Estimating Ingroup Preferences Using Ethnic Housing Quotas in Singapore

Maisy Wong [*,†]
Massachusetts Institute of Technology

November 13, 2007

Job Market Paper—Draft
Preliminary & Incomplete

Abstract

Many countries try to promote group integration in areas such as housing, education, immigration, employment etc. While several studies have focused on the consequences of segregation, there is little work studying the causes of segregation. One potential cause is ingroup preference. This paper estimates ethnic ingroup preferences (taste to interact with ethnic ingroup members) over residential neighborhoods using variation from an ethnic housing quotas policy in Singapore. My approach combines policy induced variation akin to a regression discontinuity identification strategy with a structural demand estimation framework à la Berry, Levinsohn, and Pakes (1995). I find that all groups have strong preferences to live with at least some other members of their ethnic group. However, the majority group (Chinese) exhibits preferences that are inverted U-shaped so that after a neighborhood reaches 43% Chinese, they would rather add a new neighbor from the other group.

[*]I wish to thank my advisors Professors Esther Duflo, Amy Finkelstein, Panle Jia and Bill Wheaton for guidance and support. I am also deeply indebted to Professors Peter Diamond, Michael Greenstone, Whitney Newey and Stephen Ryan for their helpful comments. I have benefited from conversations with Professor Chua Beng Huat, Jessica Cohen, Greg Fischer, Raymond Guiteras, Kam Lee Ching, Trang Nguyen and Wong Liang Kit. 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. All remaining errors are my own.

[†]E-mail: maisy@mit.edu/ Address: 50 Memorial Drive, Department of Economics, Cambrige MA 02142.

1 Introduction

Many countries try to promote group integration in areas such as housing, education, immigration, employment, etc. For example, in Chicago, the Gautreaux program (the predecessor of the Moving Towards Opportunity program) tried to encourage residential desegregation using housing vouchers (Jennan, 2000). In Singapore, the government instituted a set of controversial ethnic housing quotas to encourage residential integration. In education, immigration and employment, similar integration policies are constantly being debated. Economists have studied extensively the consequences of segregation (Cutler and Glaeser, 1997; Ananat, 2007). In this paper, I explore one potential cause of residential segregation–ingroup preferences.[1] This paper estimates ethnic ingroup preferences over residential neighborhoods (marginal willingness-to-pay to live with members of one's own ethnic group) using variation from the ethnic housing quotas policy in Singapore.

Understanding ingroup preferences is important for at least three reasons. First, segregation is a prominent pattern in residential neighborhoods of ethnically diverse countries but we know relatively little about what causes this pattern. Second, we need to understand whether residential segregation is driven by tastes for ethnic interactions or tastes for ethnic-specific amenities because the policy implications are different: If segregation is driven by tastes for interaction with one's own ethnicity, integration policies should focus on incentivizing residential location choices. If, instead, segregation patterns arise because ethnic group members share the same tastes for amenities (correlated tastes), then reducing the spatial clustering of ethnic-specific amenities can encourage integration with a potentially lower deadweight loss. Finally, my approach allows me to recover the elasticities of demand for ingroup interactions. These are crucial inputs to measure potential deadweight losses from regulating location choices, which will have to be weighed against potential benefits of

[1] Another cause of residential segregation is discrimination (Cutler, Glaeser and Vigdor, 1999). I discuss the implications of discrimination below.

integration.

To estimate ingroup preferences over residential locations, I build and estimate a discrete choice model à la Berry, Levinsohn, and Pakes (1995) in which individuals choose residential neighborhoods as a function of the proportion of ethnic ingroup members in a neighborhood. The empirical challenge is that the explanatory variable of interest, neighborhood ethnic proportions, could be correlated with unobserved ethnic-specific neighborhood quality (as measured by ethnic-specific amenities, for example). An amenity such as mosques could be more attractive to a specific ethnic group if a majority of that ethnic group is Muslim. In this case, just addressing omitted variables that are common across groups is insufficient if some neighborhoods have systematically more mosques (a measure of ethnic neighborhood quality) that will attract a specific ethnic group. Put simply, the challenge in this demand estimation exercise is not just that there are omitted variables but that the omitted variables are group-specific.

The ethnic housing quotas in Singapore provide a source of identification. The quotas were introduced in 1989 to encourage residential integration amongst the three major ethnic groups in Singapore – Chinese (77%), Malays (14%) and Indians (8%) (Singapore Census, 2000). The quotas were designed as upper bounds on block level and neighborhood level ethnic proportions. In practice, the Housing Development Board (HDB) did not want to evict owners in units that were in violation of the quotas. Therefore, these owners could remain in their homes. To this day, there exist units in violation of the quota. Any transactions that forced these blocks and neighborhoods farther above the upper bound, however, would be barred. For example, when Chinese quotas are binding, non-Chinese sellers cannot sell to Chinese buyers because this transaction increases the Chinese proportion farther above the Chinese quota. These ethnic-specific restrictions on transactions play a central role in my analysis.

The quotas help identification of ingroup preferences in two ways. First, the quotas imply that prices faced by different ethnic groups differ in equilibrium when the quota binds. The intuition is similar to that of price discrimination models. In such models, different groups can be charged different prices in equilibrium as long as there are no arbitrage opportunities. The quotas impose ethnic-specific restrictions on transactions that prevent some sellers from arbitraging away price differences across groups (when the Chinese quotas bind, non-Chinese sellers cannot sell to Chinese buyers). Therefore, prices can differ across groups within the same neighborhood when the quotas bind. Since prices are positively correlated with quality, I exploit information from group-specific prices to recover group-specific neighborhood quality.

The second reason the ethnic quotas help identification of ingroup preferences is that the policy rule is a step function of ethnic proportions. Units above the upper bound on ethnic proportions are constrained (the quota status is 1) while units below the upper bound are unconstrained (the quota status is 0). This step function is an ideal set up for a regression discontinuity design (Angrist and Lavy, 1999). The idea behind regression discontinuity is that the treatment effect of the quotas can be identified close enough to the discontinuity (the upper bound), assuming omitted variables do not change discontinuously at the upper bound.

I incorporate within neighborhood, across group variation in prices in a structural demand estimation framework à la Berry, Levinsohn and Pakes (1995), hereafter BLP. I use the method of simulated moments to estimate ingroup preferences. My instruments for the endogenous variables (price and ethnic proportions) are exogenous characteristics of nearby neighborhoods (following BLP, 1995 and Bayer et al., 2007) and historical data on early ethnic settlements. However, there is less a priori reason to believe that these variables are good instruments for group-specific prices. Here too, the quotas play a crucial role.

I use the step function to construct an instrument for group-specific prices. By preventing arbitrage, the quotas generate group-specific price variation, and hence, are good instruments for group-specific prices. But, actual quota status is correlated with neighborhood quality.[2] To construct my instrument for group prices, I first estimate ethnic proportions using the set of exogenous instruments. Then, I assign the quota status to be 1 if the predicted ethnic proportions are above the upper bound as defined by the quota policy, and 0 otherwise. Constructing instruments using this step function is akin to the regression discontinuity identification strategy: Exogenous characteristics used to estimate the quotas could affect prices but the effect will not be discontinuous at the upper bound of the quota.

To implement this analysis, I collected individual data on residential location choices by matching names from the 2005 and 2006 Singapore residential phonebooks. I hand matched more than 589,000 names to ethnicities. I also collected data on neighborhood characteristics, prices and quotas for 170 neighborhoods. My price data consists of weighted averages of group-specific prices.[3]

I find that all groups have strong preferences to live with at least some other members of their ethnic group. Moreover, the Chinese and the Indians have ingroup preferences that are inverted U-shaped. For example, the average Indian household living in a neighborhood with 5% Indians (the 10th percentile) is willing to substitute to a neighborhood that is 2.56km *further* from the closest subway station in exchange for living in a neighborhood with a 1 standard deviation (3%) increase in the proportion of Indians. However, after a neighborhood reaches 8% Indians and 43% Chinese respectively, Indians and Chinese would rather add a new neighbor from the other groups. These estimates are significant at the 5% level. Malays appear to have strong ingroup preferences, although the estimates for their

[2] Chinese quotas are more likely to bind in neighborhoods with a high Chinese quality.

[3] Not observing actual group-specific prices is unfortunate, but this is a limitation that is specific to my dataset. For the rest of the paper, I discuss the identification strategy assuming I observe group-specific prices. In the data section, I discuss how I estimate group-specific prices using a weighted average of prices and observed weights.

taste parameters are not significant.

This paper makes two contributions. First, the estimates on ingroup preferences have implications for research and policy. The finding that some households in Singapore exhibit ingroup preferences that are inverted U-shaped suggests that models that estimate ingroup preferences assuming monotonicity could be biased. In addition, policy makers could use these estimates on ingroup preferences to determine subsidies for housing voucher programs that aim to encourage desegregation.

The second contribution of this paper is that I combine policy induced variation akin to regression discontinuity design with structural methods. I embed the regression discontinuity identification strategy within a structural demand estimation framework by using the step function to construct exogenous quota variables that can be used as instruments for group-specific prices.

The backbone of the empirical analysis rests on the literature on discrete choice theory (see for example McFadden, 1974; Berry, Levinsohn, and Pakes, 1995, 2004). This is not the first paper in the literature that estimates multiple unobserved characteristics (see Das, Olley, and Pakes, 1993; Athey and Imbens, 2007). This paper demonstrates that in markets with price discrimination, variation in prices across groups can provide additional variation for identification.[4]

This paper builds on the rich literature on social interactions and residential segregation that dates back to Manski (1993), Schelling (1971) and more recently, Cutler, Glaeser and Vigdor (1999) and Card, Mas and Rothstein (2007). Bayer, Ferreira and McMillan (2007) and Bajari and Kahn (2005) estimate ethnic preferences using random coefficients. Both papers estimate a rich set of taste parameters using many demographic variables. My paper

[4]Some products studied in the demand estimation literature (such as cars and cereal) lack variation in product characteristics across markets which implies that estimation will likely require large datasets.

complements these two papers by identifying policy variation that I exploit to allow neighborhood quality to vary across groups, thereby identifying taste for ethnic interactions from correlated tastes for ethnic-specific neighborhood quality.

In the next section, I discuss the background of ethnic quotas in Singapore. Then, I describe the data (Section 3) and the results from the regression discontinuity analysis (Section 4). I then build a model of individual utility over residential locations with group-specific prices (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%). The Chinese have the highest median monthly income (S$2335), followed by the Indians (S$2167) and the Malays (S$1790) (Singapore Census, 2000).

Public housing is the most popular choice of housing in Singapore with 82% of the resident population living in public housing (HDB, 2006). The units are built and managed by the Housing Development Board (HDB). There are three ways Singapore residents can live in a HDB flat. They may apply through the primary allocation system for new HDB flats, they may purchase existing HDB flats in the resale market or they may rent. The rental market is negligible: 98% percent of the HDB flats are owner-occupied (HDB, 2006). This paper focuses on the resale market which is the relevant market for the ethnic quotas. 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 *flat*. A group of HDB flats constitute an HDB *block*. A group of HDB blocks make up a *neighborhood.* 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. 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 debate, 1989). The policy was announced in a parliamentary debate on February 16, 1989 and implemented starting March 1, 1989. The Policy is a set of quotas 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 national ethnic proportions. Block quotas are 3% above neighborhood quotas, allowing more flexibility at the block level (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. Therefore, these owners could remain in their homes. To this day, there exist blocks and neighborhoods above the quota.

The quotas are upper bounds on ethnic proportions to *prevent* HDB communities that are already segregated from becoming more segregated. Once a community hits the upper bound, transactions that make the community more segregated will be blocked. However, other transactions will be allowed. In particular, transactions involving buyers and sellers from the same ethnicity will always be allowed because this does not increase the ethnic proportion. As an example, Table 1 shows that the Chinese neighborhood quota is set at 84%. Once 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 barred by 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 group-specific restriction prevents arbitrage and thus allows prices to differ across groups in equilibrium.

3 Data

I use data covering 170 neighborhoods and 7 markets for resale transactions in the public housing market in Singapore, between April 2005 and March 2006. This dataset encompasses virtually all of Singapore.[5] A neighborhood is comparable to a US Census Tract (4500 households, on average). 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, with a mean of 24 and a standard deviation of 10.

Prices

I collected data on 25,182 actual transaction prices from the HDB website between March 2005 and April 2006. I aggregate these monthly transaction level prices into average, annual neighborhood prices. These are actual purchase prices, not rents. The rental market is very small (<5%). Unfortunately, I do not observe the ethnicity of the buyers and sellers. Hence, for each neighborhood j, I only observe a weighted average of the actual group-specific prices. I discuss how I estimate group prices in the following sub-section.

Neighborhood characteristics

Neighborhood characteristics include ethnic proportions (of the initial *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

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

structure of Chinese, Malay and Indian names. 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 Census, 2000); Indian names are matched according to popular first and last names. I collected the remaining neighborhood attributes from online street directories, the HDB website and a non-public dataset purchased from HDB. See the Data Appendix for definitions of these variables and their sources.

Choice data

I collected data on individual residential location choices by matching names from the 2005 and 2006 Singapore residential phonebooks.[6] I define *movers*, as individuals whose postal code in 2005 did not match with their postal code in 2006. A postal code uniquely identifies an HDB block. In cases with multiple names, I match movers to their choices randomly.[7] There are 16,092 movers.[8] Using the ethnicity of movers, I calculated ethnic shares. For example, the percent of the *flow* of Chinese buyers who chose a neighborhood. 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 ethnic share for each ethnic group.[9] Note that ethnic proportions describe the ethnic distribution of the *stock* of residents while ethnic shares refer to the *flow* of movers. In my analysis, I use *ethnic shares* as a proxy for demand and *ethnic proportions* as the primary neighborhood characteristic. The assumption is that the flow of movers is so small

[6] In principle, one could use just 1 phonebook to analyze the residential location choices of all residents in Singapore (using the choice data of both movers and stayers). I focus on using the choices of movers instead of all residents because choices of movers are actual location choices involving a relatively high purchase price. Also, I use actual transaction prices instead of imputed rents. These prices should reflect the choice of movers more than the choice of residents.

[7] Multiple names will not affect my analysis because only the ethnicity of the buyer matters in the calculation of the ethnic shares. For my purposes, ethnicity depends only on names.

[8] The actual number of transactions (25,182) is higher than the number of movers (16,092). This could be due to spelling errors between names in the 2005 phonebook and the 2006 phonebook. I assume that the spelling errors are random such that the choice data is not a selected sample.

[9] 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.

that the ethnic proportion of the stock of residents is essentially constant within a year.

Early ethnic settlements

I use data on early 19th century ethnic settlements in Singapore to instrument for ethnic proportions. The Jackson Plan, which was commissioned in the 1820s, specified that the west of the Singapore River should be reserved for Chinese and Indian communities while the east of the river was reserved for the Malay communities (see Figure 1a). I argue that the historical assignment of ethnic settlements to opposite banks of the river is an important instrument for current ethnic proportions.

Figure 1a shows the map of early 19th century Singapore according to the Jackson Plan (Crawfurd, 1828). The Jackson Plan was formulated by a committee led by Lieutenant Philip Jackson to set up ethnic functional subdivisions within the growing port city. The map illustrates early settlements in Singapore around the Singapore River. 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. Figure 1b shows the distribution of existing quota-constrained neighborhoods. Malay-constrained neighborhoods are primarily in the east while the Chinese neighborhoods are in Central, South and West Singapore and the Indian neighborhoods are in Central and North Singapore. This suggests that the Malay neighborhoods expanded to the east of the river. I define a dummy variable that is 1 when the entire area of the neighborhood is in the east of the eastern end of early Malay settlements, and 0 otherwise.

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 buyers 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.

3.1 Estimated Prices

I estimate group-specific prices for each neighborhood using the weighted average of prices at the block level as well as the observed ethnic weights (from the data on movers from the phonebook). This estimation procedure essentially solves a system of equations where the variables of interest are the block level ethnic weights and the unknowns are the neighborhood level group prices. The assumption is that within a neighborhood, controlling for block level characteristics, prices vary across blocks only because the ethnic weights vary (a block with 20% Chinese buyers and 80% Malay buyers will have a different price from a similar block with 80% Chinese buyers and 20% Malay buyers).

I estimate the following equation for the average price of block b in neighborhood j:

$$\ln \bar{P}_{bj} = \pi^C I_j * w_{bj}^C + \pi^M I_j * w_{bj}^M + B_{bj} * I_j \gamma + I_j + \upsilon_{bj} \quad (1)$$

where π^C and π^M are the unknowns I want to solve for, w_{bj}^C and w_{bj}^M are the Chinese and Malay buyer weights, B_{bj} is a set of block-level characteristics (the block quotas, the number of 1-room flats, 2-room flats etc.), I_j is a neighborhood dummy and the Indians are the omitted group. Notice that the neighborhood dummy is interacted with each explanatory variable. Essentially, for each neighborhood, I am solving a system of N_j equations and 3 unknowns for each neighborhood j where N_j is the number of blocks in neighborhood j and the 3 unknowns are the Chinese, Malay and Indian prices for neighborhood j. Using estimates from equation (1), I substitute $w_{bj}^C = 1$, $w_{bj}^M = 0$ to obtain the Chinese price, $\hat{P}_j^C$ and likewise for the Malay and Indian prices.[10]

[10] I calculate $\hat{P}_j^C = \exp\left(\ln \overline{P}\left(\hat{\pi}, \hat{\gamma}; w_j^C = 1\right)\right)$. I use log of prices to ensure that the price estimates are positive. In ongoing work, I estimate group prices using average prices (instead of log prices) but constraining price estimates to be positive.

4 Regression Discontinuity

In this section, I use results from a regression discontinuity framework to analyze average and estimated group prices around the quota (Angrist and Lavy 1999). I argue that findings on average prices offer suggestive evidence that prices do differ across groups in equilibrium. If prices differed across groups, then this within neighborhood, across group variation in prices could help recover group-specific neighborhood quality.

The regression discontinuity method relies on the step function. Recall, units are constrained (the quota status is 1) above the upper bound on ethnic proportions and units are unconstrained (the quota status is 0) below the upper bound. The idea behind regression discontinuity is that the treatment effect of the quotas can be identified close enough to the discontinuity (the upper bound), assuming omitted variables are similar right above and right below the upper bound. For example, this assumes that the number of mosques (a measure of unobserved Malay neighborhood quality) could be different above and below the upper bound but the number of mosques does not change discontinuously at the Malay neighborhood quota (22% Malays). If this assumption holds, comparisons of units right above and right below the quota offers one way to address the omitted variable problem. In practice, when the quotas started, the Housing Development Board (HDB) did not want to evict households from constrained areas. Hence, to this day, I still observe households above the quotas.

4.1 Regression discontinuity analysis using average prices

I estimate the following equations

$$\ln \bar{P}_{bjit} = \alpha + \gamma QC_{bjit} + \sum_{k=1}^{4} \phi_k pctC_{bji}^k + \varepsilon_{bjit} \quad (2)$$

$$\ln \bar{P}_{bjit} = \alpha + \gamma QC_{bjit} + \sum_{k=1}^{4} \phi_k pctC_{bji}^k + B_{bji}\beta + \tau_t + \omega_i + \varepsilon_{bjit} \quad (3)$$

where $\ln \bar{P}_{bjit}$ is the log of the average 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, $pctC_{bji}^k$ are kth order polynomials of the percent of Chinese; B represents other observable characteristics 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 this equation for units that are 10% above and below the Chinese quotas. Similarly, I repeat the analysis for the Malay and Indian quotas. The coefficient of interest is γ, which summarizes the effect of the quota on average prices, at the upper bound. I report results from estimating the discontinuity in average prices at the block quotas.

I assume that the policy rule is perfectly enforced. Since these are public housing units, all resale transactions need to be approved by the HDB. Part of the approval process involves checking that buyers and sellers of a transaction do not violate the ethnic quota rule. It is possible that households sort around the discontinuity. In this case, if households have incentives to undo the discontinuity, then any discontinuity in prices that I estimate would be a lower bound of the actual discontinuity caused by the quotas.

4.2 Discussion on group-specific prices

Equilibrium prices differ across groups when the quotas bind because the quotas prevent arbitrage. In this section, I will argue that results from the regression discontinuity analysis using average prices are suggestive that the quotas prevent arbitrage. The idea is that if sellers can arbitrage, then, they cannot be worse off so that average prices cannot be lower in the presence of arbitrage. For example, arbitrage opportunities exist when the Chinese quota binds because non-Chinese sellers cannot sell to Chinese buyers, who are likely to have a higher willingness-to-pay (WTP) for units in Chinese-constrained areas.[11] Sellers could engage in inter-temporal arbitrage by waiting for the quotas to become unconstrained, and then sell to the high-WTP buyers. Therefore, to show that prices differed across groups when the quota was binding, I need to show that buyers were willing to sell to the low-WTP buyers instead of engaging in inter-temporal arbitrage.[12]

If I observed group-specific prices, then, I could test whether prices differed significantly across groups for observationally identical units in quota-constrained areas. In particular, I could test if Chinese buyers paid a higher price than Malay and Indian buyers when the Chinese quota binds (where the Chinese are likely the high-WTP group) but a lower price when the Malay and Indian quotas bind (where the Chinese are likely the low-WTP group). Table 2 provides a summary of the relationship between group prices and quotas. Finding that Chinese buyers paid a lower price than Malay and Indian buyers in Malay- and Indian-constrained neighborhoods suggests that sellers in these neighborhoods are willing to accept the lower price of the Chinese buyers instead of engaging in arbitrage. If sellers instead engaged in arbitrage, then I would not observe transactions involving these lower-WTP

[11] The link between group-specific prices and quotas is premised on two correlations: (i) Chinese prices are high when the Chinese neighborhood quality is high; (ii) Chinese quotas are more likely to bind in neighborhoods with high Chinese quality because these neighborhoods attract relatively more Chinese. These two correlations imply that Chinese prices are likely to be positively correlated with Chinese quotas. In addition, if we assume the Chinese, Malay and Indian quality are not perfectly correlated, then, the prices that Chinese buyers are willing-to-pay would be higher than Malay and Indian prices in Chinese-constrained neighborhoods, and lower than Malay and Indian prices in Malay- and Indian-constrained neighborhoods.

[12] One reason sellers do not arbitrage is waiting costs. If waiting costs were high, then, it would be costly to wait for units to become unconstrained, then sell to the high-WTP buyers. Another reason is that arbitrage is risky if there is lack of coordination. Suppose the Malay quotas were binding at 22% Malays, and, the non-Malay sellers were waiting for the unit to become unconstrained. Then, once the unit becomes unconstrained, there could be a sudden excess supply of units, which may exert a downward force on prices.

Chinese buyers.

What happens to the weighted average of prices is less clear. I argue that the Chinese are an overwhelming majority of the buyers such that the Chinese price closely reflects the average price. Since Chinese prices are *higher* above the Chinese quota but *lower* when Malay and Indian quotas bind, a weighted average of prices is likely to be higher when the Chinese quota binds and lower when the Malay and Indian quotas bind.[13] Admittedly, these are merely suggestive evidence that prices differed across groups since the predictions rely on assumptions that cannot be proven without actual group-specific prices.

The findings are at least consistent with price variation across by groups — average prices are higher when the Chinese quota binds and lower when the Malay and Indian quotas bind. I report results from the regression analysis 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 characteristics (such as age, number of 1-room flats, number of 2-room flats etc.) and month and town fixed effects (5th column).

Without controlling for building characteristics and fixed effects, average prices are 10% to 11% higher when the Chinese quota binds, but 4% and 3% lower when the Malay and Indian quotas bind. All estimates are significant at the 1% level. In Figure 2, I plot the predicted log prices using the estimates from the regression with polynomials of the ethnic proportions up to the 4th order (from columns 4, 9 and 14 in Table 4).

[13]Since the average price is a weighted average of prices of high-WTP buyers and low-WTP buyers, I am assuming that in the case of binding Chinese quotas, the large weight on the high-WTP Chinese buyers are more than enough to compensate the fact that the low-WTP buyers pay a lower price, so that the net effect should still be an increase in average prices.

These findings are robust to including 1st to 4th order polynomials of the ethnic proportions and for samples restricted to 5% above and below the quotas. Once I add other controls, the estimates for the Indian quotas are not significant. This could be because Indians are such a minority that almost 95% of the neighborhoods fall within 10% of the Indian quota. Hence, restricting the analysis close to the Indian quotas would essentially be an OLS analysis, since almost all units fall within the 10% window.

4.3 Regression discontinuity analysis with estimated group prices

To analyze the estimated group prices (discussed in Section 3.1), notice that knowing the prices that buyers from each ethnic group paid implies knowing the ethnicity of the buyer. In the following specification, I assigned dummy variables for buyer ethnicity and stacked the estimated group prices in a similar regression discontinuity set up

$$\begin{aligned}\ln \hat{P}_{bjit} &= \alpha + \rho_1 QC_{bjit} + \rho_2 QC_{bjit} * QI_{bjit} + \eta_1 buyM_{bjit} + \eta_2 buyI_{bjit} \\ &\quad + \gamma_1 QC_{bjit} * buyM_{bjit} + \gamma_2 QC_{bjit} * buyI_{bjit} + \gamma_3 QC_{bjit} * QI_{bjit} * buyM_{bjit} \\ &\quad + \sum_{k=1}^{4} \phi_k pctC_{bji}^k + B_{bji}\beta + \tau_t + \omega_i + \varepsilon_{bjit} \qquad (4)\end{aligned}$$

where now, QC_{bjit} is a dummy that is 1 when only the Chinese quota in block b in neighborhood j, town i and month t is binding; $QC_{bjit} * QI_{bjit}$ is a dummy when both the Chinese and Indian quotas are binding, $buyM_{bjit}$ is a dummy variable that is 1 when the buyer is Malay, $buyI_{bjit}$ is a dummy for an Indian buyer, $pctC_{bji}^k$ is the percent of Chinese in the block, B is a matrix of observable block characteristics, τ_t and ω_i are the month and town fixed effects, respectively.[14] 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 block quotas. Each time interacting the Malay (Indian) quota, with a dummy for non-Malay

[14]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.

(non-Indian) buyers.

The key coefficients of interest here are the $\gamma's$ and the $\rho's$. The idea is to test if Chinese buyers paid a higher price for Chinese-constrained blocks ($\rho's > 0$) and non-Chinese buyers paid a lower price for Chinese-constrained blocks ($\gamma's \leq 0$). This would indicate that group prices differed for the same neighborhood. Table 5 shows the results from the estimation. The 3 columns correspond to the regression for the Chinese, Malay and Indian quotas.

Column 1 shows that Chinese buyers paid 6% more in Chinese-constrained blocks, Chinese and Indian buyers paid 20% more in blocks where both the Chinese and Indian quotas were binding (these estimates of $\rho_1$and ρ_2 are significant at the 1% level). Non-Chinese buyers did not seem to pay a higher price for Chinese-constrained blocks. Moreover, Malay buyers paid 6% less for Chinese- and Indian-constrained blocks (this estimate of γ_3 is significant at the 5% level). So, the signs of the coefficients for the estimated Chinese prices are as expected. The Malay quota has a similar effect on estimated group prices except the coefficients are less significant and I find that blocks where both the Malay and Indian quotas bind, Malay and Indian buyers paid a significantly lower price. The results for the estimation close to the Indian quota (column 3) do not follow the same pattern. Indian buyers paid 4% less (the estimate is significant at the 1% level) when I expected Indian buyers to pay more for Indian-constrained neighborhoods. These findings could again be attributed to the earlier finding that the discontinuity generated by the Indian quotas disappears after controlling for unit characteristics and fixed effects.[15]

At this point, one could take the regression discontinuity analysis one step further and use the hedonic method to estimate ethnic ingroup preferences (Rosen, 1974; Chay and Greenstone, 2004). The idea is to address omitted variable bias in demand estimation by

[15] In the estimation of the BLP model, since I essentially estimate a separate BLP model for each group (ie. the Indian prices only enter the BLP model for the Indians), this should not affect estimates for the BLP models of the Chinese and the Malays.

examining estimates of the hedonic price function right above and below the discontinuity. This approach comes with two limitations. First, using regression discontinuity design to estimate preferences implies that the estimated taste parameters are only valid for households who are at the discontinuity. There are reasons such as sorting that may suggest that these households are different. A second limitation is that choosing residential neighborhoods is essentially a discrete demand problem while the hedonic method applies to cases with continuous demand. In ongoing work, I explore the possibilities of estimating ethnic preferences using the hedonic model versus the discrete choice model (see Bayer et al., 2007 for a comparison of estimates from hedonic methods and structural methods).

5 Utility Specification

To recover ingroup preferences of households away from the discontinuity, I make some assumptions on the functional form of the utility and distributional assumptions on the heterogeneity of individuals. The goal of the structural estimation is to recover individual preferences from aggregate data using the method of simulated moments. I begin with a random coefficients model of individual utility for residential neighborhoods that is then aggregated to obtain market-level demands to be matched with aggregate sample moments.

Suppose we observe 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

$$U_{ijm}^G = X_{jm}^G \beta_i^G - \alpha_i^G P_{jm}^G + \xi_{jm}^G + \varepsilon_{ijm}^G \quad (5)$$

for j=1, ..., J_m $\forall m$, where X_{jm}^G is a K-dimensional (row) vector of observed neighborhood characteristics (including ethnic proportions), P_{jm}^G is the price that a buyer of group G has to pay for a unit in neighborhood j in market m, ξ_{jm}^G is the group-specific preference for the unobserved neighborhood attribute and ε_{ijm}^{G} represents mean-zero, idiosyncratic individual preferences for a buyer of group G, assumed to be independent of neighborhood characteristics, prices and of each other. Note that prices are now indexed by G. This is a consequence of the quotas that I exploit to recover group-specific neighborhood quality $\left(\xi_{jm}^{G}\right)$. To keep notation simple, I will drop the market subscript from here on.

We can write the buyers' taste parameters as a mean component and an individual-specific deviation from the mean

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

where $\nu_{i1}^{G}, ..., \nu_{iK}^{G}$ is individual i's unobserved taste for characteristic K, drawn independently (for each individual in each group) from a standard normal distribution and ν_{iP}^{G} is drawn from a log normal distribution. $\sum$ is a (K+1)x(K+1) dimensional scaling matrix whose diagonal elements are denoted by σ_k and σ_P. Note that I assume mean preferences vary by group but the standard deviation does not ($\sum$ is not indexed by G).

To estimate ethnic preferences, I include the ethnic proportions of the initial stock of owners in neighborhood j. Specifically, the primary neighborhood characteristics of interest are $percent\ Ingroup, percent\ Ingroup^2$ which are the initial percent of ingroups in neighborhood j and its squared. The parameters that represent ethnic preferences are $\bar{\beta}_{percent\ Ingroup}^{G}$ and $\bar{\beta}_{percent\ Ingroup^2}^{G}$. For example, $\left(\bar{\beta}_{percent\ Ingroup}^{C} + 2\bar{\beta}_{percent\ Ingroup^2}^{C}\overline{percent\ Chinese}\right)*0.01$ is the average Chinese buyer's marginal utility to live in a neighborhood with 1% more ingroup members (Chinese) relative to a neighborhood with 1% more outgroup neighbors, evaluated for the households living in the average Chinese neighborhood. Allowing neighborhood quality (ξ) to vary by group allows the interpretation of these parameters as taste for ethnic ingroup interactions, that is separate from the taste for ethnic-specific neighborhood quality. Other observed neighborhood attributes included in X_{jm}^{G} are school quality, access to public transportation, average number of rooms and average age of HDB blocks in neighborhood j.

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

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

Since market shares depend only on differences in utilities, the actual estimation algorithm subtracts U_{i0}^G from U_{ij}^G such that utility is defined relative to the outside good.

Substituting (6) into (5) and grouping individual-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

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

and an individual-specific deviation from that mean

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

There are two features in (8) that will be relevant for estimation. First, utility from neighborhood quality only depends on the quality of that neighborhood alone.[16] This implies that using exogenous attributes of nearby neighborhoods as instruments will satisfy the exclusion restriction. Second, the omitted variable (ξ) enters the mean utility (δ) linearly. This allows the estimation of (8) using linear instrumental variable techniques (Berry and Pakes, 2007).

The first two terms in (9) are the interaction between consumer tastes and neighborhood characteristics that determine substitution patterns in discrete choice models (McFadden

[16] This excludes utility specifications, for example, where buyers have higher utility if their neighborhood is better than adjacent neighborhoods.

et al., 1977; Hausman and Wise, 1978; BLP, 1995). As heterogeneity (σ) in the unobserved tastes for observed product characteristics increases, neighborhoods that are similar (in characteristics space) become better substitutes.

A potential weakness of the specification is that equation (5) has no income. The lack of income data is not particularly helpful for models of residential location choices. In ongoing work, I explore the possibility of proxying for the income of buyers using the income distribution of the entire population.

To return to the model of individual utility, market-level aggregate consumer behavior is obtained by aggregating the choices implied by the individual utility model over the distribution of consumer attributes. Let F_{μ^G} be the population distribution function of individual-level attributes for group G. The fraction of households of group G that choose neighborhood j (aggregate demand) is obtained by integrating over the set of individual attributes that imply a preference for neighborhood j. Let the group G share for neighborhood j be

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

where

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

and $\theta^G = \{\bar{\beta}^G, \bar{\alpha}^G, \sigma\}$.

Following the literature, I assume that the idiosyncratic errors, the ε_{ij}^G, have an independently and identically distributed Type I extreme value distribution. This assumption yields the Logit form for the model's choice probabilities. Letting y_i^G denote the choice of buyer i of group G,

$$\Pr(y_i^G = j|\boldsymbol{\nu}_i^G, \theta^G, x^G, P^G) = \frac{exp(\delta_j^G + \sum_k \sigma_k x_{jk}^G \nu_{ik} + \sigma_P P_j^G \nu_{iP}^G)}{1 + \sum_{j'} exp(\delta_{j'}^G + \sum_k \sigma_k x_{j'k}^G \nu_{ik} + \sigma_P P_{j'}^G \nu_{iP}^G)} \quad (12)$$

Note that the omitted variable (ξ), enters the mean utility term (δ), linearly but it enters demand non-linearly. This complicates the use of standard non-linear instrumental variables method to address omitted variable bias. To address the problem of non-linear omitted variables (ξ), Berry (1994) shows that we can invert the market share function to recover the choice specific constant (δ). Since the choice specific constant is a linear function of quality, once we have recovered δ, we can use standard instrumental variables methods to estimate the mean utility equation (8) by finding instruments for endogenous neighborhood attributes that are orthogonal to quality (ξ) .

6 Estimation

Using the model of individual utility above, I recover the taste parameters, $\left\{\bar{\beta}^G, \bar{\alpha}^G, \sigma\right\}$ by matching aggregate moments predicted from the model to sample moments using the Method of Simulated Moments (MSM).

6.1 Method of Simulated Moments

The following moment condition is assumed to hold at the true parameter value, $\theta_0 \in R^p$:

$$E[g(\theta_0)] \equiv E[\xi(\theta_0)|Z] = 0 \quad (13)$$

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

For each ethnic group G, I first guess values for θ^G which I use to calculate the share function using equations (10) and (12) . I use the contraction mapping provided in Berry (1994) to find the value of δ^G that makes the observed ethnic shares, s^G, equal to the predicted shares defined in equation (10). Notice that the integral in the share function, no longer has a closed form solution. I simulate the integral by drawing R=10,000 $\nu_i^{G\prime}s$ independently for each group G and calculating the Logit form in equation (12) ,which is then aggregated out to obtain the market level shares.. After recovering the mean utility, δ^G, by inverting the ethnic share function, I calculate ξ^G using equation (8) and the estimated prices. Now that we have an estimate of the (structural) error, ξ^G,we are ready to form the moments.

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

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

An MSM estimator, $\widehat{\theta}$, minimizes a weighted quadratic form in $\sum_j^J g_j(\hat{\theta})$:

$$\theta = \arg\min_{\theta \in \Theta} \frac{1}{J} \left[\sum_j^J \hat{g}_j(\theta) \right]' \Omega_J \left[\sum_j^J \hat{g}_j(\theta) \right] \tag{15}$$

where Ω_J is an LxL positive, semi-definite weighting matrix. Assume $\Omega_J \xrightarrow{p} \Omega_0$,an LxL positive definite matrix. Define the LxP matrix $G_0 = E[\nabla_{\theta'} g(\theta_0)]$. Under some mild regularity conditions, Pakes and Pollard (1989) and McFadden (1989) show that:

$$\sqrt{J}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N(0, \left(1 + R^{-1}\right) * A_0^{-1} B_0 A_0^{-1}) \tag{16}$$

where R is the number of simulations, $A_0 \equiv G_0' \Omega_0 G_0$, $B_0 \equiv G_0' \Omega_0 \Lambda_0 \Omega_0 G_0$ and $\Lambda_0 = E\left[g\left(\theta_0\right) g\left(\theta_0\right)'\right] = Var\left[g\left(\theta_0\right)\right]$. If a consistent estimate of Λ_0^{-1} is used as the weighting matrix, the MSM estimator, $\hat{\theta}$, is asymptotically efficient, with its asymptotic variance being

$$Avar(\hat{\theta}) = (1 + R^{-1}) * \left(G_0' \Lambda_0^{-1} G_0\right)^{-1} / J.$$

To account for the error from using estimated prices instead of actual prices, I follow the discussion in Newey (1984) on sequential estimators and method of moments. We can think of the exercise above in the context of GMM where there are 2 moments and parameters are estimated sequentially. First, we estimate prices with parameters π and γ (the first moments are $g^1(\pi, \gamma)$). Then, using these parameters as inputs, the second moments are just the standard BLP moments, $g^2(\theta, \hat{\pi}, \hat{\gamma})$. To calculate the standard errors, I stack the BLP moments and the moments from the estimation of prices and calculate standard errors using equation (16).

6.2 Instruments

In this section, I discuss the instruments for ethnic proportions and price. Note that the exclusion restriction holds by definition for most of the instruments discussed below because utility from neighborhood j only depends on the characteristics of neighborhood j that are in the utility function.

Ethnic proportions

I instrument for ethnic proportions using attributes of nearby neighborhoods (as in Bayer et al., 2007) as well as historical settlement data. I use the sum of the exogenous characteristics of rival products (in my case, I use attributes of neighborhoods in 1-3km rings, 3-5km rings and 5-7km rings). The attributes include average number of rooms, average age of building, school quality and average distance to the closest subway station. I chose 1km as the cutoff because the neighborhoods would be far enough to avoid spatial correlation with own neighborhood. I chose 2km widths so that all neighborhoods would have at least one nearby neighborhood within the ring.[17] The idea of using attributes of nearby neighborhoods is that Chinese, Malays and Indians have different preferences for neighborhood attributes, perhaps due to demographics such as family sizes. 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 (HDB, 2000). The thought experiment involves 2 similar neighborhoods where one is surrounded by neighborhoods with many big units and the other is surrounded by neighborhoods with few big units. Malay households would tend to sort into the neighborhood surrounded by neighborhoods with few big units since many of these large Malay households will prefer big units.

In addition to using attributes of nearby neighborhoods to instrument for ethnic proportions, I also use a dummy variable on whether units are to the east of the early Malay settlement. The idea is that exogenous assignment of Malay settlements to the east of the Singapore River imply that subsequent Malay neighborhoods were more likely to develop on the east side of the Singapore River.

Group-specific prices

I instrument for group-specific prices using exogenous characteristics of nearby neighborhoods and estimated quotas. Specifically, I follow BLP (1995) and use the sum of the exogenous characteristics of rival neighborhoods. Attributes of nearby neighborhoods are valid instruments for prices if markets are competitive.

While attributes of nearby neighborhoods could instrument for common prices, it is hard to think of them as good instruments for group-specific prices. One would expect ethnic quotas to be highly correlated with group-specific prices (Chinese prices are high when the Chinese quota binds). However, actual quotas are not valid instruments because they

[17] 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.

are positively correlated with the structural error term, ξ_j^G in equation (8). Therefore, I use estimated quotas where the quotas are estimated using only exogenous variables. This ensures that variation from the estimated quotas only derives from exogenous variation in neighborhood characteristics.

To estimate quotas, I exploit the step function in the policy rule. First, I estimate the block (neighborhood) level ethnic proportions using own block (neighborhood) characteristics and nearby neighborhood characteristics. Then, I assign the estimated block (neighborhood) quotas to be 1 if the estimated block (neighborhood) ethnic proportions are above the block (neighborhood) level quotas and 0 otherwise. The estimation equation for block and neighborhood proportions are

$$pctG_{bj} = \gamma_0 + X_{bj}^{ex}\gamma_1 + \gamma_2 East_j + Z_j\gamma_3 + u_{bj} \quad (17)$$

$$pctG_j = \rho_0 + X_j^{ex}\rho_1 + \rho_2 East_j + Z_j\rho_6 + v_j \quad (18)$$

where G=(C)hinese, (M)alays and (I)ndians, b indexes blocks and j indexes neighborhoods. The variable, $pctG$ is the percent of residents from group G, X^{ex} is the set of exogenous observed characteristics, $East$ is a dummy for whether the neighborhood is to the east of the early Malay settlements, Z is the set of exogenous characteristics of nearby neighborhoods.

Constructing instruments using this step function is akin to the regression discontinuity identification strategy. Quotas affect prices according to the step function (above the upper bound, prices differ across groups). The instruments, X^{ex}, Z and $East$, could affect prices but the effect is not discontinuous at the upper bound. Therefore, even though the quotas were estimated by projecting neighborhood proportions onto the space of X^{ex}, Z and $East$, the estimated quotas should still have power to predict group prices. To test this, I estimate the following equations

$$QG_j = \chi\widehat{QG}_j + \upsilon_j \quad (19)$$

$$QG_j = \phi_0 + X_j^{ex}\phi_1 + \phi_2 East_j + Z_j\phi_3 + \phi_4\widehat{QG}_j + \omega_j \tag{20}$$

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 in a month.[18] Also, I regress actual quotas on the estimated quotas,controlling for the full set of exogenous variables. If quotas have power above and beyond the exogenous characteristics 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 the dependent variable is not group-specific prices (what the quotas instrument for) because I do not observe group-specific prices.

7 Results

Table 6 reports results from regressions of actual quotas on estimated quotas. The first 3 columns do not control for other exogenous instruments that I used to predict ethnic proportions. The coefficients on the estimated quotas are all positive and significant. After controlling for exogenous instruments (columns 4-6), Chinese and Malay estimated quotas remain significant but Indian quotas are negative and not significant. This could be because the discontinuity from the Indian quotas essentially disappears after controlling for fixed effects and unit characteristics as shown in the regression discontinuity results (the last column in Table 4). The idea of using estimated quotas to instrument for prices relies on the step function of the quotas. Since the step function from the Indian quotas disappears after adding controls, it is not surprising that the coefficient on the estimated Indian quota is insignificant after adding controls.

[18] This is a percentage instead of a dummy because there are block and neighborhood quotas. This number is 1 when the neighborhood quota binds in a month (all blocks are constrained) and less than 1 when some blocks are hitting the block quota. For the dependent variable, I use data from March 2005, the earliest month in my dataset that is relevant for this group of movers.

Table 7 reports results from estimating a Logit model with OLS (columns 1-3) and IV (columns 4-6) where the dependent variables are the log of the ethnic shares, $\ln(s_{jm}^{G})$ subtracted by the log of the ethnic share for the outside good, $\ln\left(s_{0m}^{G}\right)$ and G indexes for the (C)hinese, the (M)alays and the (I)ndians. I use estimated group-specific prices in this regression. Most of the coefficients are of the right sign in OLS but this does not mean estimates are not biased due to omitted variables.

Table 8 reports results from estimating the random coefficients model using group-specific prices for 170 neighborhoods. 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.

Interpreting the magnitudes of the taste parameters, living in a neighborhood where the average building is 10 years older is as bad as living 2.3 km further away from the subway station. These parameters are significant at the 1% and 5% level respectively. On average, households prefer the outside good (not moving) since the marginal utility of the constant term is significantly negative. There is substantial heterogeneity in the taste for rooms. The coefficient on price enters negatively but it is not significant. Most coefficients enter with the right sign except the marginal utility for rooms which is negative and significant.

All groups want to live with at least some members of their own group. The Chinese and Indians have ingroup preferences that are inverted U-shaped such that in neighborhoods where there are enough members of their ingroup, Chinese and Indians prefer neighbors from other groups, on the margin (the parameters are significantly positive for $percent\ Ingroup$ but significantly negative for $percent\ Ingroup^2$). The estimated marginal utilities for Malays are positive for both terms, albeit not significant.

Using the estimates on marginal utilities from Table 8, I calculate the marginal rates of substitution (MRS) between *percent Ingroup* and distance to the subway station as well as the MRS between *percent Ingroup* and the average age of a building (where older is considered worse).Because of the quadratic term on *percent Ingroup*, the MRS changes when the ethnic proportions in a neighborhood change. The MRS is calculated as the marginal utility for a neighborhood with a 1 standard deviation increase in the percent of ingroup members divided by the marginal utility for a neighborhood that is 1km closer to the closest subway station.[19] I calculate the MRS with respect to age of buildings in a similar way.

Figure 3 plots the MRS with respect to the distance to the subway station as a function of the neighborhood ethnic proportion. Each point on the line is an MRS. The plots show that Chinese and Indian ingroup preferences are inverted U-shaped because the marginal utilities for ingroup neighbors are positive below 43% Chinese and 8% Indians respectively but are negative for neighborhoods with a higher concentration of ingroup members.

Table 9 shows the MRS's evaluated at neighborhoods with the mean, the 10th percentile, and the 90th percentile of ethnic proportions. The MRS for the Indians (the 2nd and 3rd panels in the last column) suggests that Indians have ingroup preferences that are inverted U-shaped. The average Indian household living in a neighborhood with 5% Indians (the 10th percentile) is willing to substitute to a neighborhood that is 2.56km *further* from the closest subway station in exchange for a neighborhood with a 1 standard deviation (3%) increase in the proportion of Indians. This distance is relatively far considering the average household reported that the maximum acceptable walking distance to a subway station is 0.53km (HDB, 2000).[20] On the contrary, the average Indian household living in a neighborhood

[19]Since distance to subway and age of building are bad characteristics, I used the negative of their marginal utilities in the MRS calculation so that a positive MRS reflects ingroup preferences and a negative MRS reflects outgroup preferences for the marginal neighbor. For example, the MRS with respect to distance to the subway is calculated as $\frac{(\bar{\beta}^{C}_{pctIngroup}+2\bar{\beta}^{C}_{pctIngroup^2}\overline{pctChinese})*0.01}{-\bar{\beta}_{subway}}$

[20]This is based on a survey conducted by the HDB to study the profile of public housing residents in Singapore. The distance is calculated by the surveyors in HDB based on a rate of conversion of 10 minutes to 500m of walking distance. There are 2 types of subway stations; the Mass Rapid Transit (MRT) is the primary subway while the Light Rapid Transit (LRT) is relatively new and limited. The number given above

with 8% Indians (the mean) is willing to substitute to a neighborhood that is 0.21km *closer* to the subway station for the same 1 standard deviation increase in the percent of ingroup neighbors.

Although the Chinese have ingroup preferences that are inverted U-shaped as shown in Figure 3, the MRS's in Table 9 for the Chinese (the first column) are all negative. This is because the estimates suggests that the Chinese have strong ingroup preferences for at least 43% of Chinese but once a neighborhood has more than 43% Chinese, the average Chinese household desires a neighbor from the outgroups. Since all neighborhoods in my sample have more than 43% Chinese (the minimum is 61%), the MRS for the Chinese evaluated at all neighborhoods in the sample are negative. The Malays who live in a neighborhood with an average percent of Malays (13%), are willing to substitute to a neighborhood that is 2.1km further away from subway stations to live in a neighborhood with a 1 standard deviation (7%) increase in the percent of Malay neighbors, although the estimates for the Malays are not significant.

The finding that Chinese exhibit positive outgroup preferences at the margin is suggestive evidence against ethnic discrimination. The concern with ethnic discrimination is that Malay enclaves may form even when Malays have no ingroup preferences, simply because the Chinese are discriminating against Malays and forcing them into Malay enclaves. This could mean that my estimate of Malay ingroup preferences is an overestimate. However, to the extent that the Chinese are discriminating against outgroups, the Chinese should have strong negative outgroup preferences, which is not what I find.

Policy makers could use these estimates on ingroup preferences to design integration policies such as housing voucher programs. The finding that some households exhibit ingroup preference that are inverted-U shaped suggests that the minimum required value to incen-

refers to the distance to an MRT station.

tivize households to move depends on the group and the existing equilibrium distribution of neighborhood ethnic proportions. For example, the estimates suggest that the average Indian household living in a neighborhood with 8% Indians would require lower subsidies to move to a neighborhood with one standard deviation fewer Indians compared to the average Indian household living in a neighborhood with 5% Indians. Estimates of Malay preferences (although not significant) suggest that the policy was successful at preventing the formation of Malay enclaves such that Malays have positive marginal utility at the margin, even for Malays living in neighborhoods with almost a maximum (90th percentile) proportion of Malays.

8 Conclusion and Future Research

In this paper, I build and estimate a discrete choice model of residential location choices by combining policy variation akin to a regression discontinuity framework with a structural demand estimation framework, motivated by the ethnic housing quotas in Singapore. I exploit the step function in the quota rule as well as within neighborhood, across group variation in prices to identify ethnic preferences from correlated tastes for ethnic neighborhood quality.

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 reaches 43% Chinese and 8% Indians respectively, they would rather add a new neighbor from the other group.

In ongoing work, I explore the possibilities of combining hedonic methods and regression discontinuity to estimate ethnic preferences as well as to compare estimates from the hedonic method and the discrete choice method. Future work will also include the use of estimates on ethnic preferences in this paper to simulate counterfactuals and estimate deadweight losses from integration policies that will need to be weighed against social benefits of integration.

The challenge in performing such welfare calculations is that sorting models typically have multiple equilibria. Findings from this type of simulation can inform the relative deadweight losses and the distributional implications of various integration policies.

References

Ananat-Oltmans, Elizabeth. 2007. "The Wrong Side(s) of the Tracks: Estimating the Causal Effects of Racial Segregation on City Outcomes." NBER Working Paper No. 13343.

Angrist, Josh and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics*, 114(2): 533-75.

Athey, Susan and Guido Imbens. 2007. "Discrete Choice Models with Multiple Unobserved Choice Characteristics." Forthcoming, *International Economic Review.*

Bajari, Patrick and Matthew E. Kahn. 2005. "Estimating Housing Demand With an Application to Explaining Racial Segregation in Cities." *Journal of Business and Economic Statistics*, 23(1): 20-33.

Bayer, Patrick and Christopher Timmins. 2007. "Estimating Equilibrium Models of Sorting Across Locations." *The Economic Journal*, 117(518): 353-374.

Bayer, Patrick, Fernando Ferreira and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." NBER Working Paper No. 13236.

Bayer, Patrick, Robert McMillan, and Kim Rueben. 2004. "An Equilibrium Model of Sorting in an Urban Housing Market." NBER Working Paper No. 10865.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *The Rand Journal of Economics*, 25(2): 242–62.

Berry, Steven T., James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica,* 63(4): 841–90.

Berry, Steven T., James Levinsohn, and Ariel Pakes. 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.

Brennan, Brian. 2000. "Background on MTO." *Moving To Opportunity Research.* http://www.wws.princeton.edu/~kling/mto/background.htm

Card, David, Alexandre Mas and Jesse Rothstein. 2007. "Tipping and the Dynamics of Segregation." NBER Working Paper No. 13052.

Chay, Kenneth and Michael Greenstone. 2004. "Does Air Quality Matter? Evidence from the Housing Market." *Journal of Political Economy*, 113(2): 376-424.

Crawfurd, John. 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." URL: http://dlxs.library.cornell.edu/s/sea/index.php (Cornell University, Southeast Asia Visions)

Cutler, David M.; Glaeser, Edward L. 1997. "Are Ghettoes Good or Bad?" *Quarterly Journal of Economics*, 112(3): 827-72.

Cutler, David M.; Glaeser, Edward L.; Vigdor, Jacob L. 1999. "The Rise and Decline of the American Ghetto." *Journal of Political Economy*, 107(3): 455-506.

Das, S., G. Steven Olley, and Ariel Pakes. 1995. "The Market for TVs." Yale Working Paper.

Hausman, Jerry A., and David A. Wise. 1978. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." Econometrica 46 (March): 403–26.

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

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

Lum, Sau Kim and Tan Mui Tien. 2003. "Integrating Differentials: A Study on the HDB Ethnic Integration Policy." *Crestimes*, Singapore Center for Real Estate (CRES), 3(1): 5.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies,* 60 (3): 531-42.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by Paul Zarembka. New York: Academic Press.

McFadden, Daniel, Antti Talvitie and Associates. 1977. *Demand Model Estimation and Validation.* Berkeley: The Institute of Transportation Studies, University of California, Berkeley and Irvine.

McFadden, Daniel. 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57(5): 995-1026.

Newey, Whitney. 1984. "A Method of Moments Interpretation of Sequential Estimators." *Economic Letters*, 14: 201-206.

Pakes, Ariel, and David Pollard. 1989. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica,* 57(5): 1027–57.

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

Singapore Department of Statistics. 2006. "Singapore Resident Population, 1990 - 2006."

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

Virtual Map Online Street Directory. http://www.streetdirectory.com/

Table 1: Neighborhood and Block Level Quotas Relative to National Ethnic Proportions

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

Source: 2000 Census (Singstat), Lum and Tan (2003)

Table 2: The Relationship between Quotas, Buyer Ethnicity, Seller Ethnicity, and Prices

Binding Quota	Buyer Ethnicity	Seller Ethnicity	Status	Group-Specific Prices
Chinese	Chinese	Chinese	Allowed	High Chinese Prices
	Non-Chinese	Non-Chinese	Allowed	
	Non-Chinese	Chinese	Allowed	
	Chinese	Non-Chinese	Not Allowed	
Malay	Malay	Malay	Allowed	High Malay Prices
	Non-Malay	Non-Malay	Allowed	
	Non-Malay	Malay	Allowed	
	Malay	Non-Malay	Not Allowed	
Indian	Indian	Indian	Allowed	High Indian Prices
	Non-Indian	Non-Indian	Allowed	
	Non-Indian	Indian	Allowed	
	Indian	Non-Indian	Not Allowed	

Notes: The link between group-specific prices and quotas is premised on two correlations: (i) Chinese prices are high when the Chinese neighborhood quality is high; (ii) Chinese quotas are more likely to bind in neighborhoods with high Chinese quality because these neighborhoods attract relatively more Chinese. These two correlations imply that Chinese prices are likely to be positively correlated with Chinese quotas. In addition, if we assume the Chinese, Malay and Indian quality are not perfectly correlated, then, the prices that Chinese buyers are willing-to-pay would be higher than Malay and Indian prices in Chinese-constrained neighborhoods, and lower than Malay and Indian prices in Malay- and Indian-constrained neighborhoods.

Table 3: Summary Statistics

Variable	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 an average flat 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: Regression Discontinuity Results on Average Prices

	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	0.10***	0.10***	0.11***	0.11***	0.06***										
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)										
Percent Chinese	-0.17***	1.70	328.45***	219.38***	42.51										
	(0.06)	(2.03)	(50.25)	(33.43)	(50.71)										
Percent Chinese^2		-1.10	-383.21***	-192.06***	-36.90										
		(1.20)	(58.72)	(29.30)	(43.82)										
Percent Chinese^3			148.63***	dropped	dropped										
			(22.84)												
Percent Chinese^4				43.27***	8.21										
				(6.63)	(9.68)										
Malay Quota						-0.04***	-0.04***	-0.04***	-0.04***	-0.03***					
						(0.01)	(0.01)	(0.01)	(0.01)	(0.01)					
Percent Malay						0.10	-1.73***	-8.15**	-77.81***	-3.10					
						(0.06)	(0.51)	(3.49)	(21.28)	(22.82)					
Percent Malay^2							4.01***	32.13**	488.59***	11.61					
							(1.10)	(15.16)	(138.35)	(147.89)					
Percent Malay^3								-39.77*	1,338.48**	-15.08					
								(21.39)	(391.85)	(416.34)					
Percent Malay^4									1,354.56**	1.15					
									(408.09)	(429.84)					
Indian Quota											-0.03***	-0.03***	-0.03***	-0.03***	-0.01
											(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Percent Indian											-0.28***	-0.11	2.32***	-10.64***	1.71
											(0.06)	(0.23)	(0.72)	(2.02)	(1.81)
Percent Indian^2												-0.85	-25.42***	177.01***	-16.35
												(1.09)	(6.94)	(30.25)	(25.18)
Percent Indian^3													73.11***	1,189.35**	45.84
													(20.38)	(184.76)	(145.45)
Percent Indian^4														2,691.83**	-7.05
														(391.55)	(303.64)
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	14136	14136	14136	14136	14136	11471	11471	11471	11471	11471	23871	23871	23871	23871	23871
R-squared	0.01	0.01	0.02	0.02	0.74	0.004	0.01	0.01	0.01	0.71	0.003	0.004	0.004	0.01	0.72

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Note: The first 5 columns are restricted to be 10% above and below the Chinese quota, the next 5 columns, correspond to the Malay quota, and the last 5 columns correspond to the Indian quota. Columns 1-4, 6-9, 11-14 include results from regressions with a quota dummy and polynomials (up to 4th order) of the block level ethnic proportions. Columns 5, 10 and 15 control for unit characteristics (average age of building, its squared, number of 1-room flats, 2-room flats etc.), month and town fixed effects.

Table 5: Regression Discontinuity Results on Estimated Group-Specific Prices

	Dependent variables		
	Predicted Ln Price	Predicted Ln Price	Predicted Ln Price
	(1)	(2)	(3)
Chinese Quota	0.06***		
	(0.01)		
Chinese Quota * Indian Quota	0.20***		
	(0.01)		
Chinese Quota * Malay Buyer	-0.02**		
	(0.01)		
Chinese Quota * Indian Buyer	-0.003		
	(0.01)		
Chinese Quota * Indian Quota * Malay Buyer	-0.06***		
	(0.02)		
Malay Quota		0.03***	
		(0.01)	
Malay Quota * Indian Quota		-0.07***	
		(0.01)	
Malay Quota * Indian Buyer		0.01	
		(0.01)	
Malay Quota * Chinese Buyer		0.01	
		(0.01)	
Malay Quota * Indian Quota * Chinese Buyer		-0.01	
		(0.01)	
Indian Quota			-0.04***
			(0.005)
Indian Quota * Chinese Quota			0.18***
			(0.01)
Indian Quota * Malay Quota			-0.09***
			(0.01)
Indian Quota * Chinese Buyer			-0.005
			(0.01)
Indian Quota * Malay Buyer			0.002
			(0.01)
Indian Quota * Chinese Quota * Malay Buyer			-0.03**
			(0.01)
Indian Quota * Malay Quota * Chinese Buyer			-0.02
			(0.01)
Chinese Buyer		0.004	0.01***
		(0.004)	(0.003)
Malay Buyer	-0.01***		-0.004
	(0.004)		(0.003)
Indian Buyer	-0.01	-0.01	
	(0.004)	(0.003)	
Controls	Y	Y	Y
Obs	10767	10149	17394
R-squared	0.24	0.35	0.28

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

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 1-room flats, 2-room flats etc. The omitted group is the Chinese buyer (column 1), the Malay buyer (column 2) and the Indian buyer (column 3).

Table 6: Regression of Actual Quota Status on Estimated Quota Status

	Dependent variables					
	Actual Chinese Quota	Actual Malay Quota	Actual Indian Quota	Actual Chinese Quota	Actual Malay Quota	Actual Indian Quota
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Chinese Quota	0.48***			0.17*		
	(0.05)			(0.07)		
Predicted Malay Quota		0.65***			0.42***	
		(0.12)			(0.12)	
Predicted Indian Quota			0.30**			-0.14
			(0.10)			(0.13)
Controls	N	N	N	Y	Y	Y
Obs	170	170	170	170	170	170
Fstat	85.63	31.50	8.61	7.23	3.82	2.33
R-squared	0.34	0.16	0.05	0.46	0.31	0.22

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Note: Results in columns 1-3 do not include other controls that will be in the full model, columns 4-6 control for own attributes and the instruments. 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.

Table 7: Logit and IV Logit

	Dependent variables					
	Ln Chinese Share	Ln Malay Share	Ln Indian Share	Ln Chinese Share	Ln Malay Share	Ln Indian Share
	(1)	(2)	(3)	(4)	(5)	(6)
School Quality	0.12***	0.09***	0.11***	0.12***	0.09***	0.10***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
Distance to Subway	-0.42**	-0.14	-0.39**	-0.42**	-0.09	-0.36*
	(0.19)	(0.16)	(0.18)	(0.21)	(0.20)	(0.22)
Average No. of Rooms	-0.08***	-0.05***	-0.08***	-0.09***	-0.05**	-0.04
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Average Age of Buildings	-0.09	-0.26	-0.24	-0.77	-0.85*	-0.01
	(0.30)	(0.24)	(0.27)	(0.48)	(0.46)	(0.40)
Percent Ingroup	76.90***	17.54***	22.68**	97.32*	26.87**	28.34
	(24.81)	(5.05)	(9.25)	(52.71)	(12.25)	(22.75)
Percent Ingroup^2	-51.40***	-34.78**	-86.67**	-67.61**	-59.27	-230.05**
	(15.77)	(17.14)	(40.13)	(34.09)	(39.15)	(107.16)
Price	-11.72***	-4.68**	-6.63***	-0.76	5.71	-1.14
	(2.85)	(2.14)	(2.42)	(6.64)	(7.06)	(4.75)
Constant	-27.58***	-2.29*	-0.89	-33.35*	-3.13**	-3.43*
	(9.73)	(1.16)	(1.33)	(20.15)	(1.50)	(1.94)
Obs	170	170	170	170	170	170
R-squared	0.37	0.38	0.27	0.29	0.29	-0.06

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: The dependent variable is the log of the Chinese share for neighborhood j subtracted by the log of the Chinese share for the outside good (columns 1and 4). I define the dependent variables for the other columns in a similar manner. 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 as well as the estimated quotas.

Table 8: Random Coefficients Logit

Variables	Units	Common Taste Parameters (1)	Chinese Taste Parameters (2)	Malay Taste Parameters (3)	Indian Taste Parameters (4)
Means $(\bar{\beta}, \bar{\alpha})$					
Constant		-5.69***			
		(1.70)			
School Quality	.1 awards	1.46***			
		(0.15)			
Distance to Subway	1 km	-0.24**			
		(0.13)			
Average No. of Rooms	0.1 rooms	-6.77***			
		(2.34)			
Average Age of Buildings	0.01 years	-5.63***			
		(1.43)			
Percent Ingroup			4.83**	4.00	5.64**
			(2.70)	(7.85)	(1.46)
Percent Ingroup^2			-5.58**	1.17	-3.62**
			(2.96)	(2.71)	(1.09)
Price	S$million	-2.11			
		(3.29)			
Heterogeneity (σ)					
Constant		-1.30***			
		(0.23)			
Average No. of Rooms	0.1 rooms	-3.78***			
		(0.77)			
Price	S$million	0.29			
		(0.39)			

Standard errors in parentheses adjusted for sequential estimators, using Newey (1984)
* significant at 10%; ** significant at 5%; *** significant at 1%

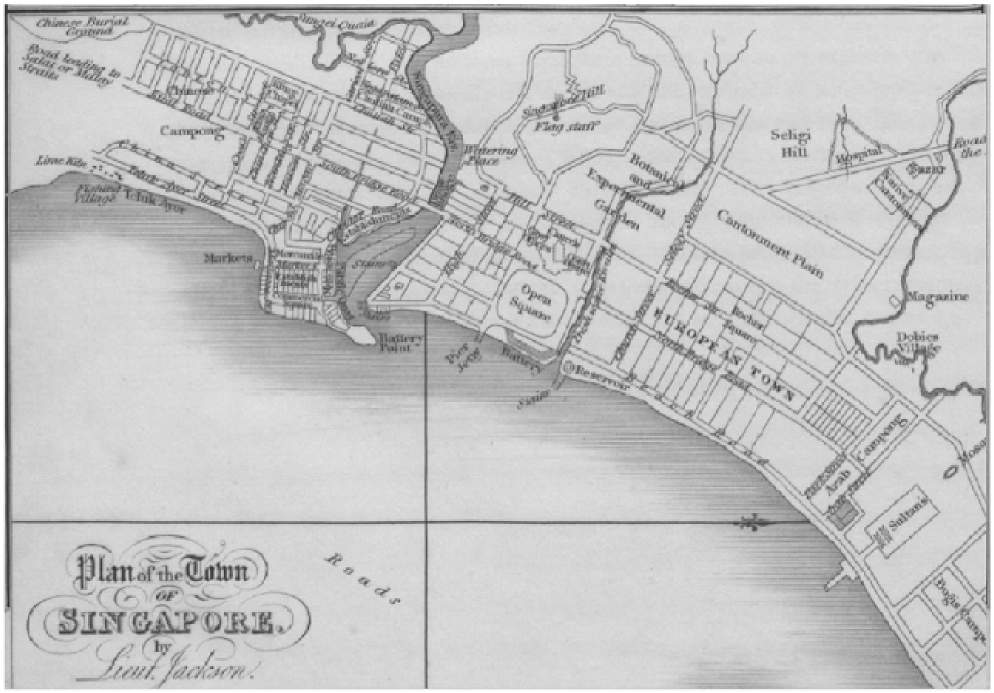
Notes: Variables are scaled so that the mean is between 0 and 1. The units are in the table. For example, the coefficient on *School Quality* implies that an increase in 10 awards is associated with an increase of 1.46 utils. For the variable *Percent Ingroup*, *percent Chinese, percent Chinese^2* and *percent Malay* are not scaled; *percent Malay^2* and *percent Indian* are multiplied by 10; *percent Indian^2* is multiplied by 100. The instruments are the sum of school awards, the distance to the closest subway station, the average age of buildings, the average number of rooms for nearby neighborhoods within 1-3km, 3-5km and within 5-7km, a dummy for being in the east of the early Malay settlements as well as the estimated quotas.

Table 9: MRS Evaluated at Various Ethnic Proportions in the Sample

	Chinese (1)	Malays (2)	Indians (3)
Relevant statistics for ethnic proportions			
Mean of *Percent Ingroup*	79%	13%	8%
10th percentile of *Percent Ingroup*	70%	5%	5%
90th percentile of *Percent Ingroup*	88%	22%	12%
Standard Deviation of *Percent Ingroup*	7%	7%	3%
MRS relative to distance to subway (km)			
MRS at mean of *Percent Ingroup*	-1.18	2.09	-0.21
MRS at 10th percentile of *Percent Ingroup*	-0.88	1.54	2.56
MRS at 90th percentile of *Percent Ingroup*	-1.48	2.72	-3.90
MRS relative to age of building (years)			
MRS at mean of *Percent Ingroup*	-4.95	8.76	-0.86
MRS at 10th percentile of *Percent Ingroup*	-3.70	6.43	10.73
MRS at 90th percentile of *Percent Ingroup*	-6.20	11.38	-16.32

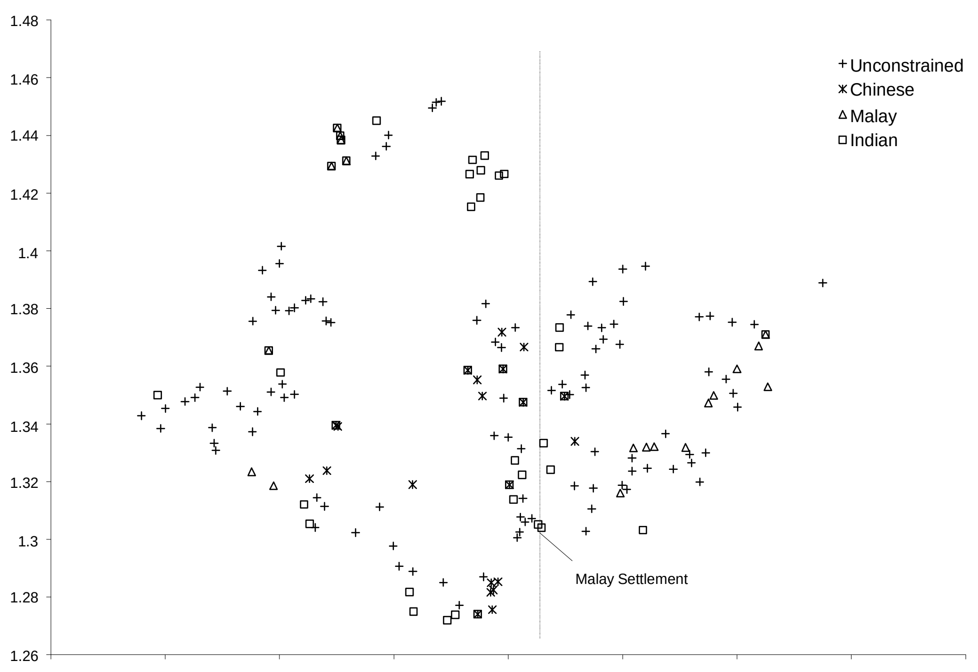
Note: This table shows calculations of the MRS evaluated at different ethnic proportions. The top panel shows the relevant statistics for the ethnic proportions. The second panel represents the MRS relative to the distance to the subway station and the third panel represents the MRS relative to the average age of the building. Since ingroup preferences are quadratic in percent ingroup, the marginal utilities vary with the percent of ingroups in the neighborhood. Each number in the 2nd and 3rd panel represents the marginal rate of substitution evaluated at neighborhoods with the mean, the 10th percentile, and the 90th percentile of *Percent Ingroup.* The MRS is calculated as the estimated marginal utility to live in a neighborhood with a one standard deviation increase in percent ingroup divided by the (negative of the) marginal utility for distance to the closest subway station (2nd panel) as well as the (negative of the) marginal utility for the average building age (3rd panel). Since distance to the subway and building age are both bad attributes, I use the negative of their marginal utilities in the denominators so that a positive MRS reflects ingroup preferences and a negative MRS reflects outgroup preferences. Although the MRS's for the Chinese are negative for all values of *percent Chinese* calculated for this table, Chinese do have positive MRS's when the percent of Chinese in a neighborhood is less than 43% (see Figure 3). However, all neighborhoods in the sample have more than 43% Chinese, which is why the MRS shown here for the Chinese are all negative.

Figure 1a: Map of Ethnic Settlements in Early 19th Century Singapore



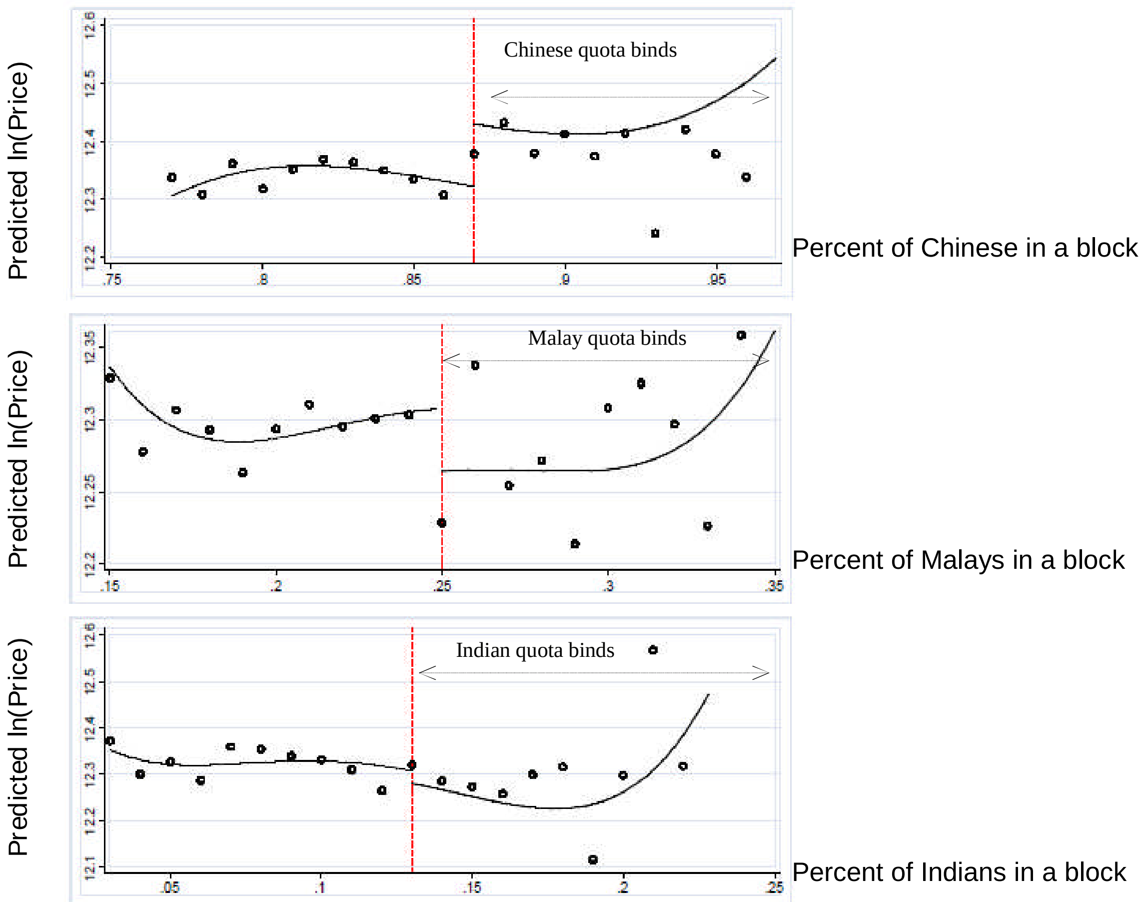
Source: Crawford (1828)
Notes: The Malay settlements ("Arab Kampong" and "Bugis Kampong") are in the south east corner of the map, just east of the European Town. The Chinese and Indian areas are to the west of the Singapore River.

Figure 1b: Map of Current Neighborhood by Quota Status



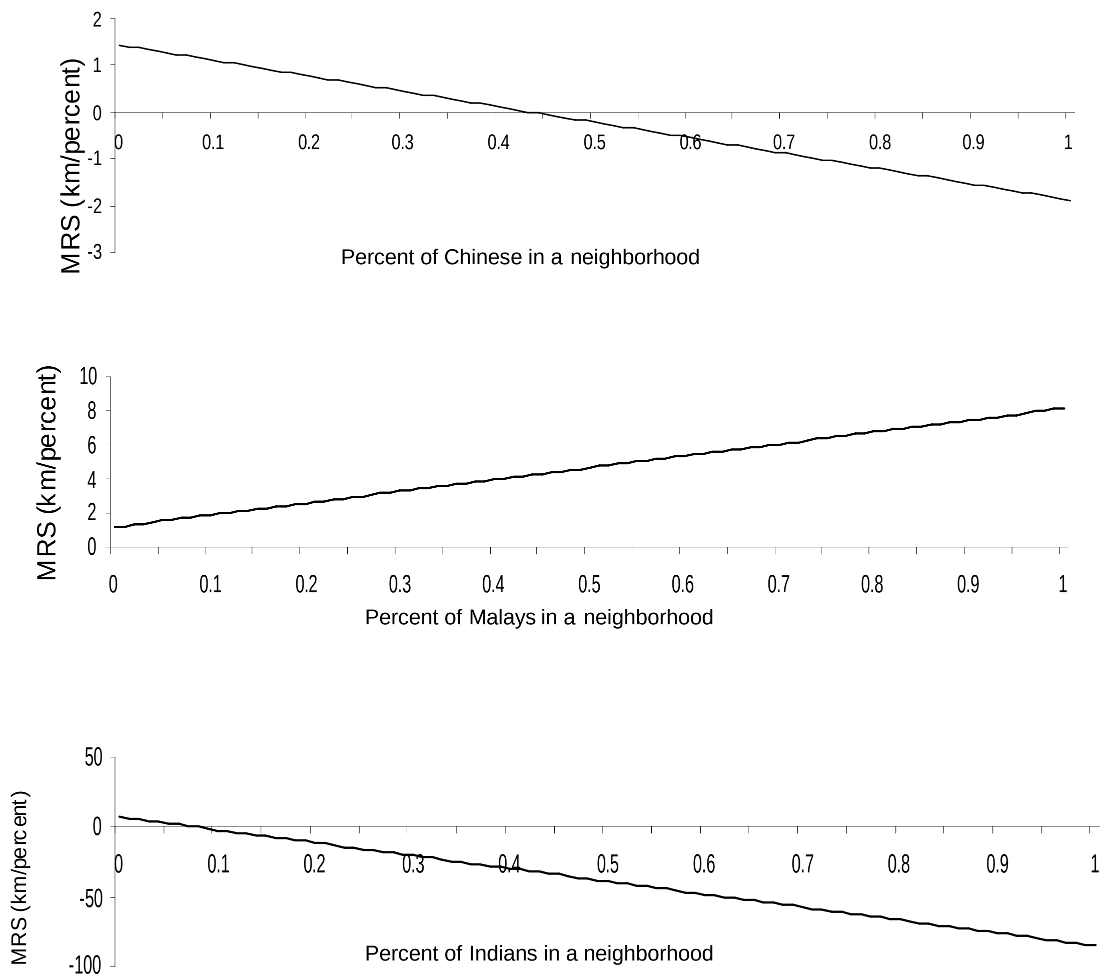
Source: Virtual Map Online Street Directory
Note: This is a map of 170 neighborhoods comprising unconstrained neighborhoods, Chinese-, Malay- and Indian-constrained neighborhoods. The line indicates the eastern tip of the early Malay settlements.

Figure 2: Average Transaction Prices above and below Chinese, Malay and Indian Quotas



Notes: Each panel in this figure is constructed using the following procedure for observations within 10% of the ethnic quotas: (i) regress the log of average transaction prices on the corresponding running variable (to a 4th order polynomial) and a dummy that is one when the corresponding block quota is binding; (ii) plot the predicted prices above and below the quota separately (iii) plot means of ln(price) for each 1% bin. I repeat the exercise for the Malay quotas and Indian quotas.

Figure 3: MRS between *Percent Ingroup* and Distance to the Subway Station for the Chinese, Malays and Indians as a Function of Neighborhood Ethnic Proportions



Note: Each point on the line represents the ratio of the estimated marginal utility to live in a neighborhood with a one standard deviation increase in *Percent Ingroup* divided by the marginal utility for distance to the closest subway station. Since ingroup preferences are quadratic in *Percent Ingroup*, the marginal utility varies with the percent of ingroup in the neighborhood (x-axis). The plot shows that in neighborhoods with less than 43% Chinese and 8% Indians respectively, the Chinese and Indians have a preference for a marginal neighbor who is Chinese. Above 43%, Chinese have a preference for a marginal neighbor who is from the outgroup. The minimums and maximums of the ethnic proportions in the sample are 61% and 98% for the Chinese; 0.6% and 33% for the Malays; 1.3% and 26% for the Indians.

Data Appendix

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 movers moved between April 2005 and March 2006 and they changed their phone records immediately after they move.

Neighborhoods

I use six-digit postal codes to define neighborhoods. Blocks that are within the same sector (defined by the first 2 digits of the postal code) and whose 3 digit block numbers share the same first digit are assigned to the same neighborhood.

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/.

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 March 2005. These are dummy variables for whether a block was constrained. If all blocks were constrained in a neighborhood, I say the neighborhood quota is binding.