Income Segregation and Local Progressive Taxation: Empirical Evidence from Switzerland

Kurt Schmidheiny*
Department of Economics, Tufts University
and CESifo

October 11, 2005

Forthcoming in *Journal of Public Economics*

Abstract

This study investigates spatial income segregation in fiscally decentralized urban areas. The theoretical part proposes the progressivity of local income taxes as a new theoretical explanation for income segregation. The empirical part studies how income tax differentials across municipalities affect the households' location decisions. I use data from the Swiss metropolitan area of Basel that contains tax information on all moving households in 1997. The location choice of the households is investigated within the framework of the random utility maximization model. Different econometric specifications of the error term structure, such as conditional logit, nested logit and multinomial probit are compared. The empirical results show that rich households are significantly and substantially more likely to move to low-tax municipalities than poor households. This result holds after controlling for alternative explanations of segregation. Social interactions and distance from the central business district are established as other major factors for income segregation. Households in general tend to choose locations close to other households like themselves.

Key Words: Location Choice, Segregation, Fiscal Federalism, Progressive Taxation, Discrete Choice

JEL-Classification: H71, H73, R20, R23

*The discussions with Klaus Neusser, Mike Gerfin, and Veronika Aegerter, greatly improved this paper. I would like to thank the Tax Administration and the Statistical Office of the Canton of Basel-Stadt for the provision of the unique household data set, my thanks to Martin Sandtner in particular. I am grateful to Wüest und Partner, and especially to Jürg Mosimann, for the provision of the housing price data. The hospitality of the HWWA Hamburg Institute of International Economics is gratefully acknowledged. I thank Katherine Baicker, Michel Bierlaire, Denis Bolduc, Marius Brülhart, Dennis Epple, Yannis Ioannides, Mario Jametti, Holger Sieg, Hans-Werner Sinn, Michael Smart, Kenneth Train, and the participants of the TAPES Conference on Fiscal Federalism, the Crans-Montana Workshop on Globalization and Economic Geography and of seminars at Carnegie-Mellon, University of Toronto and McMaster for valuable comments and suggestions. Heather Murray helped me find the right words.

1 Introduction

"To secure an efficient outcome, the provision of public services should be determined and paid for by those who benefit", as Musgrave (Buchanan and Musgrave 1999, p. 156) pointedly remarks. Oates (1972) also argued that local units deciding upon public programs are more likely to trade off costs against benefits if these programs are financed by local taxes.

While the virtues of decentralized financial responsibility are uncontested, the resulting tax differentials are highly disputed. Tax differentials are often seen as consequence of different preferences for the locally provided public goods. However, different tax rates can also be the result of different economic resources of the local population, since rich local jurisdictions can raise the same revenue with lower tax rates than poor ones. While the effect of the tax base on tax rates is trivial, the opposite effect is less evident. This paper addresses the question whether tax differentials across local jurisdictions are not just the consequence, but also the cause of differences in local average income.

The theoretical part of this paper proposes the progressivity of a local income tax as a new theoretical explanation for income segregation of the population. The empirical part studies the community choice of households in Switzerland. Swiss metropolitan areas are a laboratory for federal systems as they are divided into a multitude of municipalities with extensive political and fiscal autonomy. Moreover, the main local tax is on income rather than on property. The estimated multinomial response models show that rich households are significantly and substantially more likely to move to low-tax municipalities than poor households. This result holds after controlling for alternative explanations of segregation. Social interactions and distance from the central business district are established as other major factors for income segregation.

The theoretical literature on the local provision of local public goods goes back to Tiebout (1956). Tiebout showed that fiscal decentralization leads to an efficient provision of local public goods because people with similar preferences would settle in particular municipalities and vote for their desired level of public goods provision. Tiebout's result rests on the assumption that households have equal incomes. The location of households and the local provision of public goods when households differ in incomes was studied by Ellickson (1971), Westhoff (1977) and in the literature surveyed in Ross and Yinger (1999).

The *segregation hypothesis* is one of the central propositions in multi-community models in the tradition of Tiebout. Endogenous segregation means that different people choose different locations in equilibrium. While the Tiebout model focuses on preference heterogeneity, Ellickson and Westhoff turned their attention to income as the main dimension of difference. Several mechanisms have been proposed to explain why rich households make different choices than poor households (see Ross and Yinger, 1999, for property tax models and Schmidheiny, 2002, for income tax models). The nature of the local public good, ranging from a monetary transfer to a non-substitutable pure public good, induces a self-sorting of the population when rich households esteem public goods relatively more than poor households. Another mechanism draws on the income elasticity of housing: If housing expenditures become relatively less important with increasing income, rich households are less concerned about high housing prices than poor households.

The segregation mechanism in this paper builds on the empirical fact that most income tax schedules are progressive and that local jurisdictions can often only set the tax level within a given federal tax schedule. The high priority of tax rates in rich households' decisions is explained by the progressivity of the tax schedule.

The segregation hypothesis of Tiebout type models has been challenged by a series of empirical studies.[1] A first strand of research investigates the equilibrium predictions of multi-community models using data on aggregate community characteristics. Epple and Sieg (1999) and Epple, Romer and Sieg (2001) develop a strategy for estimating the household preference parameters in a equilibrium model where the local income distribution and local policy variables are simultaneously determined. They show that the differing income quantiles across 92 local jurisdictions in the Boston area are well explained by the model predictions. Feld and Kirchgässner (2001) regress the share of various income classes in Swiss cantons and main cities on income tax rates. They find a strong negative relationship between the tax rate and the share of rich households. However, their treatment of the generic endogeneity of tax rates by instrumental variables from mainly lagged observations may not be sufficient to solve the endegoneity problem, as the general equilibrium of tax rates and income segregation is most likely a long-run phenomenon. Rhode and Strumpf (2003) assess the importance of the segregation mechanism in Tiebout type models from a historical perspective. They collected an impressive data set with measures of heterogeneity in the population over a period of 140 years. Given that the costs of moving dramatically declined during this time, multi-community models predict that the population within local units should have become more homogeneous while the differences across local units should have aggravated. They conclude that their data do not support the model predictions on a national scale. For metropolitan areas,

[1]The early empirical literature on multi-community models investigated the relationship between local tax differentials, public goods provision and housing prices. Oates (1969) and and a multitude of subsequent studies (surveyed in Ross and Yinger, 1999) strikingly confirm the so-called capitalization hypothesis, which predicts that low taxes and attractive public goods provision should be reflected in high housing prices.

however, the observed pattern does not contradict the segregation hypothesis.

The second empirical approach - also used in this paper - directly targets the location choice of individual households using a multinomial response framework. This approach circumvents the endogeneity problem because, from the perspective of a single household, the community characteristics can be taken as given. Friedman (1981) used a conditional logit model to study the location choice of 682 households among nine residential areas in the San Francisco area. Nechyba and Strauss (1998) use the same model to study the choice of over 22,000 households among six school districts in the suburbs of Philadelphia. Both studies show that public expenditures are an important locational factor. The segregation hypothesis needs explicit consideration as household specific variables are not identified in linear conditional logit models (see Section 5.1). In need of a variable that depends on both household and community characteristics, Nechyba and Strauss calculate the households' hypothetical consumption of private goods for all school districts. This variable depends on after-tax local housing prices assuming that households consume the same amount of housing in all school districts. They therefore implicitly assume that the price elasticity of housing is zero.[2]

Bayer, McMillan and Rueben (2004) attempt a combination of the two empirical approaches. Following Berry, Levinsohn and Pakes (1995), they first estimate the households' choice of a neighborhood, using neighborhood fixed effects and a multitude of interaction effects between household and neighborhood characteristics. In a second step they explain the neighborhood fixed effects by neighborhood characteristics using instrumental variables which make use of an explicit general equilibrium model. The predictions of the estimated model therefore adequately take into consideration the (long-run) adjustment of the endogenous aggregate neighborhood characteristics.

This study follows Friedman (1981) and Nechyba and Strauss (1998) but shifts the focus to assessing the (income) segregation hypothesis. The general locational attractiveness of a community is considered in community specific constants. This is equivalent to the first stage estimation in Bayer, McMillan and Rueben (2004). The specification of the model is drawn on an explicit theoretical multi-community model.

The paper is organized as follows: Section 2 describes fiscal decentralization in Switzerland. A theoretical model of location choice based on progressive income taxation is proposed in Section 3. Alternative explanations of income segregation are discussed in Section 4. The econometric model is outlined in Section 5, while Section 6 describes the data. The empirical results are presented in Section 7. Section 8 draws conclusions.

[2] This assumption is relaxed by using community-specific coefficients for household income. Alternative specific coefficients, however, cannot be derived from a random utility framework.

2 Fiscal Decentralization and Progressive Taxation in Switzerland (and elsewhere)

Switzerland is an exemplary federal fiscal system. The Swiss federation comprises 26 states, the so-called cantons. The cantons are divided into roughly 3000 municipalities of varying size and population. All three state levels finance their expenditures essentially by their own taxes and fees. The total tax revenue of all three levels was 77 billion CHF in 1997, of which 45% is imposed by the federation, 32% by the cantons and 23% by the municipalities.[3] While the federal government is mainly financed by indirect taxes (58% of federal tax revenue) such as the VAT, the cantons and municipalities largely rely on direct taxes. Income taxes account for 65% of cantonal and 73% of municipal tax revenue. In total, 45% of the income tax revenue go to the cantons, 37% to the municipalities and only 18% to the federal government. Transfers between the three levels are not a major part of the budgets of cantons (27% of total revenue) and municipalities (15%).

The cantons organize their tax systems autonomously. For example, they decide upon the level of income and corporate taxes and the degree of tax progression. The individual municipalities in turn can generally set a tax shifter for income and corporate taxes. The municipal tax is then the cantonal tax rate multiplied by the municipal tax shifter. In some cantons, for example in the Canton of Basel-Stadt before 2001, the individual municipalities also have some freedom in setting the tax schedule. The decisions in the cantons and municipalities are made by the legislative body and are subject to referendums. Federal and cantonal systems of fiscal equalization limit the tax differences across cantons and across municipalities within the same canton to some extent, but still leave room for considerable variation.

The above outlined federal system leads to ample differences of average income taxes across Swiss municipalities. For example, for a two-child family with a gross income of 60,000 Swiss francs (CHF) the sum of cantonal and municipal average income tax ranged from 2.3% in Baar in the canton of Zug to 8.2% in Lauterbrunnen in the Canton of Bern in the year 1997 (see the data sources in the appendix). The federal income tax for this household was 0.7%. With an income of 500,000 CHF a two-child family faced much higher average tax rates due to the progressive federal and cantonal tax schedules, namely ranging from 10.9% in Wollerau in the Canton of Schwyz to 28.7% in Onex in the Canton of Geneva. The federal average income tax for this household was 9.4%.

[3] All figures in this paragraph apply to 1997. Source: Eidgenössisches Finanzdepartement (Swiss Federal Department of Finance), Öffentliche Finanzen 1999: Bund, Kantone, Gemeinden.

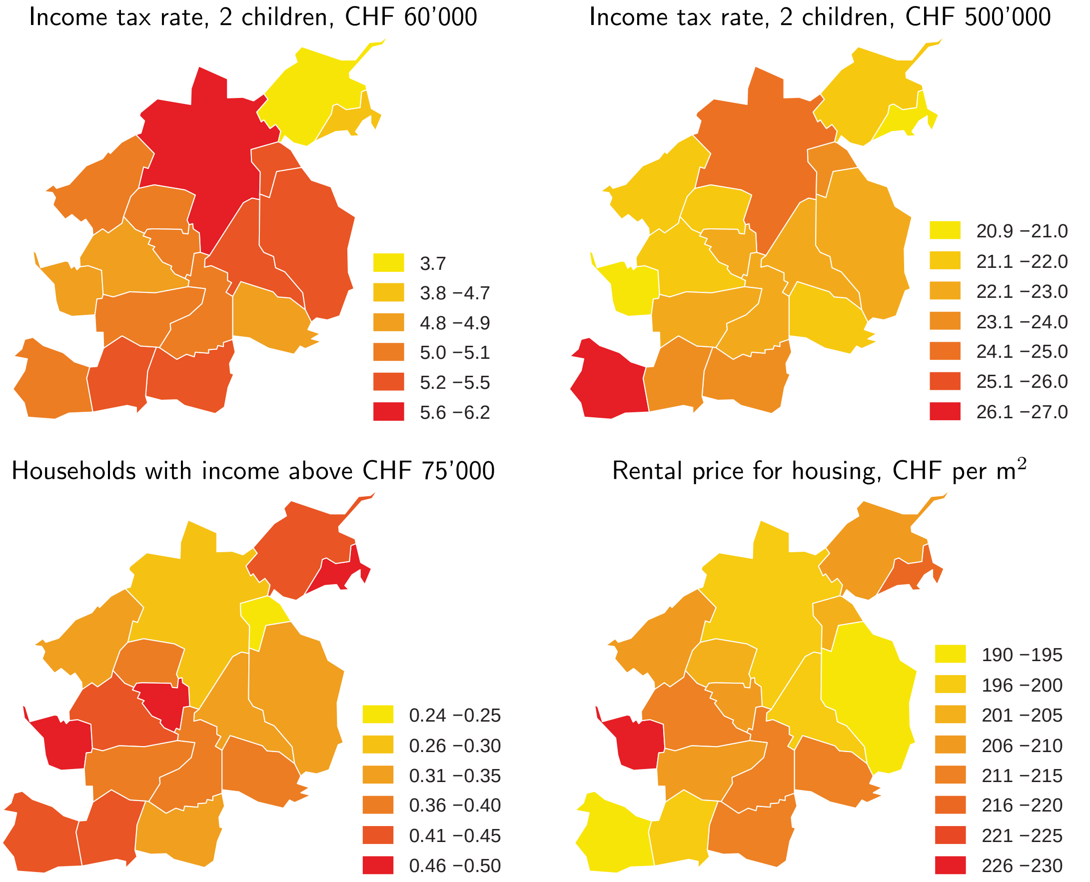

Figure 1: Characteristics of municipalities in the metropolitan area of Basel.

The tax differences across municipalities within metropolitan areas are also substantial. Figure 1 shows characteristics of the municipalities in the metropolitan area of Basel[4], the third largest Swiss metropolitan area. The Basel area depicted in Figure 1 had 340,000 inhabitants in 1997. The city of Basel with 172,000 inhabitants includes the central business district of the area. The area comprises municipalities from three cantons: Basel-Stadt, Basel-Land, and Solothurn. There is great variability in both tax levels and tax schedules. The totalled municipal and cantonal average income tax rate for a two-child family with an annual income of CHF 60,000 is depicted in the top-left map in Figure 1. The taxes are highest in the city of Basel (6.2%) and up to 41% lower in the municipalities around the center. A rich household with an income of CHF 500,000 faces a similar spatial pattern of taxes (top-right map), but much higher average tax rates due to the steep progressivity of income taxes.

It is particularly interesting to contrast the local tax rates with the income of the residents. The bottom-left map in Figure 1 shows the local share of households with incomes above 75,000 CHF. The map shows a quasi inverted picture of the

[4]See definition of the area in Section 6 and data sources in the appendix.

tax rates. The high-tax center municipality has the lowest proportion of rich households, whereas the low-tax municipalities are populated by comparatively more rich households. The rental prices for housing (bottom-right map) are also correlated with the tax rates. The low-tax fringe of municipalities around the center exhibit higher average prices than the center.

While local taxation of *property* is widespread, local taxation of *income* is rarer. Local income taxation at the municipal level is for example observed in four U.S. states (Indiana, Maryland, Ohio, and Pennsylvania) and in Denmark. These states and countries apply a *flat* local tax. Belgium is to my knowledge the only other country besides Switzerland with progressive tax schemes at the municipal level. Canada had a similar system at the provincial level between 1977 and 1996 (see Boadway and Kitchen, 1980) when personal income taxes in Canadian provinces were a percentage of the progressive federal tax.

3 A Model of Location Choice and Local Progressive Income Taxation

The theoretical model describes a metropolitan area with a fixed number J of distinct local jurisdictions, called communities. The political borders of the communities are the outcome of a historical process and thus taken as given. The area is populated by a continuum of heterogeneous households, which differ in incomes. Income is distributed according to a distribution function $f(y) > 0$ with support $[\underline{y}, \overline{y}], \underline{y} > 0, \overline{y} < \infty$. There are three goods in the economy: private consumption b, housing h and a public good g. This model is close to Epple and Romer (1991) but with income rather than property taxes.

Each community j spends the amount $n_j g$ to provide the local good g, where n_j is the measure of households living in community j. The communities levy income taxes to finance the local good. In each community j, the tax rate consists of two parts, a local tax shifter $t_j > 0$ and a progressive tax rate structure $r(y)$. I assume $r(y)$ continuous in y, $r(y) > 0$, the average tax rate $t \cdot r(y) \in [0, 1)$ and the marginal tax rate $t[r + yr'(y)] \in [0, 1)$. The tax rate structure $r(y)$ is exogenous (to the communities) and identical across communities.

I assume that the local good g is fixed and identical across communities. The local good g is therefore locally provided, locally financed but centrally decided upon. There are two reasons for assuming exogeneity of the local good: First, many locally financed goods, particularly in Switzerland, satisfy this description. Schooling, for example, accounts for the largest item in municipal budgets in Switzerland; local neighborhood schools are locally provided and locally financed. Cantonal regulations, however, leave little discretion for financially relevant decisions.[5] Second, the aim of this paper is to study the progressivity of income taxes as a source of income segregation in Switzerland. However, the model would become intractable allowing for both progressive taxation and endogenous provision of local goods.[6]

In each community j, the tax shifter t_j is determined by budget balance. The price for housing p_j in community j is established on a competitive housing market. Hence, the communities are fully characterised by their local income tax level and their local price for housing. A household can move costlessly and chooses the community which maximizes its utility as place of residence.

3.1 Household Preferences

The preferences of the households are described by a utility function[7]

$$U(h, b) \ ,$$

where h is the consumption of housing and b the consumption of the private good. The utility function is assumed to be strictly increasing, strictly quasi-concave and twice continuously differentiable in h and b.

Households face a budget constraint:

$$ph + b \leq y_d = y[1 - t \cdot r(y)] \ ,$$

where p is the price of housing.[8] The price of the private good is set to unity. Disposable income y_d depends on the local income tax shifter $t > 0$ and the exogeneous tax rate structure $r(y)$.

Maximisation of the utility function with respect to h and b subject to the budget constraint yields the housing demand function $h^* = h(p, y_d) = h(t, p\,; y)$, the demand for the private good $b^* = y(1 - t\,r) - ph(t, p\,; y)$, and the indirect utility function

$$V(t, p\,; y) := U(h^*, b^*) \,. \tag{1}$$

[5]In particular, teachers' salaries and class size are regulated by cantonal law. Furthermore, cantonal courts ruled, based on equity considerations, that schools (in rich neighborhoods) are not allowed to provide additional tutoring or extra classes for extraordinary strong or weak pupils.

[6]Schmidheiny (in press) studies endogenous local goods which are determined through municipal majority votes, but financed by *flat* local income taxes. This model exhibits very similar equilibrium properties as the one presented here.

[7]The public good does not explicitly enter the utility function because it does not affect the following considerations as it is assumed to be constant across communities.

[8]Note that the price for housing is a per-unit (for example m^2) price, which is independent of the level the consumption. Large houses therefore have the same per-unit price as small houses. Perfect price discrimination on the housing market related to the consumers income (or equivalently to the demanded house size) could offset the segregation mechanism that lead to Property 2. However, the assumption of one common price is plausible, at least in the long-run, as arbitrage opportunities would otherwise emerge.

Property 1 is a trivial result of the strictly increasing nature of the utility function and is derived by applying the implicit function theorem and the envelope theorem:

Property 1 (MRS between community characteristics)

$$M(t, p, y) := \left.\frac{dt}{dp}\right|_{dV=0} = -\frac{\partial V/\partial p}{\partial V/\partial t} = -\frac{h^*}{y \cdot r(y)} < 0\,.$$

The marginal rate of substitution (MRS) between community characteristics reflects a household's trade-off between taxes and housing prices. Property 1 simply follows from the fact that households dislike both high taxes and high housing prices. A household can therefore be compensated for a tax increase by a decline in housing prices and vice-versa.

The following two assumptions about the form of the indirect utility function generate the segregation by income.

Assumption 1 (Income elasticity of housing)

$$\varepsilon_{h,y_d} := \frac{\partial h^*}{\partial y_d}\frac{y_d}{h^*} \leq 1 \quad \textit{for all } y_d \textit{ and } p.$$

Assumption 1 means that housing is a normal good, i.e. the elasticity of housing with respect to disposable income is smaller or equal to unity. This implies that the share of housing in the household's budget decreases with after-tax income.

Assumption 2 (Progressive taxation)

$$\frac{\partial r(y)}{\partial y} \geq 0 \quad \textit{for all } y.$$

Assumption 2 states that the income tax schedule is proportional or progressive.

Property 2 (Relative preferences)
If Assumptions 1 and 2 hold and if and only if at least one of them holds with strict inequality, then

$$\frac{\partial M}{\partial y} = \left[1 - \frac{\partial h^*}{\partial y_d}\frac{y_d}{h^*}\frac{\partial y_d}{\partial y}\frac{y}{y_d}\right]\frac{h^*}{y^2 r(y)} + \frac{\partial r(y)}{\partial y}\frac{h^*}{y r^2(y)} > 0$$

for all y, t and p.

Proof: Assumption 1 states that $(\partial h^*/\partial y_d)(y_d/h^*) \leq 1$. The assumptions about the relation and the bounds of the average and the marginal tax rate guarantee that $(\partial y_d/\partial y)(y/y_d) = [1 - t\,r(y) - t\,y\,r'(y)]/[1 - t\,r(y)]$ lies in $[0, 1]$. If Assumption 2, $\partial r(y)/\partial y > 0$ is strictly satisfied, both addends of $\partial M/\partial y$ are strictly positive. If Assumption 2 is not strictly satisfied, $\partial r(y)/\partial y = 0$, and Assumption 1 is strictly satisfied, $(\partial h^*/\partial y_d)(y_d/h^*) < 1$, then the second addend is zero and the first addend is strictly positive. This completes the proof of the "if" statement. The "only if" statement is given by the fact that if $\partial r(y)/\partial y = 0$ and $(\partial h^*/\partial y_d)(y_d/h^*) = 1$ then both addends are zero. □

Property 2 states that the MRS between local tax levels and housing prices increases monotonically with income. This means that rich households have a relatively stronger preference for low taxes than poor households. Property 2 explains why rich households make different location decisions than poor households. It is therefore the central condition giving rise to income segregation. Westhoff (1977) called the analogous assumption 'relative preference assumption'. It is also called the single-crossing condition. In this model, relative preferences are either caused by the progressive tax schedule, the income elasticity of housing below unity or a combination of both. As will become apparent in Section 5, Property 2 plays a key role in the identification of tax rate effects in random utility maximization models of location choice.

3.2 Location Choice

A household with income y chooses the community which maximizes its utility. Hence, given the set of community characteristics (t_j, p_j) for $j \in C = (1, ..., J)$, a household prefers community j if and only if

$$V(t_j, p_j\,; y) \geq V(t_i, p_i\,; y) \quad \text{for all } i\ . \tag{2}$$

The following propositions describe the allocation of households to communities when all communities are populated and exhibit different characteristics.

Proposition 1 (Order of community characteristics)
If all communities are populated and exhibit different community characteristics, then communities with higher housing prices impose lower income tax rates.

Proof: Suppose the opposite, i.e. that one community exhibits both lower prices and lower taxes. Then all households would prefer that community for the same reason that lead to Property 1. This is a contradiction. □

Proposition 2 (Perfect income segregation)
If the relative preference property holds and all communities are populated and exhibit different community characteristics, then all households choosing a community with lower taxes are richer than all households choosing a community with higher taxes.

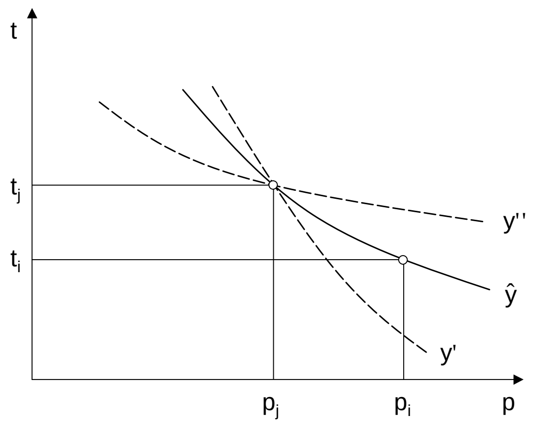

Figure 2: Indifference curves in the (t, p) space.

Proof: The proof proceeds in three steps. Firstly, it is shown that there is a 'border' household in a comparison of two communities. Secondly, income segregation is shown in a two community case. Thirdly, the result is extended to more than two communities.

(1) Define $V_j(y) := V(t_j, p_j, y)$ as a household's utility in j and $V_i(y) := V(t_i, p_i, y)$ in i. Let the household with income y' prefer j to i, hence $V_j(y') - V_i(y') \geq 0$ and a household with income y'' prefer i: $V_j(y'') - V_i(y'') \leq 0$. From the continuity of V in y follows the continuity of $V_j(y) - V_i(y)$ in y. The intermediate value theorem states that there is at least one $\hat{y}$ between y' and y'' s.t. $V_j(\hat{y}) - V_i(\hat{y}) = 0$. This household is called the border household.

(2) This part uses Figure 2. The figure shows the indifference curves in the (t, p)-space for three different income levels $y' < \hat{y} < y''$. The indifference curves represent all (t, p) pairs that households consider to be as good as community j's (t_j, p_j)-pair. Households prefer pairs south-west of the indifference curve to (t_j, p_j). Note that the indifference curves are decreasing in the (t, p)-space (Property 1). Note also that, due to Property 2, they become flatter as income rises. Imagine a community i, characterized by (t_i, p_i), $p_i > p_j$ and $t_i < t_j$, where household $\hat{y}$ is indifferent to j. All richer households, e.g. y'', prefer the low-tax community i to j and all poorer households, e.g. y', prefer the low-price community j.

(3) The proposition implies that $[\underline{y}, \overline{y}]$ is partitioned into J non-empty and non-overlapping intervals $I_j = \{y | \text{household with income}\, y\, \text{chooses}\, j\}$. Suppose the opposite, i.e. y' as well as y'' prefer community j, but an y''', $y' < y''' < y''$ strictly prefers community i. It follows from step 1 that there is an $\hat{y}$, $y' \leq \hat{y} < y'''$. Step 2 implies that $y'' > \hat{y}$ strictly prefers i to j, which is a contradiction. □

Proposition 2 claims that any community is populated by a single and distinct income class or more fundamentally that rich households choose different communities than poor households. This proposition is assessed in the empirical part of this paper.

Proposition 3 (Non-existence of income segregation)
If the local income tax rate is proportional and the household preferences are homothetic, then rich households choose the same communities as poor households.

Proof: Neither Assumption 1 nor 2 are satisfied with strict inequality. Therefore, Property 2 does not hold and the indifference curves in Figure 2 coincide. Hence, all households are, independently of their income, either indifferent between all communities or all prefer the same community. □

Proposition 3 shows that Property 2 is a necessary condition for income segregation. There is no systematically different behavior of rich and poor households in the absence of a 'screening device' such as progressive taxation and/or non-proportional housing demand.

3.3 Adding Taste Heterogeneity

So far, it has been assumed that households with identical preferences differ by income. This section extends the basic model by letting the households differ in both income $y \in [\underline{y}, \overline{y}]$, $0 < \underline{y}, \overline{y} < \infty$, and a parameter $\alpha \in [0, 1]$ describing their taste for housing. Income and taste are jointly distributed according to the density function $f(y, \alpha) > 0$. This extension is similar in spirit to Epple and Platt (1998).

The housing preference enters the utility function $U(h, b; \alpha)$ and the indirect utility

$$V_j = V(t_j, p_j\,; y, \alpha) = U(h_j^*, b_j^*; \alpha)\,. \tag{3}$$

Households with a larger preference parameter α are assumed to spend, ceteris paribus, more on housing than households with a small α. The housing demand function thus increases with α:

Assumption 3 (Housing taste)

$$\frac{\partial h^*}{\partial \alpha} = \frac{\partial h(t, p; y, \alpha)}{\partial \alpha} > 0 \quad \textit{for all } t,\ p,\ y \textit{ and } \alpha.$$

This specification of preference heterogeneity preserves income segregation within a subpopulation with identical preferences. Moreover, segregation of preferences emerges:

Proposition 4 (Preference segregation)
Consider a subpopulation with equal income y. If all communities are populated and exhibit different community characteristics, then all households choosing a community with higher housing prices have a weaker taste for housing than all households choosing a community with lower housing prices.

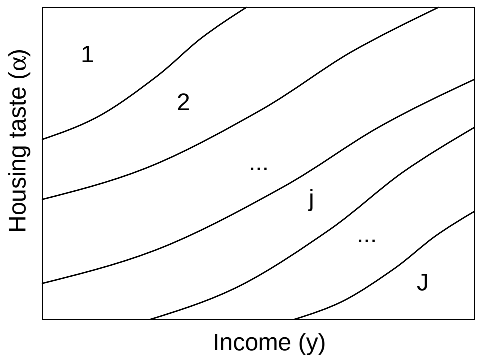

Figure 3: Simultaneous income and preference segregation. The areas denoted by $j = 1, ..., J$ show the attributes of the households that prefer community j.

Proof: The proof is analogous to Proposition 2 using the counterpart to Property 2,

$$\frac{\partial M}{\partial \alpha} = -\frac{\partial h^*}{\partial \alpha}\frac{1}{y \cdot r(y)} < 0\,. \quad \square$$

Simultaneous heterogeneity by incomes and tastes leads to a more realistic pattern of household segregation in a metropolitan area. Although income groups tend to gather, the segregation is not perfect. Figure 3 shows the resulting allocation of household types to communities. The households on the borders are indifferent between neighboring communities j. Community 1 with the lowest housing prices is populated by the poorest households with strong taste for housing, while the richest households with low housing taste are situated in community J with the lowest tax rate and the highest housing price. However, rich households with strong taste for housing prefer lower-priced communities and poor households with weak taste for housing group with relatively rich households in the lower-tax communities.

3.4 A Benchmark Case

This section presents the model with a specified utility function for homothetic preferences. Income segregation is therefore solely induced by the progressivity of the tax schedule. The derived indirect utility function will serve as a benchmark in the empirical study.

Household preferences are described by a Cobb-Douglas utility function

$$U(h, b; \alpha) = \alpha\, log(h) + (1 - \alpha)\, log(b)\,.$$

The resulting demand for housing

$$h^* = h(t, p\,; y, \alpha) = \alpha y[1 - t\, r(y)]p^{-1}$$

increases with α. The parameter $\alpha \in (0,1)$ can therefore be seen as a measure for housing taste as defined in Section 3.3.

The indirect utility function in community j is

$$V_j = V(t_j, p_j\,; y, \alpha) = k - \alpha\, log(p_j) + log(y) + log[1 - t_j\, r(y)]\,, \qquad (4)$$

where $k = \alpha\, log(\alpha) + (1-\alpha)\, log(1-\alpha)$.

3.5 Closing the Model

Sections 3.1 to 3.4 described individual location decisions given local tax rates and housing prices. This section outlines the complete general equilibrium model, in which local tax rates and housing prices arise endogenously from aggregate behavior. Individual location choices and community characteristics are thus simultaneously determined. As the empirical part of this papers solely relies on the location decision, this section shall only be a brief presentation.

Local public expenditure $G_j = G(g, n_j)$ is a function of exogenous public good provision g and the number of local residents n_j. In equilibrium, the jurisdiction's budget is balanced, i.e. local expenditure G_j equals local tax revenue T_j:

$$G_j = G(g, n_j) = T_j = \int_0^1 \int_{\underline{y}_j(\alpha)}^{\overline{y}_j(\alpha)} t_j\, y\, f(y,\alpha)\, dy\, d\alpha = n_j\, t_j\, Ey_j\,, \qquad (5)$$

where $\underline{y}_j(\alpha)$ and $\overline{y}_j(\alpha)$ are the lowest and highest income for the subpopulation with taste $\alpha \in [0,1]$ in community j. The double integral aggregates individual tax payments over the local population. Aggregate tax revenue can be expressed as the population weighted tax payment of a household with average income Ey_j. For a given local population, equation (5) determines the local tax rate t_j. The model therefore does not require us to assume a local public choice mechanism such as a majority vote.

Housing supply $H_j^s = H^s(p_j, L_j)$ in each community j is a function of the local housing price p_j and land area L_j. In equilibrium, p_j clears the local housing market:

$$H_j^s = H^s(p_j, L_j) = H_j^d = \int_0^1 \int_{\underline{y}_j(\alpha)}^{\overline{y}_j(\alpha)} h(t_j, p_j, g_j; y, \alpha) f(y,\alpha)\, dy\, d\alpha, \qquad (6)$$

where H_j^d is aggregate housing demand in community j.

A set of local characteristics (p_j, t_j), $j = 1, ..., J$, and an allocation of individual households across communities is an equilibrium if (a) all households choose their location to maximise their utility, (b) the housing market clears in all communities, and (c) the budget is balanced in all communities. Hodler and Schmidheiny (2005) proof existence of an asymmetric equilibrium in the model with taste homogeneity.[9] In equilibrium, some communities show low tax rates and high housing prices while other communities show high tax rates and low housing prices. The low-tax communities attract richer households which allows them to finance the public good and balance their budget with lower tax rates than the (poor) high-tax communities.

4 Alternative Explanations for Income Segregation

The theoretical model in the previous section relates income segregation solely to institutions (in the case of taxes) and markets (in the case of housing prices). There are other mechanisms which potentially explain income segregation in metropolitan areas. This section discusses two leading alternative explanations which are considered in the empirical implementation.

4.1 Income Segregation and Distance

The classic framework for the study of residence choice and housing prices in urban areas was developed by von Thünen (1826) and formalized by Alonso (1964), Beckmann (1969), Muth (1969) and Mills (1972). In the basic model, a monocentric city is characterized by a single central business district where all job opportunities are located. Fujita (1989, chapter 2) nicely shows how this basic structure explains segregation of the population by incomes: If commuting costs are monetary expenditures, the rich locate farther from the center than the poor. The opposite holds if commuting costs are time lost while travelling. The former proposition was often used to explain the affluent suburbs in U.S. cities by high monetary commuting costs. The latter proposition corresponds with the rich Japanese city centers and the fact that Japanese firms used to pay the monetary commuting costs. The von Thünen approach seems to have strong explanatory power even in its basic form. However, Wheaton (1977, p. 631) already concluded that this type of model "empirically contributes little to the explanations of American location-income patterns" and that fiscal and other externalities are more important.

4.2 Income Segregation and Social Interactions

There is growing literature in economics and mathematical sociology that models the emergence of segregation from social interactions. Schelling (1969, 1971)

[9] Proof of existence and uniqueness of equilibria in multi-community models with taste heterogeneity is notoriously difficult (see e.g. Epple and Platt, 1998). Hodler and Schmidheiny present simultated equilibria in a calibrated version of the model with taste heterogeneity.

showed how an (even minor) intrinsic preference of white households for living with other white households leads to endogenous racial segregation. Bénabou (1993, 1996a, 1996b) and Durlauf (1996a, 1996b) integrate models of fiscal decentralization with models of social interactions. They assume that the productivity of schooling depends on the social mix of the neighborhood, and that investment in education depends on the local tax base. The income distribution of the neighborhood is then the determinant of households' location decisions but also their consequence. This feedback effect of the aggregate outcome on the individual decisions is typical for models with social interactions. In equilibrium, rich households cluster in order to supply a better education to their offspring, while poor households stay away because of high housing costs. There is only little empirical evidence on the role of social interactions for the emergence of (income) segregation. Bayer, McMillan, and Rueben (2004) establish neighborhood effects in the location choice of households. Ioannides and Zabel (2002, 2003) show that both the demographic composition of the neighborhood and the actual behavior of the closest neighbors are important determinants for the choice of a dwelling.

5 The Econometric Model

If spatial income segregation is the result of individual households' decisions it can be studied by looking at the decisions when they are made. Households do not daily decide upon their place of residence. There are specific moments in any individual's life when the decision about where to live becomes urgent. Such moments occur for example when an individual leaves college to start a first job, when a couple decides to move together or when the family size changes. Career opportunities later in life and changing budget constraints further trigger the decision to move. These are the moments when households have to evaluate potential residential locations. Limiting the analysis to moving households therefore eliminates the bias when including households that stay in a per-se sub-optimal location because of high monetary and psychological costs of moving. However, the limitation to moving households introduces a potential selection bias when the unobserved individual factors that trigger the decision to move are correlated with the unobserved individual taste for certain locations.

The location choice in the theoretical model in Section 3 naturally leads to a multinomial response model based on random utility maximization (see McFadden 1974, 2001 and Train 2003). The indirect utility V_{nj} of a household n in location j is the sum of a systematic and a stochastic part

$$V^*_{nj} = V_{nj} + \varepsilon_{nj} \,.$$

V_{nj} is a deterministic function of observable household and community characteristics and ε_{nj} is a household and community specific error term. Equivalent to equation (2) in the theoretical model, a household n chooses community j as its place of residence if it offers the highest value of indirect utility, i.e.

$$V^*_{nj} \geq V^*_{ni} \quad \text{for all } i \in C = (1, ..., J) \ . \tag{7}$$

5.1 Functional Form and Identification

The indirect utility function in the theoretical part guides the choice of systematic factors in the indirect utility function

$$V_{nj} = V(t_{nj}, p_j, a_j, y_n, h_n) \,, \tag{8}$$

where t_{nj} is the average income tax rate of household n in community j, p_j is the housing price in community j, a_j indicates further location specific characteristics, y_n is household income and h_n indicates further household characteristics. Note that from the point of view of an individual household, the community characteristics p_j, t_{nj} and a_j are exogenous, although they are the (long-term) aggregate of the agents' decisions.

The functional form of the deterministic part (equation 8) of the latent variable needs to be specified for the empirical implementation. Starting point is the indirect utility function (4) from the benchmark case presented in Section 3.4,

$$V_{nj} = \beta_0 + \beta_1 \, log(1 - t_{nj}) + \beta_2 \, log(p_j) + \beta_3 \, log(y_n) \,.$$

In this homothetic specification, income segregation can only be generated through the tax rate t_{nj}, which depends on the community j, but also on the household n through the progressivity of the tax scheme and through household type specific deductibles. The theoretical model introduces two additional mechanisms for segregation: income elasticity of housing below unity and heterogenous taste for housing. I introduce these two mechanism by varying the coefficient β_2 with household income and other household characteristics:

$$\beta_2 = \alpha_{20} + \alpha_{21} \log(y_n) + \alpha_{22} h_n \,.$$

Housing consumption as a function of disposable income $y_{dn} = y_n(1 - t_{nj})$ is implied by Roy's identity (taking $\beta_3 = \beta_1$). It is indeed monotonic in the taste parameter α_{22}: $\partial h / \partial \alpha_{22} = -\alpha_{22}/(\alpha_{21} log(p_j) + \beta_3) \cdot y_{dn}/p_j$ (satisfying assumption 3). The elasticity of housing consumption with respect to disposable income $\varepsilon_{h,y_{dn}} = (\beta_2 + \alpha_{21})/\beta_2$ is below 1 as long as the effect of prices on utility is negative, $\beta_2 < 0$, and decreasing in absolute value in income, $\alpha_{21} > 0$ (satisfying assumption 1).

I also allow the effect of taxes to vary with household income:

$$\beta_1 = \alpha_{10} + \alpha_{11} \log(y_n) \,.$$

Location specific constants δ_j capture all observed and unobserved local characteristics that attract households to a location. The effect of any observed community characteristic such as housing prices, p_j, cannot be distinguished from the constant of this community and α_{20} is thus not identified. Furthermore, factors that shift the indirect utility of all alternatives in the same way are not identified, hence β_3 cannot be estimated. The resulting identified *base model* is therefore

$$V_{nj} = \delta_j + \alpha_{10}\, log(1 - t_{nj}) + \alpha_{11}\, log(1 - t_{nj}) log(y_n) + \alpha_{21}\, log(p_j) log(y_n) + \alpha_{22}\, log(p_j) h_n \,.$$

In the *full model* of the econometric specification, I allow for segregation through the interaction of additional location specific characteristics a_j ={share rich, share foreigners, share foreign pupils, distance, distance^2} with household income $log(y_n)$ and further household characteristics h_n ={household size, foreign status, children}.

5.2 Modelling the Stochastic Part

The stochastic part ε_{nj} captures all factors of community choice that are hidden from the researcher but known to the household. It therefore represents all *unobserved* factors such as more detailed socio-demographic information about the household as well as all *unobservable* factors such as the household members' attachment to a certain location. Several specifications are used and compared in the empirical analysis.

The first specification assumes that the error terms follow independently and identically an extreme value distribution. The cumulative distribution function is

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}} \,.$$

This leads to the *conditional logit* model. The probability that household n chooses community j is

$$P_{nj}(\theta) = \frac{e^{V_{nj}}}{\sum_{i=1}^{J} e^{V_{ni}}} \,, \tag{9}$$

where V_{ni} is the deterministic part of the utility of household n in community i and $\theta = \beta$ is the set of parameters to be estimated. The independence of the error term implies that a household's stochastic, i.e. unobserved, preference for a certain location is fully independent of its stochastic preference for other locations. The strong and unpleasant consequences of this assumption are discussed in the literature as *independence of irrelevant alternatives* (IIA).

The *nested logit* model is a generalization of the conditional logit model that avoids IIA by allowing a specific pattern of correlations across the error terms (see McFadden, 1984). The vector of all location specific error terms $\varepsilon_n = (\varepsilon_{n1}, ..., \varepsilon_{nJ})$

follows the generalized extreme value distribution (GEV) introduced by McFadden (1978):

$$F(\varepsilon_n) = e^{[-\sum_k \left(\sum_{i \in C_k} e^{-\varepsilon_{ni}/\lambda_k}\right)^{\lambda_k}]} .$$

The choice set $C = (1, ..., J)$ is divided into K mutually exclusive subsets C_k, called nests. The unobserved portions of utility ε_{ni} are correlated within the same nest k and independent across nests. The parameter λ_k captures the correlation within nest k, i.e. $corr(\varepsilon_{ni}, \varepsilon_{nj}) = 1 - \lambda_k^2$ for all $i, j \in C_k$ (see Ben-Akiva and Lerman, 1985). The extreme case $\lambda_k = 1$ means that there is no correlation within nest k. The nested logit model is consistent with random utility maximization if (but not only if; see Börsch-Supan, 1990) $\lambda_k \in [0, 1]$. Setting all λ_k to unity leads to the conditional logit model. The probability that household n chooses location j is

$$P_{nj}(\theta) = \frac{e^{V_{nj}/\lambda_l} \left(\sum_{i \in C_l} e^{V_{ni}/\lambda_l}\right)^{\lambda_l - 1}}{\sum_{k=1}^{K} \left(\sum_{i \in C_k} e^{V_{ni}/\lambda_k}\right)^{\lambda_k}} , \tag{10}$$

where l is the nest of location j and $\theta = (\beta, \lambda)$.[10] The nested structure of the error term can be looked at as the result of a two-part choice: households choose a certain nest first and afterwards an alternative within the nest. In the empirical study the first step is naturally the decision whether to move to a borough in the center community or to move to a community in the periphery. Households with a large unobserved preference for a location in the periphery (center) therefore also have a higher preference for all other communities in the periphery (center).

The *multinomial probit* model enables a more flexible specification of the error term compared to the previous two models. The vector of error terms across alternatives is assumed to follow a J-variate normal distribution

$$\varepsilon_n \sim N(0, \Omega) ,$$

where Ω is the $J \times J$ variance-covariance matrix.[11] This study uses a very parsimonious specification of Ω. Following Bolduc (1992) and Bolduc, Fortin and

[10] Note that this form of the likelihood function is directly derived from the random utility model and the generalized extreme value distribution. Some software packages, e.g. the `nlogit` command in Stata, use a slightly different likelihood function in their implementation of nested logit. These likelihood functions are not consistent with random utility maximisation. See Hensher and Greene (2002) for a critical discussion.

[11] This general form allows for all possible correlation patterns across the unobserved part of utility. This flexibility, however, comes at a price: the estimation of multinomial probit models is numerically demanding (see Section 5.3) and the general variance-covariance needs to be restricted for both theoretical and practical reasons. Due to the fact that the agents only care about the utility differences across alternatives, Ω needs normalizing and only a maximum of $[(J-1)J/2] - 1$ parameters can be estimated compared to the $J(J+1)/2$ distinct elements in Ω (see Train 2003). In the case of e.g. 17 alternatives there are still 135 parameters to be estimated. These parameters are in practice hardly identified.

Gordon (1997),[12] the alternative specific error terms follow a first order spatial autoregressive process (SAR)[13]

$$\varepsilon_n = \rho\, W\, \varepsilon_n + \xi_n\,,$$

where $\xi_n \sim N(0, I)$ and $\rho \in (-1, 1)$ is a parameter to be estimated. W is an exogenous $J \times J$ weighting matrix where the weight w_{ji} is a decreasing function of the distance d_{ij} between community j and i

$$w_{ji} = \frac{1/d_{ji}}{\sum_{s=1}^{J} 1/d_{js}}$$

and satisfies $w_{ji} = w_{ij}$, $w_{ii} = 0$ and $\Sigma_s w_{is} = 1$ by construction. The variance-covariance of the error term can be derived as

$$\Omega(\rho) = (I - \rho\, W)^{-1}(I - \rho\, W)^{-1}$$

because $\rho \in (-1, 1)$ guarantees the nonsingularity of $(I - \rho\, W)$ (see Berman and Plemmons, 1994, p.133). The probability that household n chooses community j is

$$P_{nj}(\theta) = Prob\,[\varepsilon_1 - \varepsilon_j > V_{nj} - V_{n1}, \ldots, \varepsilon_J - \varepsilon_j > V_{nj} - V_{nJ}]\,, \qquad (11)$$

where $\theta = (\beta, \rho)$. The above spatial pattern means that households with a strong unobserved taste for a certain community also like other communities geographically close to that community.

5.3 Estimation

The conditional and nested logit models are estimated using maximum likelihood (ML) and full information maximum likelihood (FIML) respectively. The log likelihood function is

$$log\, \mathcal{L}(\theta) = \sum_{n=1}^{N} \sum_{j=1}^{J} z_{nj}\, log\, P_{nj}(\theta)\,, \qquad (12)$$

where $z_{nj} = 1$ if the household n chooses location j and $z_{nj} = 0$ otherwise. The choice probabilities P_{nj} of the conditional logit and nested logit model are defined in equations (9) and (10), respectively. The maximum likelihood estimator $\hat{\theta} = (\hat{\beta}, \hat{\lambda})$ is consistent, asymptotically efficient and normally distributed.

The multinomial probit model is estimated with maximum simulated likelihood (MSL, see Hajivassiliou and Ruud, 1994) based on equation (12). The

[12]Bolduc, Fortin and Fournier (1996) present one of the rare applications of SAR in multinomial response models. They use a slightly different specification and mix the multivariate normal SAR process with an extreme value distribution.

[13]See Anselin and Florax (1995) for a general treatise of SAR processes.

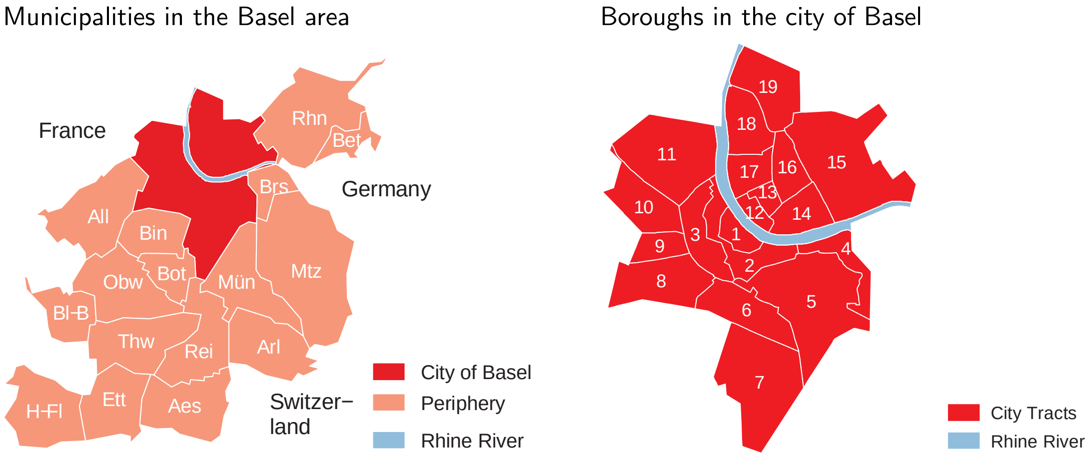

Figure 4: Municipalities and city boroughs in the Basel metropolitan area.

calculation of the likelihood requires the integration of a 16-variate normal distribution. As there is no analytic solution to this problem numerical integration routines or simulation methods are applied. A standard method is the Geweke-Hajivassiliou-Keane GHK choice probability simulator (see Geweke, Keane and Runkle, 1994 and Börsch-Supan, and Hajivassiliou, 1993). GHK simulates the choice probabilities P_{nj} in equation (11) by recursively drawing from univariate normal distributions. This study uses $R = 1000$ pseudo-random draws in each dimension. The properties of the MSL estimator $\hat{\theta} = (\hat{\beta}, \hat{\rho})$ are equivalent to standard ML if the number of draws R grows faster than $\sqrt{N}$ (see e.g. Train, 2003).[14]

6 Data

The empirical investigation is based on non-public household data from the Tax Administration of the Canton of Basel-Stadt. The data contain information of all households in the city of Basel that moved within the city or from the city to a municipality in the periphery in the year 1997.

The decision maker in the theoretical model is a household.[15] The choice set of

[14]All estimations are performed with the author's own programs in MATLAB. A toolbox with programs for conditional logit, nested logit, multinomial probit and mixed logit models is available on request. The Newton-Raphson algorithm with the Broyden-Fletcher-Goldfarb-Shanno method (BFGS) for updating the hessian matrix was used for numerical maximization. All parameters, including the coefficients of the correlation structure, have been appropriately scaled during optimization. The estimation of the multinomial probit model runs approximately 70 hours on a Sun Fire V880.

[15]Households are operationalized as all individuals who moved from a common old address to a common new address: families in a narrower sense, married and unmarried couples as well as people who simply share a flat. Married couples that move from single households to the same street address are treated as one household. Unmarried individuals who move to the same

these households is restricted to the city of Basel and a circle of 16 municipalities around it.[16] This leaves 7,872 households with 11,540 members in the data set. The included municipalities belong to three different cantons, Basel-Stadt (BS), Basel-Land(BL) and Solothurn (SO) and thus exhibit great variability in tax levels and tax schedules. In the estimation of the full specification, the center municipality is divided into 19 city boroughs. See Figure 4 for a map of the included municipalities and city boroughs.

In addition to income, there is some limited demographic information on these households:

- *inc*: total gross income of all household members,[17]
- *hhsize*: number of persons living in the household,
- *child*: one or more under-age children in household,
- *foreign*: primary earner is not a Swiss citizen.

The universe of moving households with 11,540 members (7.6% of the city populatoin) differs from the population in the city of Basel. The average household size of moving household is 1.5 while it is 1.8 in the whole population (Census 2000). The dataset reports an average of 1.67 under-age children (as defined by the tax assessment) in 19% of all households, compared to 1.69 children (unmarried and below 20) in 18% of households in the resident population (Census 2000). The share of foreign moving households is 20% while the share of foreigners in the city of Basel is 24% (Census 2000). The fraction of households with an income above CHF 75,000 is 26% and exactly the same for movers (based on the cantonal tax assessment in my data) and residents (based on the federal tax assessment 1997/98). However, the fraction of households with incomes below CHF 40,000 is 37% for the movers but only 16% for the whole city population.[18] The reference person in moving households is with an average of 36 years considerably younger than in the resident population with 51 years.

street address are treated as independent households.

[16]These municipalities are defined as all municipalities of whom more than 36% of the working population is commuting to the center (Census 1990). This admittedly arbitrary cut-off point leads to a well-shaped geographic area and a tractable number of choice alternatives. The five smallest municipalities are omitted as they are not covered in the tax schedule data. A narrow definition of the metropolitan area is also appropriate because Tiebout type models ignore the location of the work place. When households decide upon their place of residence on a national or global scale, job opportunities are naturally very important. In narrow metropolitan areas, however, it is reasonable to assume that any municipality is a feasible place of residence for households whose members are working in the central business district. Changing the choice set did not qualitatively change the results of the analysis.

[17]According to the last tax assessment before moving. The relevant gross income would be the gross income after moving, which is not available. Income before moving is a good proxy if a household's decision to move does not coincide with a change in its income.

[18]Note that federal data do not report very low income households which are exempted from federal income taxes.

Table 1: Characteristics of in-movers and locations.

	in-movers			locations				
	moved in	mean inc	mean hhsize	tax mid[a]	tax high[a]	median rent	share rich[b]	foreign pupils[c]
Whole Area	7872	61.6	1.5					
City of Basel	6370	59.3	1.5	6.2	24.4	197	0.26	0.41
- Alt Grossbasel (1)	134	77.8	1.3			203	0.39	0.21
- Vorstaedte (2)	191	66.5	1.2			200	0.34	0.34
- Am Ring (3)	440	59.3	1.4			197	0.30	0.25
- Breite (4)	328	60.5	1.4			202	0.19	0.35
- St. Alban (5)	279	72.0	1.5			201	0.36	0.18
- Gundeldingen (6)	820	54.6	1.4			198	0.21	0.58
- Bruderholz (7)	210	93.4	1.8			220	0.44	0.09
- Bachletten (8)	439	67.1	1.5			206	0.33	0.11
- Gotthelf (9)	247	60.1	1.3			204	0.24	0.19
- Iselin (10)	696	51.5	1.4			191	0.19	0.38
- St. Johann (11)	728	53.3	1.5			191	0.21	0.56
- Alt Kleinbasel (12)	112	51.2	1.3			192	0.22	0.38
- Clara (13)	158	55.8	1.4			190	0.21	0.59
- Wettstein (14)	220	63.8	1.5			192	0.25	0.18
- Hirzbrunnen (15)	189	67.1	1.8			192	0.25	0.21
- Rosental (16)	170	54.8	1.4			192	0.19	0.72
- Matthaeus (17)	661	53.7	1.5			188	0.21	0.68
- Klybeck (18)	252	54.3	1.6			188	0.20	0.58
- Kleinhueningen (19)	96	59.1	1.5			188	0.20	0.52
Periphery	1502	71.3	1.5					
- Hofstet.-Flueh (H-FL)	18	64.9	1.9	5.1	26.4	190	0.40	0.10
- Bettingen (Bet)	9	70.0	1.3	4.7	20.9	220	0.47	0.13
- Riehen (Rhn)	280	83.9	1.6	3.7	21.8	206	0.41	0.13
- Aesch (Aes)	57	63.0	1.5	5.3	23.3	213	0.33	0.25
- Allschwil (All)	251	69.3	1.6	5.1	21.8	207	0.32	0.17
- Arlesheim (Arl)	56	57.6	1.3	4.9	21.9	215	0.37	0.19
- Biel-Benken (Bl-B)	18	88.6	1.6	4.9	20.9	226	0.48	0.04
- Binningen (Bin)	165	73.4	1.4	5.0	21.8	205	0.36	0.23
- Birsfelden (Brs)	98	52.4	1.3	5.5	23.2	200	0.24	0.37
- Bottmingen (Bot)	43	76.4	1.6	5.1	22.1	206	0.47	0.18
- Ettingen (Ett)	24	61.5	1.7	5.3	23.5	197	0.41	0.08
- Muenchenstein (Mn)	92	59.0	1.5	5.2	22.3	198	0.31	0.18
- Muttenz (Mtz)	114	63.3	1.6	5.3	22.7	192	0.35	0.16
- Oberwil (Obw)	80	77.0	1.4	4.9	21.1	211	0.42	0.15
- Reinach (Rei)	151	72.2	1.5	5.1	22.5	212	0.36	0.14
- Therwil (Thw)	46	91.7	1.8	5.1	22.7	207	0.38	0.12

[a] Cantonal and communal average income tax rate for married couple with two children and income of CHF 60,000 and CHF 500,000 respectively.
[b] Share of households with annual income above CHF 75,000.
[c] Share of pupils with non-Swiss citizenship in primary schools.

Table 1 reports descriptive statistics of selected variables by chosen locations. From the total of 7,872 households that stayed within the choice set, 4/5 moved within the center municipality whereas only 1/5 moved to one of the 16 communities in the periphery. The income of the latter households was on average 20% higher than that of the ones remaining in the center. The very low tax communities attracted particularly rich households. The size of households (as well as the number of children, not reported) is about the same in the center and in the periphery. However, there is substantial variation across the municipalities in the periphery and across city boroughs.

As this study uses location specific constants, only variables that are potentially accountable for residential income segregation are considered. I collected the following location specific data from various sources:

- *tax* (household and municipality specific): hypothetical tax rate for totalled cantonal (state) and municipal income taxes, reflecting municipality/state specific progressive tax schemes, municipal tax shifters and household/municipality/state specific tax deductions,
- *rent* (municipalities and city boroughs): average offer price per m^2 for a rented flat,
- *share rich* (municipalities and city boroughs): share of resident households with income above CHF 75,000,
- *share foreigners* (municipalities and city boroughs): population share of non-Swiss citizens,
- *share foreign pupils* (municipalities and city boroughs): share of non-Swiss pupils at primary school,
- *distance* (municipalities and city boroughs): distance in km between a municipality and the central business district.

See the appendix for a detailed description of variables and data sources. Descriptive statistics of selected location specific variables are presented in Table 1.

7 Results

7.1 Estimating the base model

This section presents the estimates of the base model reported in Table 2. The base model is a direct implementation of the theoretical model in Section 3. In this section, I model the location choice of moving households as a choice among 17 alternative communities: the city of Basel and 16 peripheral communities. Three specifications of the error term structure are reported in the respectively labelled columns: the conditional logit model, the nested logit model and the multinomial probit model with a spatial autoregressive process (SAR).

Table 2: Estimation of the Base Model.

	Conditional Logit	Nested Logit	Mult. Probit, SAR[d]
Average Income Tax			
log(1-tax)[a]	17.8 (2.6) ***	15.7 (2.2) ***	13.8 (1.7) ***
log(1-tax)×[log(inc)-log($\overline{inc}$)]	11.3 (2.0) ***	11.0 (1.8) ***	9.8 (1.6) ***
Housing Price			
log(rent)×[log(inc)-log($\overline{inc}$)]	-0.13 (0.7)	0.28 (0.5)	0.00 (0.4)
log(rent)×hh size	0.71 (0.7)	0.04 (0.5)	0.14 (0.2)
Correlation Structure Coefficients			
Log-sum periphery (λ)		0.58 (0.1) **,c	
Spatial autocorrelation (ρ)			0.25 (1.1)
Community Specific Constants			
City of Basel (reference)	0	0	0
Mean periphery constants[b]	-5.18	-3.73	-4.10
Log likelihood	-7440.0	-7438.6	-7444.4
Number of alternatives	17	17	17
Number of observations	7872	7872	7872

Standard errors in brackets, ***, **, * significance at the 0.1%, 1% and 5% level.

[a] The coefficient gives the effect for a household with mean income (CHF 61,612)

[b] Mean of 16 individual community effects. All effects are significantly smaller than 0 (city).

[c] Tested against 1, i.e. the error terms are independent across communities in the periphery.

[d] Slope coefficients and constants of the multinomial probit are multiplied by $\pi/\sqrt{6}$ such that the error term variance of the multinomial probit equals the one in the logit models.

The specification of the error term is discussed first. The significant log-sum coefficient in the nested logit model shows the violation of the IIA assumption in the conditional logit model.[19] The estimate $\hat{\lambda} = 0.58$ implies that the error terms across municipalities in the periphery are positively correlated. This means that households with an unobserved taste for a municipality in the periphery also prefer other municipalities in the periphery. The estimated spatial autocorrelation coefficient $\hat{\rho}$ in the multinomial probit model is not significantly different from zero and does not support the proposed spatial error term structure. The nested logit model is thus the preferred model. The following discussion relates to the results of the nested multinomial logit model.

The coefficient for $log(1 - tax)$ gives the effect of the tax rate on the indirect utility function for a household with an average income of CHF 61,612. It is significantly positive on the 0.1% level and confirms that progressive taxes have a negative effect on utility. It also implies, through the progressivity of the tax scheme, that the effect is larger for rich households than for poor ones. The significantly positive coefficient of the interaction with $[log(inc) - log(\overline{inc})]$ implies that the tax effect increases with income beyond what is implied by the progressivity of the tax scheme. The quantitative implications are discussed in the full

[19] In this section, the center municipality is a nest of its own, called a degenerate nest with $\lambda_k = 1$.

model in the following subsection.

The interactions of the housing rent with income and household size turn out to be insignificant. The base model gives therefore no indication that the housing price effect varies with either income or household size. Housing prices thus cannot explain the income segregation of moving households. Note that this does not rule out that housing prices have a strong (negative) effect, common to all households, on the attractiveness of a municipality. This effect is part of the municipality specific constant and therefore not identified.

7.2 Estimating the extended model

This section presents the results of estimations with an extended choice set and several additional factors. The location choice of households is now among 35 alternatives: 16 peripheral communities and 19 boroughs in the center municipality. The subdivision of the large center municipality allows to consider the substantial intra-municipality variation especially of social context variables. This permits to disentangle the effect of social context variables from the effect of the local tax, which is constant within the center municipality.[20] Table 3 reports the estimates of nested logit models for the *base* and the *full* specification in the respectively labelled columns.

The *base model* replicates the results in the previous section, now considering intra-municipality variation of housing prices and allowing for differing error terms for the 19 city tracts. The results are almost identical to the previous findings in Table 2. Note in particular that the error terms across city boroughs are almost perfectly correlated (log-sum coeffient for the center is significantly different form 1 but not from 0), supporting the treatment of the city as a single unit in the previous section.

The *full model* controls the base model for potential segregation due to social interactions and distance. Note that both log-sum coefficients are significantly different from unity. The correlation across error terms is $1 - 0.51^2 = 0.74$ between boroughs in the center and $1 - 0.62^2 = 0.62$ between municipalities in the periphery. This means that households with an unobserved taste for a location in the center (periphery) also prefer other locations in the center (periphery, respectively).

The coefficients for the local income tax as well as for its interaction with household income are significantly different from 0. Their effect is, however, smaller than in the base model. The significantly positive coefficient of the interaction with $[log(inc) - log(\overline{inc})]$ implies that the tax effect increases with income

[20]Income tax rates and the share of rich households are highly correlated across municipalities as shown in Section 2. This poses a potential multi-collinearity problem. The consideration of intra-municipality variation helps to overcome this problem.

Table 3: Nested Logit Estimation of the Full Model.

	Base model			Full model		
Average Income Tax						
log(1-tax)[a]	16.0	(2.2)	***	12.5	(2.4)	***
log(1-tax)×[log(inc)-log($\overline{inc}$)]	11.2	(1.8)	***	5.6	(2.0)	**
Housing Price						
log(rent)×[log(inc)-log($\overline{inc}$)]	0.19	(0.4)		0.29	(0.3)	
log(rent)×hh size	-0.04	(0.1)		-0.53	(0.2)	*
Social Interactions with Residents' Income						
share rich×[log(inc)-log($\overline{inc}$)]				0.44	(0.2)	**
Social Interactions with Residents' Alien Status						
share foreigners×foreign				1.63	(0.3)	***
share foreigners×[log(inc)-log($\overline{inc}$)]				-0.21	(0.1)	*
share foreign pupils×child[b]				-0.42	(0.1)	**
share foreing pupils×child×[log(inc)-log($\overline{inc}$)]				-0.38	(0.1)	**
share foreing pupils×child×foreign				1.12	(0.2)	***
Distance from CBD in km²						
distance/100×[log(inc)-log($\overline{inc}$)]				3.52	(1.2)	**
(distance/100)^2×[log(inc)-log($\overline{inc}$)]				-0.39	(0.1)	**
distance/100×hh size				6.08	(1.4)	***
(distance/100)^2×hh size				-0.31	(0.1)	*
Correlation Structure Coefficients						
Log-sum (λ) center	0.07	(0.1)	***,c	0.51	(0.1)	***,c
Log-sum (λ) periphery	0.56	(0.1)	**,c	0.62	(0.1)	**,c
Location Specific Constants						
Mean city constants[d]	0			0		
Mean periphery constants[e]	-3.48			-2.30		
Log likelihood	-24881			-24670		
Number of alternatives	35			35		
Observations	7872			7872		

Standard errors in brackets, ***, **, *, ~ significance at the 0.1%, 1% and 5% level.

[a] The coefficient gives the effect for a household with mean income ($\overline{inc}$ = CHF 61,612)

[b] The coefficient gives the effect for a household with mean income and Swiss citizenship.

[c] Tested against 1, i.e. the error terms are independent across communities in the periphery.

[d] Mean of 19 constants of city boroughs. Mean is set to zero for identification.

[e] Mean of 16 constants of peripheral communities. All effects are smaller than 0 (center).

beyond what is implied by the progressivity of the tax scheme. When income doubles, the effect of taxes increases by $log(2) \cdot 5.6/12.5 = 31\%$ on top of the effect from the progressivity of the tax scheme. This study can only speculate about the reasons for the high priority of local taxes in the location decision of high-income households. High-income households might for example rely more on advice from professional real-estate and investment consultants who propose moving to tax heavens as a form of tax optimization.

The quantitative impact of the tax rate and its role for income segregation is best explained using an example: Consider a two-child family with average income of CHF 61,612 that compares the borough Iselin in the city of Basel (labelled "10" in the right map of Figure 4) to the neighboring municipality of Allschwil in the canton of Basel-Land (labelled "All" in the left map of Figure 4). The tax rates it faces are 6.2% in Iselin and 5.1% in Allschwil. The model predicts - ceteris paribus, i.e. all locations have identical characteristics except for taxes - that the average income household moves to the borough of Iselin with a probability of 2.1% and to the municipality of Allschwil with a probability of 3.6%; hence the odds are 1.7 in favor of Allschwil.[21] Consider now a rich two-child family with an income of CHF 500,000. This family faces tax rates of 24.4% and 21.8% in Iselin and Allschwil, respectively. For this family the model predicts - again, ceteris paribus - odds of 3.7 in favor of Allschwil (probability of 1.5% for Iselin and 5.4% for Allschwil) which is far higher (odds ratio of 2.2) than for the average income household.[22] Note that considering all factors (including location specific constants), the odds for a Swiss two-child family with average income are in favor of Iselin (odds of 0.4) and for a rich Swiss family in favor of Allschwil (odds of 1.5, odds ratio of 4).

The interaction of housing rent with household income is not significant. However, the interaction of housing rents with household size is significantly negative. Assuming that the (not identified) base effect from housing prices is negative this does mean that the price effect increases in absolute value with the size of the household. In other words: larger households are more concerned about housing prices than smaller ones.

Social interaction as mechanism for residential income segregation is considered in three context variables. Firstly as the share of rich households[23] that already live in the location, secondly as the share of foreigners in the location and thirdly as the share of foreign pupils at the primary schools. The interaction of the share of rich households with household income is significantly positive. This means that rich households are more attracted by locations with a large share of rich households than are poor households.[24]

The interaction of the local share of foreigners with the foreign status of the household is significantly positive. Foreign households therefore tend to locate close to other foreign households. The interaction with household income

[21] The difference in indirect utility between Allschwil and Iselin due to the tax differential is $[log(1-0.051) - log(1-0.062)] \cdot 12.5 = 0.15$ for the average income family.

[22] The difference in indirect utility between Allschwil and Iselin due to the tax differential is $[log(1-0.218) - log(1-0.244)] \cdot (12.5 + 5.6 \cdot log(500000/61612)) = 0.82$ for the rich family, hence 5.5 times larger than for the average income family.

[23] Typically, average or median income is used as context variable. As this information is not available, I use the share of the top income group.

[24] The interaction of household income and (average) income of the residents' income can be derived by including the squared difference between household and location income in the indirect utility function. The two resulting squared terms are not identified.

is significantly negative, implying that rich households (Swiss or foreign) shun neighborhoods with a large foreign population.

The share of foreign pupils captures a further type of social interactions. School quality may be strongly determined by peer group effects in class. In Switzerland, parents are concerned about pupils from low educated backgrounds and non-native speaking pupils as their childrens' peers. These characteristics are not directly observable but highly correlated with the observed alien status of pupils. The interaction of the share of foreign pupils with a dummy variable for households with children is significantly negative. This means that Swiss households with children and average income are seeking school districts with a small share of foreign children. The effect sharply increases with income: it triples ($-0.42 - 0.38 \cdot log(500000/61612) = -1.22 \approx 3 \cdot -0.38$) for a rich Swiss family with an income of CHF 500,000 and falls to zero for a very poor family with an income of CHF 20,000. The effect also changes its sign for foreign families with average income ($-0.42 + 1.12 = 0.70$). This significantly positive effect reveals that foreign average income families tend to locate in school districts with many other foreign pupils. Rich foreign families with income above CHF 380,000, however, also prefer districts with fewer foreign pupils.

Distance from the central business district (CBD) is interacted with household income and household size. I also include distance squared to allow for a non-linear effect of distance. The two interaction effects with household income are significantly different from zero. The interaction effect of distance and household income is positive up to 9 km from the CBD (strongest at $3.52/(2 \cdot 0.39) = 4.5$ km) and negative thereafter. Rich households are therefore more likely to move to the first ring of municipalities around the center, sometimes called the "Speckgürtel", than poor households. The interaction of distance and distance squared with household size is also significant. The interaction effect is always positive and strongest towards the edge of the metropolitan area ($6.08/(2 \cdot 0.31) = 9.7$ km from the CBD), meaning that larger households care less about distance than smaller households. This could be explained by the fact that larger households usually include children and/or a non-working spouse which depend less on the CBD than for example a working single.

7.3 Assessing the Relative Importance of Taxes, Housing Costs and Social Interactions for Income Segregation

The relative importance of the factors for income segregation in the full model is not immediately readable from the estimated coefficients. I assess the relative importance by comparing the predicted segregation pattern in counterfactual simulations. Given the attributes of the households and the locations, the model

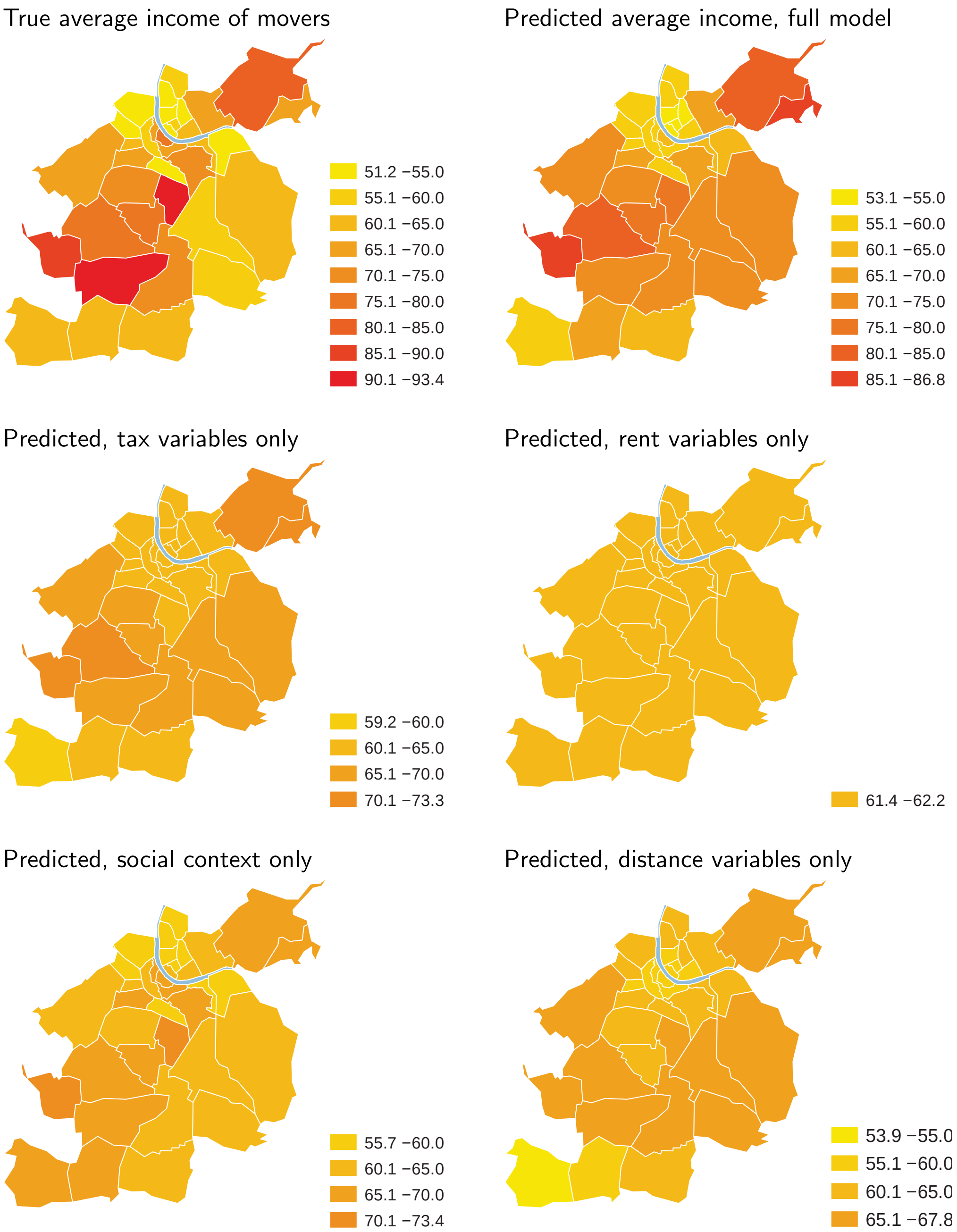

Figure 5: Observed and Predicted Average Income of Movers in Thousand CHF.

predicts the average income of the households that move to a particular municipality or city borough j:

$$\sum_{n=1}^{N} y_n \, P_{nj}(\hat{\theta}) \Big/ \sum_{n=1}^{N} P_{nj}(\hat{\theta}) \, .$$

Table 4 shows the predicted average incomes using the full model and using only a distinct causal channel, i.e. setting the other coefficients to zero. The spatial pattern of these income differentials is displayed in Figure 5. The full

Table 4: Observed and Predicted Average Income (in thousand CHF) of Movers.

	observed	full model	taxes[a]	rent[a]	social[a]	distance[a]
Whole area	61.6	61.6	61.6	61.6	61.6	61.6
City of Basel (BS)						
- Alt Grossbasel	77.8	61.0	60.3	61.8	69.0	55.4
- Vorstädte	66.5	57.6	60.3	61.7	65.1	55.4
- Am Ring	59.3	56.6	60.3	61.6	64.4	55.4
- Breite	60.5	61.3	60.3	61.7	59.6	64.9
- St. Alban	72.0	68.2	60.3	61.7	68.4	63.2
- Gundeldingen	54.6	56.2	60.3	61.6	57.6	61.5
- Bruderholz	93.4	77.9	60.3	62.2	73.4	67.8
- Bachletten	67.1	66.6	60.3	61.8	68.6	61.5
- Gotthelf	60.1	59.8	60.3	61.8	63.4	59.6
- Iselin	51.5	57.2	60.3	61.5	59.0	61.5
- St. Johann	53.3	55.8	60.3	61.5	57.4	61.5
- Alt Kleinbasel	51.2	54.7	60.3	61.5	60.3	57.5
- Clara	55.8	53.6	60.3	61.4	57.1	59.6
- Wettstein	63.8	59.7	60.3	61.5	63.7	59.6
- Hirzbrunnen	67.1	65.4	60.3	61.5	63.8	64.9
- Rosental	54.8	54.2	60.3	61.5	55.7	61.5
- Matthäus	53.7	53.1	60.3	61.4	56.5	59.6
- Klybeck	54.3	56.4	60.3	61.4	56.3	63.2
- Kleinhüningen	59.1	58.6	60.3	61.4	57.1	64.9
Periphery						
- Hofstett.-Flüh	64.9	57.0	59.2	61.4	67.2	53.9
- Bettingen	70.0	86.8	73.3	62.0	68.4	66.3
- Riehen	83.9	80.8	70.2	61.7	67.1	66.2
- Aesch	63.0	64.0	63.5	61.9	62.8	61.5
- Allschwil	69.3	73.9	68.6	61.7	63.6	65.5
- Arlesheim	57.6	74.7	68.5	61.9	64.4	65.9
- Biel-Benken	88.6	86.0	72.4	62.2	70.7	65.2
- Binningen	73.4	70.5	68.6	61.7	63.9	61.9
- Birsfelden	52.4	62.5	63.5	61.6	59.3	63.3
- Bottmingen	76.4	76.3	67.5	61.7	68.0	64.9
- Ettingen	61.5	66.0	62.8	61.5	67.4	59.5
- Münchenstein	59.0	71.6	66.8	61.6	63.0	65.8
- Muttenz	63.3	72.0	65.4	61.5	64.4	66.3
- Oberwil	77.0	81.4	71.6	61.8	67.0	66.2
- Reinach	72.2	73.1	66.1	61.8	65.0	66.3
- Therwil	91.7	73.3	65.4	61.7	66.4	65.8
Standard deviation[b]	10.5	8.2	3.1	0.2	4.7	3.1
Unexplained std. dev.[b,c]		5.5	9.2	10.3	6.8	9.5
Explained variance[b,d]		0.73	0.22	0.02	0.57	0.17
Correlation with observed[b]		0.86	0.52	0.76	0.86	0.43

[a] Uses only slope coefficients only for tax, rent, social interactions or distance, respectively.
[b] Weighted by observed migrating population.
[c] Standard deviation of the difference between true and predicted average incomes.
[d] 1-(unexplained std. dev.)^2/(total std. dev.)^2.

model predicts large income differentials across places, similar to the observed ones both in their magnitude and in their spatial pattern (compare the top two maps in Figure 5). 71% of the observed differences in average incomes across locations are explained by the full model. Taxes alone can account for 23%

of the observed variation; they are, of course, not able to explain any intra-city differences. Social interactions alone can explain 55% of the difference in average incomes; they are especially powerful in explaining intra-city variation. The distance from the city center alone predicts 17% of the observed variation; it captures the circular pattern with richer household in a close ring of municipalities around the city. Housing rents, while significantly interacting with household size, do not account for any substantial income segregation.

8 Conclusions

This study investigates spatial segregation of the population in fiscally decentralized urban areas. The theoretical part proposes the progressivity of local income taxes as a new theoretical explanation for income segregation.

The empirical part studies how income tax differentials across municipalities affect the location decisions of households in the metropolitan area of Basel. The estimation results show that rich households are substantially and significantly more likely to move to low-tax municipalities than poor households. The higher valuation of low taxes by rich households is partly explained by the progressivity of the local income tax. However, rich households prefer low-tax municipalities to a greater extent than is explained by the tax schedule. This impact of taxes holds after controlling for alternative explanations.

Distance form the city center also determines the households' locational choice. Rich households are more likely than poor ones found in places away from the city center up to the first ring of periphery communities, but less likely further out. Large households are also more likely to move out of town than small households. Housing prices as observed offer little explanation for the observed segregation pattern.

Social interactions is established as another major source of income segregation. Households in general tend to choose locations close to other households like themselves: Rich households are significantly more likely to locate in areas with a high fraction of rich residents. Foreign households seek locations with a high share of foreign residents, while rich (Swiss or foreign) households seek locations with a low share of foreigners. Swiss families and rich foreign families are avoiding school districts with a large share of foreign pupils, while foreign families with average income actively seek them.

Appendix: Data

The data used in the empirical investigation were made available by the following institutions:

Household data	Statistical Office of the Canton of Basel-Stadt, merged data from the Cantonal Tax Administration and the Residents Registration Office.
Tax schemes	Swiss Federal Tax Administration, Steuerbelastung in der Schweiz, Natürliche Personen nach Gemeinden 1997, Neuchâtel: Swiss Federal Statistical Office.
Housing prices	Wüest und Partner, Zurich.
Income Distribution	Municipalities: Swiss Federal Tax Administration, Direkte Bundessteuer 1997/98 - Gemeinden. Boroughs: Statistical Office of the Canton of Basel-Stadt.
Foreigners	Municipalities: Swiss Federal Statistical Office, Statistical Office Basel-Land, Department of Finance Solothurn. Boroughs: Statistical Office of the Canton of Basel-Stadt.

Notes on the construction of the variables:

Income (uses household data): The information on the household income is based on the tax assessment. Unmarried adult household members and children with their own income are assessed individually. The income of all individually assessed household members is added up. The income in the raw data is income before tax and deductions for children and spouse but after social security contributions and further deductions. The study uses (hypothetical) gross income which was calculated without considering further individual deductions.

Children (uses household data): Number of children that allow for tax deductions.

Tax rate (uses household and tax schedule data): The tables from the Swiss Federal Tax Administration report the totalled cantonal and municipal average tax rates for different household types (single household, married couple without children and married couple with two children) and for selected gross incomes. The tax rate for households with income between the reported income classes and for household types not listed were interpolated. The tax rates for household members with individual tax assessment were first calculated individually. The tax rate of the household is calculated from the totalled individual tax amounts.

Rent (uses housing price data): Wüest und Partner collected all rents for flats offered in newspapers and in the internet in 1997. Missing information on exact flat sizes was inferred from the information given in the advertisements.

Distance: Distance between the geographical centers of the municipalities. The center was taken as the middle of the maximal east-west and north-south extensions.

References

Alonso, William, 1964. Location and Land Use, Cambridge (MA): Harvard University Press.

Anselin, L. and Raymond J.G.M Florax (Eds.), 1995. New Directions in Spatial Econometrics, Berlin: Springer.

Bayer, Patrick, Robert McMillan, and Kim Rueben (2004), An Equilibrium Model of Sorting in an Urban Housing Market, NBER working paper 10865.

Beckmann, Martin J., 1969. On the Distribution of Urban Rent and Residential Density, Journal of Economic Theory, 1, 60-67.

Bénabou, Roland, 1993. Workings of a City: Location, Education, and Production, Quarterly Journal of Economics, 108, 619-652.

Bénabou, Roland, 1996a. Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance, American Economic Review, 86(3), 584-609.

Bénabou, Roland, 1996b. Equity and Efficiency in Human Capital Investment: The Local Connection, Review of Economic Studies, 63, 237-264.

Ben-Akiva, Moshe and Steven Lerman, 1985. Discrete Choice Analysis: Theory and Application to Travel Demand, Cambridge: MIT Press.

Berman, Abraham and Robert J. Plemmons, 1994. Nonnegative Matrices in the Mathematical Sciences, Philadelphia: Siam.

Berry, Steven, James Levinsohn and Ariel Pakes 1995. Automobile Prices in Market Equilibrium, Econometrica, 63(4), 841-890.

Boadway, Robin W. and Harry M. Kitchen, 1980. Canadian Tax Policy, Toronto, Canada: Canadian Tax Foundation.

Bolduc, Denis, 1992. Generalized Autoregressive Errors: The Multinomial Probit Model, Transportation Research B, 26, 155-70.

Bolduc, Denis, Bernard Fortin and Marc-Andre Fournier, 1996. The Impact of Incentives Policies on the Practice Location of Doctors: A Multinomial Probit Analysis, Journal of Labor Economics, 14, 703-32.

Bolduc, Denis, Bernard Fortin and Stephen Gordon, 1997. Multinomial Probit Estimation of Spatially Interdependent Choices: An Empirical Comparison of Two New Techniques, International Regional Science Review, 20, 77-101.

Börsch-Supan, Axel, 1990. On the Compatibility of Nested Logit Models with Utility Maximisation, Journal of Econometrics, 43, 373-388.

Börsch-Supan, Axel and Vassilis A. Hajivassiliou, 1993. Smooth Unbiased Multivariate Probabilities Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models, Journal of Econometrics, 58, 347-368.

Buchanan, James M. and Richard A. Musgrave, 1999. Public finance and public choice: two contrasting visions of the state, Cambridge: MIT Press.

Durlauf Steven N., 1996a. A Theory of Persistent Income Inequality, Journal of Economic Growth, 1, 75-93.

Durlauf Steven N., 1996b. Neighborhood Feedbacks, Endogeneous Stratification, and Income Inequality. In: W. Barnett, G. Gandolfo and C. Hillinger (Eds.), Dynamic Disequilibrium Modelling, Cambridge: Cambridge University Press.

Ellickson, Bryan, 1971. Jurisdictional Fragmentation and Residential Choice, American Economic Review, 61(2), 334-339

Epple, Dennis and Glenn J. Platt, 1998. Equilibrium and Local Redistribution in an Urban Economy when Households Differ in both Preferences and Income, Journal of Urban Economics, 43, 23-51.

Epple, Dennis and Thomas Romer, 1991. Mobility and Redistribution, Journal of Political Economy, 99(4),828-858.

Epple, Dennis, Thomas Romer and Holger Sieg, 2001. Interjurisdictional Sorting and Majority Rule: An Empirical Analysis, Econometrica, 69(6), 1437-1465.

Epple, Dennis and Holger Sieg, 1999. Estimating Equilibrium Models of Local Jurisdictions, Journal of Political Economy, 107, 645-681.

Feld, Lars P. and Gebhard Kirchgässner, 2001. Income Tax Competition at the State and Local Level in Switzerland, Regional Science and Urban Economics, 31, 181-213.

Friedman, Joseph, 1981. A Conditional Logit Model of the Role of Local Public Services in Residential Choice, Urban Studies, 18, 347-358.

Fujita, Masahisa, 1989. Urban Economic Theory: Land Use and City Size, Cambridge: Cambridge University Press.

Geweke, John, Michael Keane and David Runkle, 1994. Alternative Computational Approaches to Inference in the Multinomial Probit Model, Review of Economics and Statistics, 1994, 609-632.

Hodler Roland and Kurt Schmidheiny, 2005. How Fiscal Decentralization Flattens Progressive Taxes, Department of Economics Discussion Paper 2005-08, Tufts University.

Hajivassiliou, Vassilis A. and Paul A. Ruud, 1994. Classical Estimation Methods for LDV Models using Simulation. In: R. F. Engle and D.L. McFadden (Eds.), Handbook of Econometrics IV, Amsterdam: Elsevier.

Hensher, David A. and William H. Greene, 2002. Specification and Estimation of Nested Logit Models, Transportation Research B, 36(1), 1-18.

Ioannides, Yannis M. and Jeffrey E. Zabel, 2002. Interactions, Neighborhood Selection and Housing Demand, Department of Economics Discussion Paper 2002-08, Tufts University.

Ioannides, Yannis M. and Jeffrey E. Zabel, 2003. Neighborhood Effects and Housing Demand, Journal of Applied Econometrics, 18(5), 563-584.

McFadden, Daniel, 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In: P. Zarembka, Ed., Frontiers in Econometrics, New York: Academic Press.

McFadden, Daniel, 1978. Modelling the Choice of Residential Location. In: A. Karlqvist et al. (Eds.), Spatial Interaction Theory and Planning Models, Amsterdam: North-Holland.

McFadden, Daniel, 1984. Econometric Analysis of Qualitative Response Models. In: Zvi Griliches and Michael D. Intriligator (Eds.), Handbook of Econometrics II, Amsterdam: Elsevier.

McFadden, Daniel, 2001. Economic Choices, American Economic Review, 91, 351-378.

Mills, Edwin S., 1972. Studies in the Structure of the Urban Economy, Baltimore: Johns Hopkins Press.

Muth, Richard, 1969. Cities and Housing, Chicago: Chicago University Press.

Nechyba, Thomas J. and, Robert P. Strauss, 1998. Community Choice and Local Public Services: A Discrete Choice Approach, Regional Science and Urban Economics, 28, 51-73.

Oates, Wallace E., 1969. The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis, Journal of Political Economy, 77, 951-971.

Oates, Wallace E., 1972. Fiscal Federalism, New York: Harcourt Brace Jovanovich.

Rhode, Paul W. and Koleman S. Strumpf, 2003. Assessing the Importance of Tiebout Sorting: Local Heterogeneity form 1850 to 1990, American Economic Review, 93/5, 1648-1677.

Ross, Stephen and John Yinger, 1999. Sorting and Voting: A Review of the Literature on Urban Public Finance. In: Paul Ceshire and Edwin S. Mills (Eds.), Handbook of Regional and Urban Economics, Vol. 3, Amsterdam: North-Holland.

Schelling, Thomas, 1969. Models of Segregation, American Economic Review, 59/2, 488-493.

Schelling, Thomas, 1971. Dynamic Models of Segregation, Journal of Mathematical Sociology, 1, 143-186.

Schmidheiny, Kurt, 2002. Income Stratification in Multi-Community Models, VWI Diskussionsschrift 15-02, Universität Bern.

Schmidheiny, Kurt, in press. Income Segregation from Local Income Taxation When Households Differ in Both Preferences and Incomes. Regional Science and Urban Economics.

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

Train, Kenneth E., 2003. Discrete Choice Methods with Simulation, Cambridge: Cambridge University Press.

von Thünen, Johann H., 1826. Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie, Hamburg : Friedrich Perthes.

Westhoff, Frank, 1977. Existence of Equilibria in Economies with a Local Public Good, Journal of Economic Theory, 14, 84-112.

Wheaton, William C., 1977. Income and Urban Residence: An Analysis of Consumer Demand for Location, American Economic Review, 67(4), 620-631.