. This number

estimated quota status, controlling for the instruments used to estimate the quota dummies. If quotas have power above and beyond the exogenous attributes used to estimate them, then, the coefficient, ϕ_4, 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. As a robustness check, I repeat the regression above using placebo limits to test whether they have significant effects.

Panel A of Table 7 reports the results for this "first stage". The first 3 columns report results from regressions of the actual quota status on the estimated quota status (equation (18)). As expected, the estimated quota dummies are all positive and significantly correlated with actual quota status. Importantly, after controlling for the instruments used to estimate these quota dummies (columns 4-6), the estimated Chinese and Malay quota dummies still have a significant effect, indicating that the step function is powerful. But not so for the Indian quotas probably because the price discontinuities were not large for the Indian quotas (see Table 4).

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 which are expected to have no effect on prices. I use placebo limits that are 3% above and below the actual limits. Importantly, the coefficient on the predicted Chinese quota dummies no longer have statistical power, after controlling for the instruments for ethnic proportions indicating that what is identified is due to the policy rule.

The choice of placebo limits is constrained for the Malay and Indian quotas because placebo limits that are too far from the actual limits are dropped. For example, only 1 neighborhood was estimated to be Malay-constrained and all Indian quota dummies were estimated to be zero when the placebo limit was 3% above the actual limit. On the other hand, placebo limits that are too close to the actual limit would likely still have an effect because there could be measurement error in the ethnic proportions if names are not perfectly matched to ethnicities. This could be why the Malay quota dummies 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.

My second instrument for prices is the same as the price instruments in Bayer, Ferreira, and McMillan (2007). The idea is to use the equilibrium conditions of the model to find optimal weights for the price instruments. The first step uses the price instruments (including the predicted quota dummies and exogenous attributes of surrounding neighborhoods) to

is 1 when the neighborhood quota binds and less than 1 when the neighborhood quotas is unconstrained but some blocks in the neighborhood are constrained.

estimate a demand model with exogenous variation only. The second step uses the model to solve for a vector of equilibrium prices that clears the market. This 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

Columns 1-4 in Table 8 report results on taste parameters estimated with the random coefficients model.[42] 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.[43] Interpreting the magnitudes of the mean marginal utilities ($\bar{\beta}$), living 1 km further away from the subway station is as bad as living in a neighborhood where the average building is 4.6 years older (both are worth 0.31 utils less).

The estimates on ethnic preferences show that all groups want to live with at least some members of their own ethnic group. Interestingly, the shape differs significantly across ethnic groups. The Chinese and Indian preferences are inverted-U shaped: the marginal utilities for own ethnic group neighbors are positive below 45% Chinese and 12% Indians respectively but are negative for neighborhoods with a higher concentration of neighbors from the same ethnic group.[44] Using the delta method, I test whether the marginal utilities are significantly different from zero at these two turning points. The t-statistics are 0.0173 and -0.1964, respectively, indicating that the inverted-U shape is significant but the location of the turning point is not.[45] Previous research in the United States have documented evidence of tastes for diversity using data on racial attitudes from the General Social Survey

[42] See the online appendix 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).

[43] For example, this restricts Chinese, Malays and Indians to share the same tastes for school quality. If Chinese had stronger tastes for education than the other groups, neighborhoods with a high school quality would tend to be more Chinese-segregated. This could upward bias the estimates of Chinese preferences.

[44] Although 45% represents the Chinese proportion that yields the maximum utility for the average Chinese, all neighborhoods do not converge towards 45% because the Chinese make up 77% of the population and there is variation in other attributes that are desired.

[45] As a comparison, the turning points for the quadratic model using IV Logit are 47% and 12% for the Chinese and Indians.

(Aldrich, Arcidiacono, and Vigdor, 2005) but most empirical estimates of ethnic preferences have not been able to demonstrate such non-monotinicity in ethnic preferences because they have focused on linear models.

By contrast, the Malays do not exhibit such non-monotonic tastes. The estimated marginal utilities for Malays are positive for both terms, but the coefficients are only significant jointly (the F statistic is 16.44). This is also suggestive evidence against a pure income segregation story. If observed segregation patterns are driven purely by the desire to live near (unobserved) high income neighbors, then all ethnic groups should want to live with Chinese because they have the highest income.

The finding that Chinese do have preferences for new neighbors that are non-Chinese is suggestive evidence against ethnic discrimination. The concern with ethnic discrimination is that Malay enclaves may form even when Malays do not have strong Malay preferences because the Chinese are discriminating against Malays and forcing them into Malay enclaves. This could mean that my estimate of Malay preferences is an overestimate. However, to the extent that the Chinese are discriminating against Malays, the Chinese should not have these inverted U-shaped preferences.

One concern is that quadratic ethnic preferences could be too restrictive because it imposes a symmetric functional form around the turning points. Ideally, if we had infinite degrees of freedom, we could estimate the turning points using a model with higher orders of polynomials. In Table 8, columns 5-12, as an alternative to the quadratic specification, I estimated two other random coefficient models that are linear and cubic in ethnic proportions, respectively. For the linear model, the coefficients for ethnic preferences are still positive but only significant for the Malays. This is not surprising given that I find strong non-linearities for Chinese and Indian preferences in the quadratic model while Malay preferences appear to be linear and positive. The cubic model is evidently lacking degrees of freedom. The standard errors on the coefficients of the quadratic and cubic terms are big.[46] The marginal utility estimates for all other attributes are similar to the estimates in the quadratic model.

In Table 9, I use the ratio of the marginal utility estimates from the quadratic specification (the preferred specification) in Table 8 to calculate the marginal rates of substitution (MRS) between *Percent Own Ethnic Group* and other attributes. I quantify ethnic preferences in terms of two units, dollars (marginal willingness-to-pay, MWTP) and the average distance to the subway station (in kilometers). Because of the quadratic term on ethnic proportions,

[46] One feature of the cubic model is that it allows for the possibility of two turning points in ethnic preferences. Even though the coefficients are not significant, I calculated the critical points in the cubic model and found a minimum (22%) and maximum (77%) for the Chinese. This seems counter-intuitive because it implies that the marginal utility for another Chinese neighbor is *negative* below 22% and above 77% but *positive*, in between.

the MRS changes when the ethnic proportions in a neighborhood change. I report MRS calculations for different parts of the distribution of ethnic proportions (1st percentile, mean, 99th percentile). A Malay living in a neighborhood with 13% Malays (the mean) is willing to pay S$7600 (S$1=US$0.61) to live in a neighborhood with 1% more Malays. [47] In terms of distance, he is willing to move to a neighborhood where the average distance to the subway station is 0.37km (1km = 0.62 miles) farther. The average distance in the sample is 0.8km. Indians have very steep indifference curves. An Indian living in a neighborhood with 2% Indians (the 1st percentile) is willing to pay S$31,000 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 be *compensated* S$30,000 to live in a neighborhood with 1% more Indians.

7.1 First Best Simulations

In this section, I use the preference estimates above to find the first best allocation and compare it to the existing decentralized equilibrium under the quota regime.[48] The social planner's problem is to find the allocation of ethnicities into neighborhoods that will maximize a social welfare function, assumed to be utilitarian here. In a decentralized equilibrium, individuals choose the neighborhood that maximizes his own utility, without internalizing the effect of his choice on the ethnic proportions in the neighborhood. Due to externalities, the decentralized equilibrium may not achieve the first best allocation. The social planner, by contrast, may assign an individual to a neighborhood that is suboptimal for him, but optimal for his neighbors. See Appendix 3 for the simulation details.

Figure 4 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 and high Malay proportions. Ten years after the quota policy has been introduced, I find that 71% of the neighborhoods have Chinese proportions that are within 1 standard deviation (7%) of the first best allocation and 18% are within half a standard deviation. For the Malay and Indian proportions , roughly half are within

[47] This MWTP translates to 3% of the average price (S$241,000), which is of the same order of magnitude as Bayer, Ferreira, and McMillan (2007). They find that Blacks are willing to pay $98 more per month than whites to live in a neighborhood that has 10% more black versus white households. The average monthly rent is $744. Since they have a utility model that is linear in *percent Black*, this translates to a MWTP of 1.3% of rents. By contrast, Bajari and Kahn (2005) find that Blacks are willing to pay less per year than whites to live in a PUMA with a higher proportion of blacks.

[48] In theory, one can use preference estimates to estimate compensating and equivalent variation measures of 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, Hong, and Ryan (2007) for an example of how to estimate structural models with multiple equilibria.

1 standard deviation and 21% and 13% respectively, are within half a standard deviation. The standard deviation for Malay and Indian proportions are 7% and 3%.

Table 10 looks at 3 towns where I have data on ethnic proportions before the quota was implemented in 1989.[49] 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 and Yishun were almost 4 times the first best levels. Ten years after the introduction of the quotas, they were within 5 percentage points of the first best Malay and Indian proportions. The magnitude of this "improvement" towards the first best allocation is likely to be over-optimistic because these towns were very segregated to begin with. The impact on Chinese proportions is not big probably because the Chinese are such a majority (77%) that it is hard to lower Chinese neighborhood proportions despite the Chinese having inverted U-shaped preferences.

There are two major 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), the effect of more diverse neighborhoods on welfare is ambiguous. 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.

8 Conclusion and Future Research

This paper estimates WTP 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. 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

[49] Towns are bigger than neighborhoods. Data on ethnic proportions at a disaggregated level were not available in 1988.

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. However, the Chinese and the Indians exhibit preferences that are inverted U-shaped so that after a neighborhood is segregated enough, they would rather add a new neighbor from the other group. To my knowledge, this represents the first estimate of non-monotonic ethnic preferences. made possible due to the rich variation from the phonebook data. Comparing data from 3 segregated towns before the quota to the first best allocation, I find that after 10 years, the quotas have moved the Malay and Indian proportions more than half-way (within 5 percentage points) to the first best although this effect could be an upper bound due to sample selection of pre-quota towns.

In ongoing work, I explore the possibilities of comparing hedonic and discrete choice models to estimate ethnic preferences (Wong, 2010b). 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., and Benkard, C. L. (2005). Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy*, *113*(6), 1239–1276.

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.

Bartik, T. J. (1987). The Estimation of Demand Parameters in Hedonic Price Models. *Journal of Political Economy*, *95*(1), 81–88.

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.

Boisjoly, J., Duncan, G., Kremer, M., Levy, D., and Eccles, J. (2006). Empathy or Antipathy? The Impact of Diversity. *American Economic Review*, *96*(5), 1890–1905.

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.

Ekeland, I., Heckman, J., and Nesheim, L. (2004). Identification and Estimation of Hedonic Models. *Journal of Political Economy*, (S2), S60–S109.

Epple, D. (1987). Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products. *Journal of Political Economy*, *95*(1), 59–80.

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.

Imbens, G., and Newey, W. (2009). Identification and Estimation of Triangular Simulataneous Equations Models without Additivity. *Econometrica*, *77*(5), 1481–1512.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, *75*(1), 83–119.

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

Lewis, J., and Linzer, D. (2005). Estimating Regression Models in Which the Dependent Variable Is Based on Estimates. *Political Analysis*, *13*(4), 345–364.

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 P. Zarembka (Ed.) *Frontiers in Econometrics*. New York: Academy Press.

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

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 S. N. Durlauf, and H. P. Young (Eds.) *Social Dynamics*. Washington, D.C.: Brookings Institution Press.

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.

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

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, *82*(1), 34–55.

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.

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

Straits Times (7 January 1989).

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

Wong, M. (2010a). Estimating the Impact of the Ethnic Housing Quotas in Singapore. University of Pennsylvania, Real Estate Department Working Paper.

Wong, M. (2010b). The Relationship between the Hedonic and Discrete Choice Models. 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

Variables	N	Mean	Std. Dev.	Description
Chinese Share	170	0.09%	0.11%	Percent of Chinese in a market who chose a neighborhood
Malay Share	170	0.13%	0.14%	Percent of Malays in a market who chose a neighborhood
Indian Share	170	0.30%	0.31%	Percent of Indians in a market who chose a neighborhood
Price	170	239,888	50,769	Average transaction price in a neighborhood (Singapore dollars)
Chinese Neighborhood Quota	170	0.08	0.25	Percent of months Chinese neighborhood quota binds
Malay Neighborhood Quota	170	0.05	0.19	Percent of months Malay neighborhood quota binds
Indian Neighborhood Quota	170	0.17	0.33	Percent of months Indian neighborhood quota binds
Chinese Block Quota	170	0.10	0.18	Percent of months and blocks Chinese block quota binds
Malay Block Quota	170	0.05	0.12	Percent of months and blocks Malay block quota binds
Indian Block Quota	170	0.09	0.15	Percent of months and blocks Indian block quota binds
Chinese Quota	170	0.18	0.29	Percent of months and blocks any Chinese quota binds
Malay Quota	170	0.11	0.23	Percent of months and blocks any Malay quota binds
Indian Quota	170	0.25	0.35	Percent of months and blocks any Indian 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 AT THE QUOTA LIMITS

	DEPENDENT VARIABLES														
	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price	Ln Price
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Chinese Quota	0.09***	0.08***	0.09***	0.07***	0.04***										
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)										
Malay Quota						-0.03***	-0.05***	-0.05***	-0.05***	-0.04***					
						(0.01)	(0.01)	(0.01)	(0.01)	(0.005)					
Indian Quota											-0.01**	-0.01*	-0.02***	-0.02**	-0.03**
											(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Polynomial	Linear	Quadratic	Cubic	Quartic	Quartic	Linear	Quadratic	Cubic	Quartic	Quartic	Linear	Quadratic	Cubic	Quartic	Quartic
Controls	N	N	N	N	Y	N	N	N	N	Y	N	N	N	N	Y
Month	N	N	N	N	Y	N	N	N	N	Y	N	N	N	N	Y
Town	N	N	N	N	Y	N	N	N	N	Y	N	N	N	N	Y
Obs	19314	19314	19314	19314	19314	14862	14862	14862	14862	14862	32114	32114	32114	32114	32114
R-squared	0.01	0.01	0.01	0.02	0.80	0.003	0.004	0.01	0.01	0.75	0.008	0.009	0.10	0.10	0.78

NOTE. – The regression equation is $lnP_{bjit} = \alpha + \gamma QC_{bjit} + f(percentC_{bji}) + \varepsilon_{bjit}$ where lnP_{bjit} is the log of the price of units in block b, neighborhood j and town i; QC_{bjit} is a dummy that is 1 when the Chinese (C) quotas are binding, $f(percentC_{bji})$ is a flexible polynomial, estimated separately above and below the quota. The controls are other observable characteristics of the block (age of building, number of 1-room units, 2-room units etc.), month and town fixed effects. I repeat the exercise for the Malay quotas (columns 6-10) and Indian quotas (columns 11-15); Standard errors are in parentheses and are clustered at the town level for models with fixed effects (columns 5, 10 and 15).

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

TABLE 5: EFFECTS ON ESTIMATED PRICES AT THE QUOTA LIMITS BY BUYER ETHNICITY

	DEPENDENT VARIABLES		
	Estimated LnPrice	Estimated LnPrice	Estimated LnPrice
	(1)	(2)	(3)
Chinese Quota Status	0.06**		
	(0.01)		
Non-Chinese Buyer	-0.01**		
	(0.003)		
Chinese Quota Status x Non-Chinese Buyer	-0.01		
	(0.001)		
Chinese Quota Status x Indian Quota Status	0.20**		
	(0.01)		
Chinese Quota Status x Indian Quota Status x Malay Buyer	-0.06**		
	(0.01)		
Malay Quota Status		0.03**	
		(0.006)	
Non-Malay Buyer		-0.002	
		(0.003)	
Malay Quota Status x Non-Malay Buyer		0.01	
		(0.008)	
Malay Quota Status x Indian Quota Status		-0.07**	
		(0.006)	
Malay Quota Status x Indian Quota Status x Chinese Buyer		-0.002	
		(0.01)	
Indian Quota Status			-0.04**
			(0.005)
Non-Indian Buyer			0.003
			(0.003)
Indian Quota Status x Chinese Quota Status			0.18**
			(0.008)
Indian Quota Status x Malay Quota Status			-0.09**
			(0.007)
Indian Quota Status x Non-Indian Buyer			-0.0008
			(0.006)
Indian Quota Status x Chinese Quota Status x Malay Buyer			-0.04**
			(0.01)
Indian Quota Status x Malay Quota Status x Chinese Buyer			-0.006
			(0.01)
Controls	Y	Y	Y
Observations	10767	10149	17394
R-squared	0.24	0.35	0.28

NOTE. – Each column is a regression restricted to 10% above and below the Chinese, Malay and Indian quotas. Controls include average age of building, its squared, number of 3- to 6-room units (the number of 1-, 2-, 7- and 8-room units were dropped because of collinearity). The omitted group is the Chinese buyer (column 1), the Malay buyer (column 2) and the Indian buyer (column 3); Standard errors are in parentheses, corrected for using a dependent variable that is estimated (Lewis and Linzer, 2005).

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

TABLE 6: SUMMARY OF IDENTIFICATION STRATEGY

Identification Strategy	Assumptions and Thought Experiments
1. Omitted variables that are ethnic-specific (ξ_j^G instead of ξ_j)	
• Model neighborhood-by-ethnic-group fixed effects instead of neighborhood fixed effects (δ_j^G instead of δ_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 1 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 STATUS ON ESTIMATED QUOTA DUMMIES USING PLACEBOS

	DEPENDENT VARIABLES					
	Actual Chinese Quota Status	Actual Malay Quota Status	Actual Indian Quota Status[+]	Actual Chinese Quota Status	Actual Malay Quota Status	Actual Indian Quota Status[+]
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Placebo = Actual Limits					
Estimated Chinese Quota Dummy	0.48***			0.17*		
	0.05			0.07		
Estimated Malay Quota Dummy		0.65***			0.42***	
		0.12			0.12	
Estimated Indian Quota Dummy			0.30**			-0.14
			0.10			0.13
Controls	N	N	N	Y	Y	Y
Obs	170	170	170	170	170	170
Fstat	85.6324	31.50	8.61	7.2305	3.8181	2.327
R-squared	0.34	0.16	0.05	0.46	0.31	0.22
	Panel B: Placebo = Actual Limits + 3%					
Estimated Chinese Quota Dummy	0.39**			-0.07		
	(0.09)			(0.08)		
Estimated Malay Quota Dummy		0.00			-0.54	
		(0.28)			(0.28)	
Estimated Indian Quota Dummy			dropped			dropped
Controls	N	N	N	Y	Y	Y
Obs	170	170	170	170	170	170
Fstat	17.9	0.00	0.00	11.9	5.47	7.66
R-squared	0.10	0.00	0.00	0.58	0.39	0.46
	Panel C: Placebo = Actual Limits - 3%					
Estimated Chinese Quota Dummy	0.35**			-0.001		
	(0.03)			(0.07)		
Estimated Malay Quota Dummy		0.38**			0.21**	
		(0.06)			(0.08)	
Estimated Indian Quota Dummy			0.30**			0.12
			(0.03)			(0.10)
Controls	N	N	N	Y	Y	Y
Obs	170	170	170	170	170	170
Fstat	104.00	41.10	79.60	11.80	5.81	7.35
R-squared	0.38	0.20	0.32	0.58	0.41	0.47

NOTE. – Panels A and B 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 dummies; 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.

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

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

TABLE 8: RANDOM COEFFICIENTS LOGIT WITH QUADRATIC, LINEAR AND CUBIC ETHNIC PREFERENCES

		QUADRATIC				LINEAR				CUBIC			
		Common Taste Parameters	Chinese Taste Parameters	Malay Taste Parameters	Indian Taste Parameters	Common Taste Parameters	Chinese Taste Parameters	Malay Taste Parameters	Indian Taste Parameters	Common Taste Parameters	Chinese Taste Parameters	Malay Taste Parameters	Indian Taste Parameters
Variables	Units	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Means:													
Constant		-7.36***				-3.42				-11.11***			
		(2.29)				(6.11)				(2.30)			
School Quality	.1 awards	1.46***				1.48***				1.30***			
		(0.17)				(0.15)				(0.18)			
Distance to Subway	1 km	-0.31***				-0.25**				-0.20			
		(0.11)				(0.11)				(0.13)			
Average No. of Rooms	0.1 rooms	1.23				3.47				-0.68			
		(3.07)				(3.58)				(2.93)			
Average Age of Buildings	0.01 years	-6.71***				-8.51***				-7.72***			
		(2.45)				(2.77)				(2.26)			
Percent Own Ethnic Group			8.00***	8.25	5.63***		0.41	6.94***	1.81		-51.81**	87.96***	17.34***
			(3.21)	(7.15)	(1.54)		(0.80)	(1.66)	(1.66)		(22.40)	(22.35)	(3.09)
(Percent Own Ethnic Group)^2			-8.84***	1.28	-2.42						171.75	-62.81	-14.84
			(3.40)	(2.82)	(0.94)						(5610.73)	(1914.08)	(329.73)
(Percent Own Ethnic Group)^3											-115.74	14.57	3.50
											(3506.52)	(470.27)	(95.82)
Price	S$million	-15.2***				-19.69				-13.65***			
		(3.33)				(37.49)				(3.32)			
Heterogeneity:													
Constant		-2.71***				-0.36				-4.05***			
		(0.62)				(15.41)				(0.96)			
Average No. of Rooms	0.1 rooms	2.32***				-0.46				-0.05			
		(0.53)				(9.97)				(0.30)			
Price	S$million	0.26***				0.08				0.06			
		(0.07)				(22.47)				(0.19)			

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 1.46 utils. For the ethnic proportions, *percent Chinese, percent Chinese*^{2}, *percent Chinese*^{3} and *percent Malay* are not scaled; *percent Malay*^{2} and *percent Indian* are multiplied by 10; *percent Malay*^{3} and *percent Indian*^{2} are multiplied by 100; *percent Indian*^{3} is multiplied by 1000. 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: MWTP AND MRS EVALUATED AT VARIOUS ETHNIC PROPORTIONS IN THE SAMPLE

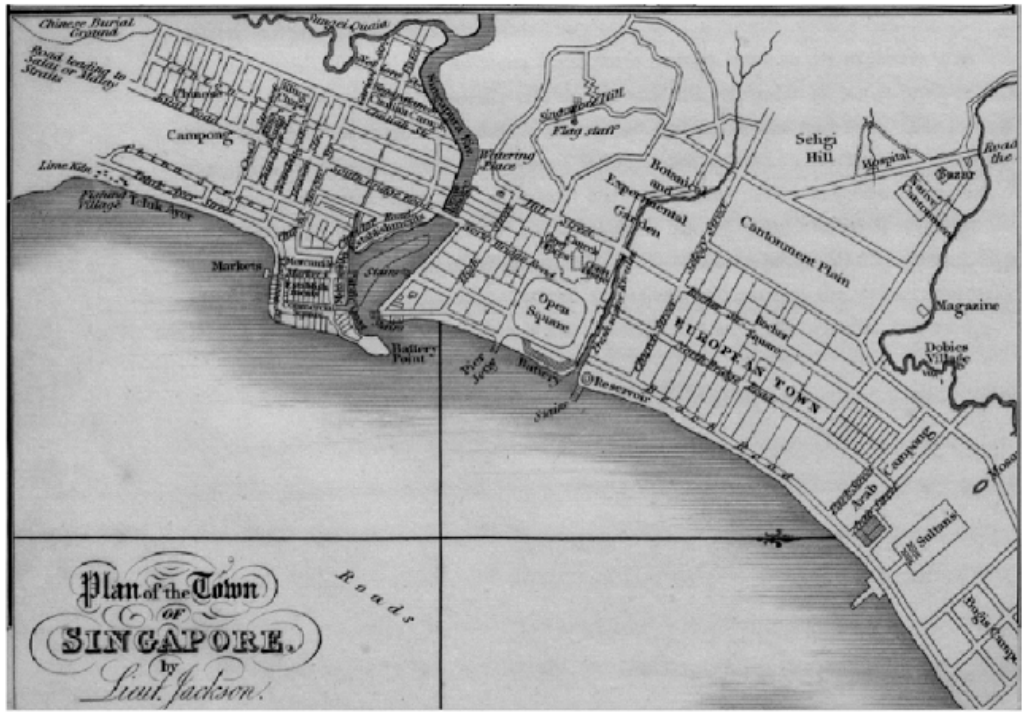
	ETHNICITY		
	Chinese	Malays	Indians
Relevant statistics for ethnic proportions:			
Mean of Percent Own Ethnic Group	79%	13%	8%
1st percentile of Percent Own Ethnic Group	63%	1%	2%
99th percentile of Percent Own Ethnic Group	98%	29%	21%
Standard Deviation of Percent Own Ethnic Group	7%	7%	3%
MWTP per 1% increase in ethnic proportion:			
MWTP at mean of Percent Own Ethnic Group	-3,926	7,617	11,566
MWTP at 1st percentile of Percent Own Ethnic Group	-2,065	5,596	30,671
MWTP at 99th percentile of Percent Own Ethnic Group	-6,136	10,312	-29,829
MRS relative to distance to subway (km), per 1% increase in ethnic proportion:			
MRS at mean of Percent Own Ethnic Group	-0.19	0.37	0.57
MRS at 1st percentile of Percent Own Ethnic Group	-0.10	0.27	1.50
MRS at 99th percentile of Percent Own Ethnic Group	-0.30	0.51	-1.46

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 2 units: dollars per 1 percentage point increase in the percent of own ethnic group neighbors (S$1=US$0.61); 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 two MRS's are calculated as the marginal utility for a 1 percentage point increase in own ethnic group neighbors divided by the negative of (i) the marginal utility for dollars (this is the MWTP in the 2nd panel); (ii) the marginal utility for distance to the closest subway station (3rd 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 price for a unit is S$240,000. 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

	Before (1988)	After (1998)	First Best
Percent Chinese in Redhill	87%	84%	75%
Percent Malay in Bedok	59%	19%	15%
Percent Indian in Yishun	24%	11%	6%

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: Crawford, 1828

FIG. 1. –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

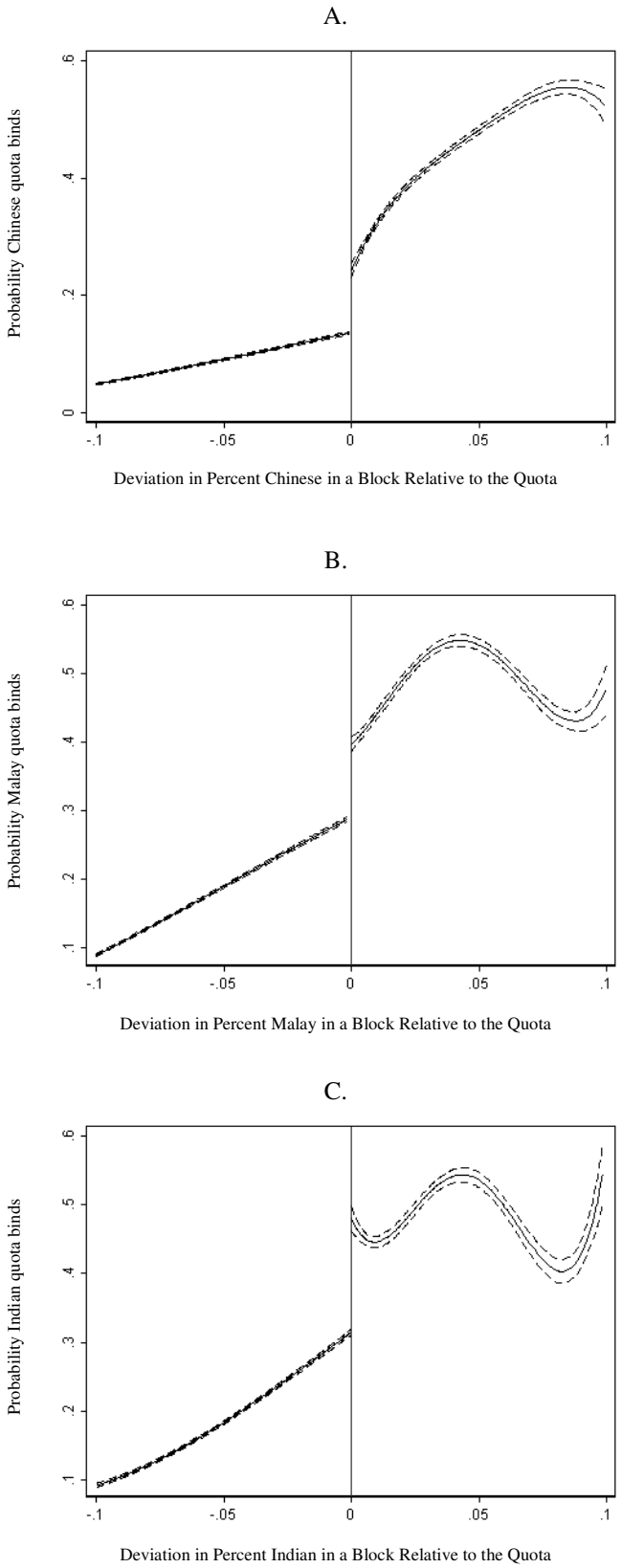
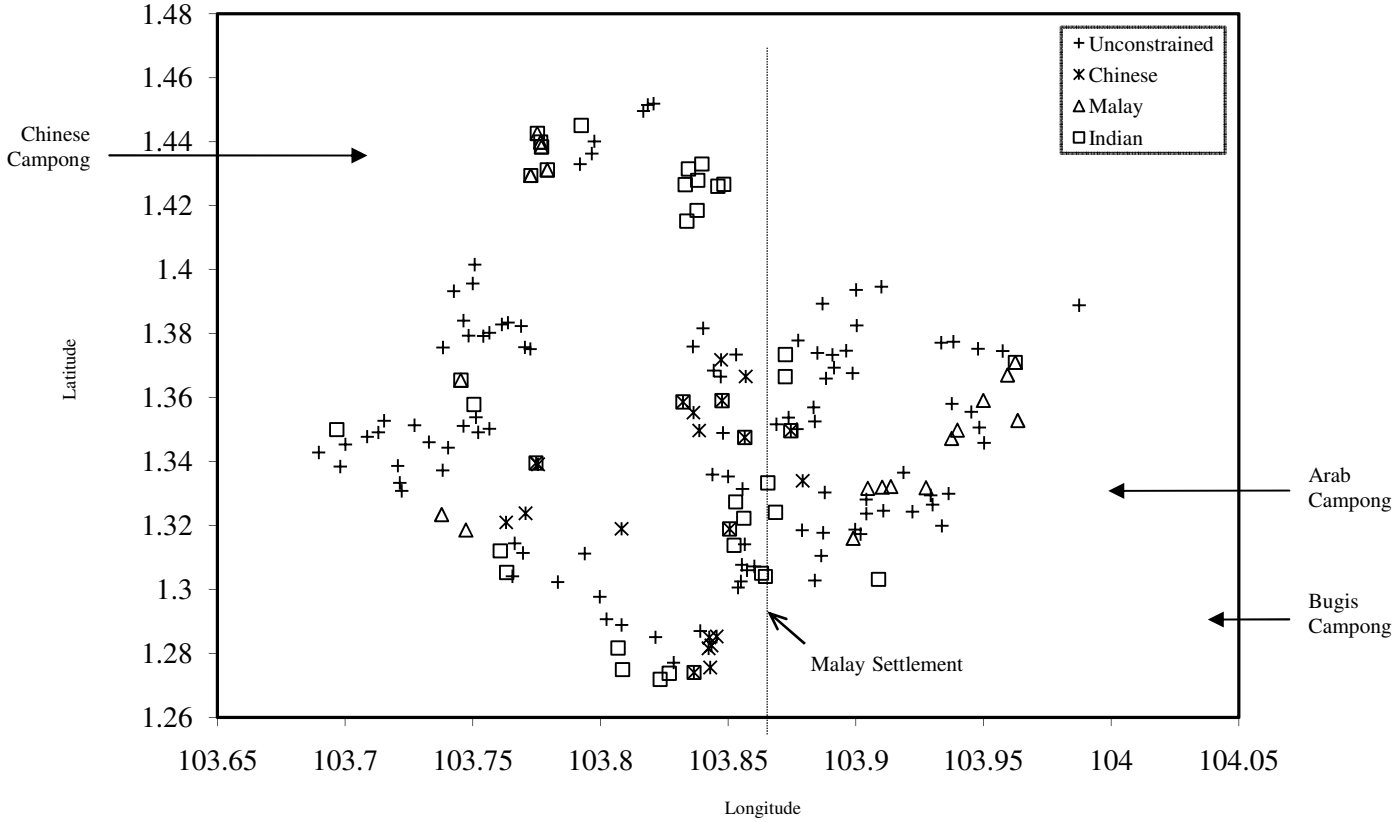
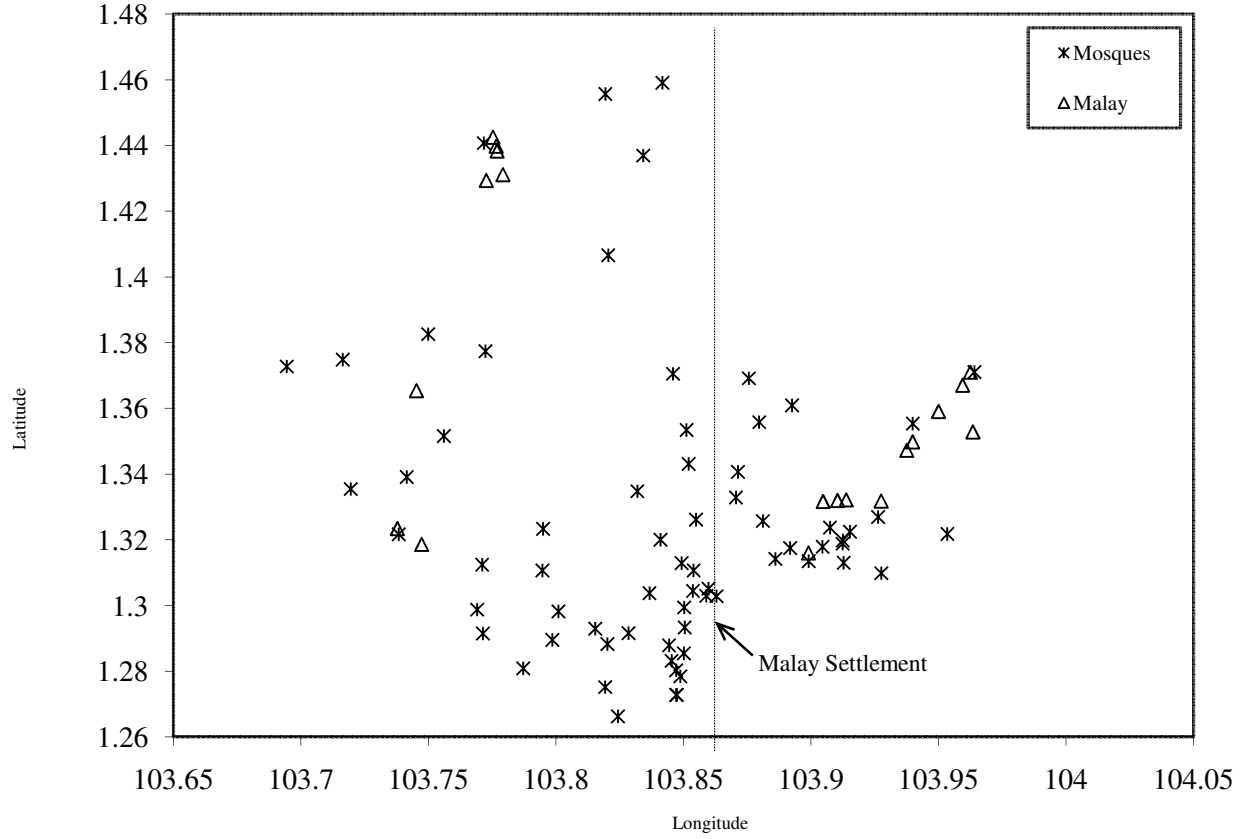


FIG. 2. – Testing for discontinuity in the probability that the quota binds, 10% above and below the quota. Each panel in this figure is constructed using the following procedure for observations within 10% of the ethnic quotas: (i) regress *Q*(a dummy for whether the quota is binding) on smooth functions of the corresponding running variable (4th order polynomials), separately, once to the left and once to the right of the quota; (ii) plot the predicted probabilities above and below the quota separately; I repeat the exercise for the Malay quotas and Indian quotas. The dashed lines represent 95% confidence intervals; *A,* probability that the Chinese quota binds, 10% above and below the quota; *B,* probability that the Malay quota binds, 10% above and below the quota; *C,* probability that the Indian quota binds, 10% above and below the quota



SOURCE: Virtual Map Online Street Directory

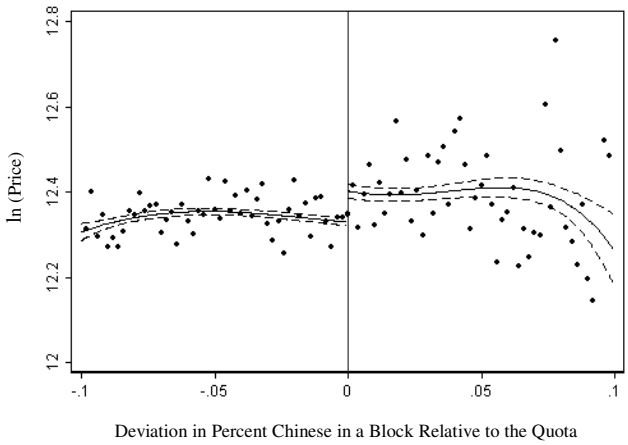
FIG. 3. — Map of 170 neighborhoods in 2005 comprising unconstrained neighborhoods, Chinese-, Malay- and Indian-constrained neighborhoods. The line indicates the eastern tip of the early Malay settlements.



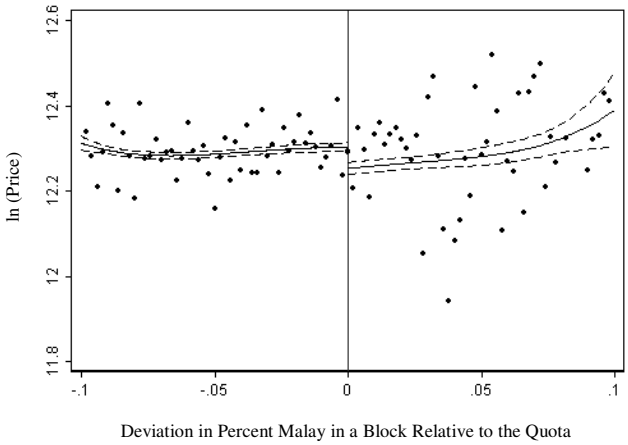
SOURCE: Virtual Map Online Street Directory

FIG.4. — Map of Malay--constrained neighborhoods and the location of mosques. The line indicates the eastern tip of the early Malay settlements.

A.



B.



C.

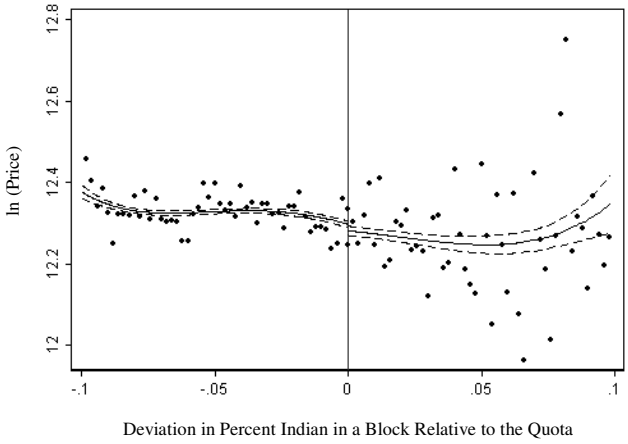
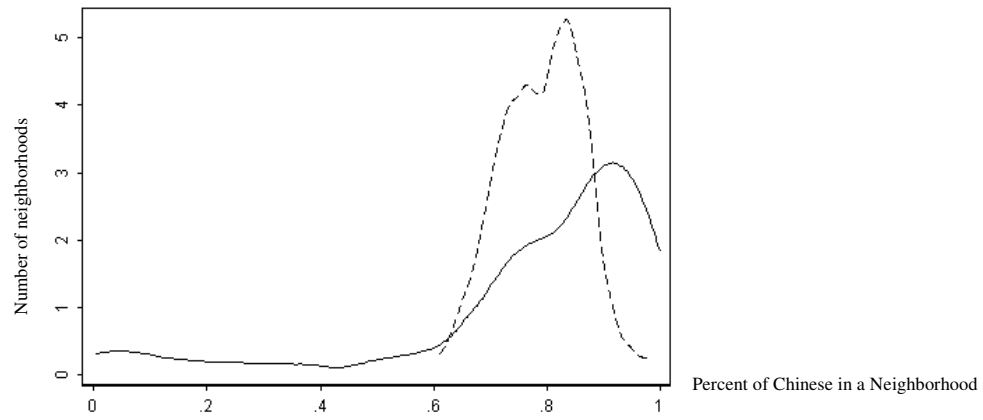
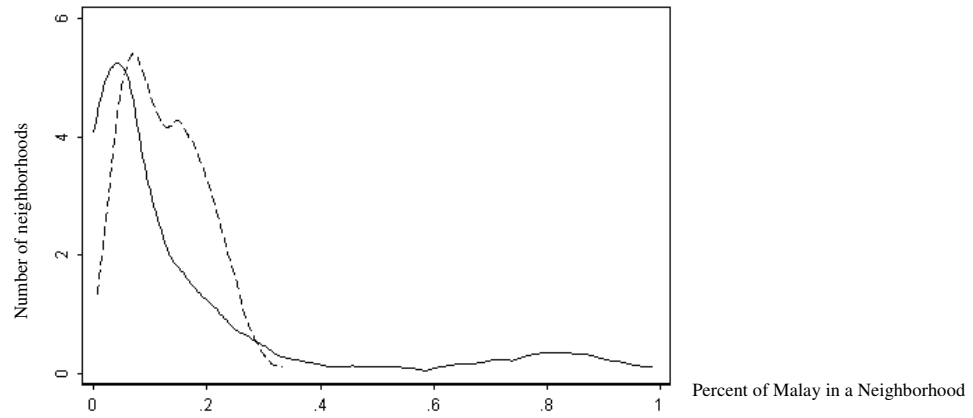


FIG. 5. – Impact of block quotas on lnPrice, 10% above and below the quota. Each panel in this figure is constructed using the following procedure for observations within 10% of the ethnic quotas: (i) regress the log of transaction prices on smooth functions of ethnic proportions (two 4th order polynomials, above and below the quota) and a dummy that is one when the corresponding block quota is binding; (ii) plot the predicted prices above and below the quota separately (solid line) as well as the 95% confidence interval (dashed lines); (iii) plot means of ln(price) for each 1% bin. I repeat the exercise for the Malay quotas and Indian quotas; *A,* the impact of Chinese block quotas on ln(price); *B,* the impact of Malay block quotas on ln(price); *C,* the impact of Indian block quotas on ln(price).

A.



B.



C.

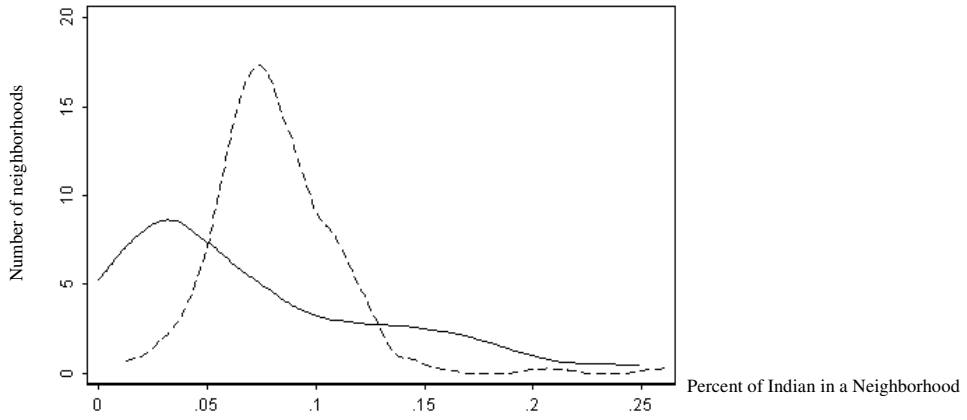


FIG. 6. – Density of ethnic proportions in a neighborhood, first best and now. Note that dashed lines represent the density now (--); Solid lines (–) represent the first best density. *A,* Percent Chinese in a neighborhood. *B*, Percent Malay in a neighborhood. *C*, Percent Indian in a neighborhood.

Appendix 1: Data

In this section, I describe some variables in more detail and list the corresponding data sources.

Choice data

I match the postal codes of individuals in the 2005 and the 2006 phonebook. Movers have to update their contact information within a month of moving. 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 phone company updates the data every year on April 1st. For my dataset, I assume that movers moved between April 2005 and March 2006 and that they changed their phone records immediately after they move.

Neighborhoods

I use six-digit postal codes to define neighborhoods. The first two digits denote the *sector*. The fourth digit denotes the neighborhood within the postal sector while the last two digits denote the block within the neighborhood. Occasionally, some blocks have an alphabet as a suffix, in which case the third digit would be non-zero. For example, postal codes 101106 and 102106 correspond to blocks 106A and 106B in postal sector 10, neighborhood 1.

School quality

I obtain data on awards given to primary, secondary schools and tertiary institutions from the Singapore Ministry of Education website. The school quality is defined as the total number of awards received from all schools and tertiary institutions in a neighborhood.

Access to subway

For each neighborhood, I calculate the distance (in kilometers) from the midpoint of the neighborhood to the closest Mass Rapid Transit (MRT) or Light Rapid Transit (LRT) station using latitude and longitude data obtained from a popular local online street directory, http://www.streetdirectory.com/ (Virtual Map).

Age

This is obtained from the resale transactions data on the HDB website. Since all blocks in the resale market were sold at some point in my dataset, I observe the age of each HDB block. I use the average age of HDB blocks in a neighborhood.

Rooms

I purchased this data from the HDB. For each HDB block, I have the number of type 1 flats, type 2 flats, etc. There are 8 types of HDB flats comprising 1-room to 5-room

flats, executive flats, HUDC and multi-generational flats. 1-room flats are studios, 2-room flats are 1 bedroom flats and so on. Executive flats, HUDC and multi-generational flats are defined as 6-room flats in my dataset.

Quotas

I collected monthly data on the ethnic quotas from the public HDB website, beginning in March 2005. These are dummy variables for whether a block was constrained. If all blocks were constrained in a neighborhood, I define the neighborhood quota as binding.

Appendix 2: Estimating Ethnic-Specific Prices

I estimate ethnic-specific prices for each neighborhood using observed transaction prices at the block level as well as the observed ethnic weights (from the data on movers from the phonebook). This estimation procedure is akin to fitting a system of equations where the variables of interest are the block level ethnic weights and the unknowns are the neighborhood level ethnic prices. The assumption is that controlling for observed block characteristics, observed prices within a neighborhood varies because the proportions of Chinese, Malay and Indian movers in each block is different. For example, a block with buyers who are 20% Chinese and 80% Malay will have a different average price from a similar block with 80% Chinese buyers and 20% Malay buyers. An HDB block is comparable to a US Census block group, with an average of 70 households. An HDB neighborhood is comparable to a US Census tract, comprising an average of 60 HDB blocks.

I estimate the following equation

$$\ln \bar{P}_{bj} = \pi^C N_j * w^C_{bj} + \pi^M N_j * w^M_{bj} + B_{bj} * N_j \gamma + \pi N_j + \upsilon_{bj}$$

where w^C_{bj} and w^M_{bj} are the Chinese and Malay buyer weights from the phonebook data; B_{bj} is a set of block-level characteristics (the block quotas, the number of 1-room units, 2-room units, etc.); N_j is a neighborhood dummy and the Indians are the omitted group. Notice that the neighborhood dummy is interacted with each explanatory variable. Using the estimates from this equation, I substitute $w^C_{bj} = 1$, $w^M_{bj} = 0$ to predict the price paid by the Chinese buyer, $\hat{P}^C_{bj}$, and likewise for the Malay and Indian prices.

Appendix 3: Simulation

The idea of the simulation is to find the allocation of ethnic groups across neighborhoods, $\{percent\ Chinese_j, percent\ Malay_j\}_{j=1}^{J}$ that maximize a utilitarian social welfare function, using the following steps:

1. For each market m, find the number of neighborhoods, J_m in that market.
2. Assume each neighborhood has 100 units. Let the population in that market be $N_m = 100 * J_m$. The number of Chinese, Malays and Indians are $0.77N_m$, $0.14N_m$ and $0.09N_m$ respectively.
3. Randomly draw J_m Chinese and Malay proportions, $\{percent\ Chinese_j,\ percent\ Malay_j\}_{j=1}^{J_m}$ such that the mean Chinese proportion in the market is 77% and the Chinese and Malay proportions sum to less than 1 for each neighborhood.
4. Assign Chinese, Malays and Indians to live in each neighborhood, where the number assigned to each neighborhood is determined using the Chinese and Malay proportions drawn in the previous step. This step determines the function $j(i)$ for each i.
5. Draw the corresponding idiosyncratic taste shocks for each ethnic group, where the individual taste for characteristics, υ_{ik}^{G}, is common across neighborhoods and the logit error, ε_{ij}^{G}, is neighborhood-specific.
6. Calculate utility, $U_i = U_{ij(i)}$ where the assignment of individual i to neighborhood j is determined in step 4 and utility is defined in the model to include utility from observed and unobserved neighborhood attributes.
7. Sum the individual utilities to get the social welfare function.
8. Repeat steps 3-7 10,000 times for each market. Determine the ethnic proportions that maximize a utilitarian social welfare function.