HWWA HAMBURG

DISCUSSION PAPER

Income Segregation and Local Progressive Taxation: Empirical Evidence from Switzerland

Kurt Schmidheiny

HWWA DISCUSSION PAPER

248

Hamburgisches Welt-Wirtschafts-Archiv (HWWA)
Hamburg Institute of International Economics
2003
ISSN 1616-4814

Hamburgisches Welt-Wirtschafts-Archiv (HWWA)
Hamburg Institute of International Economics
Neuer Jungfernstieg 21 - 20347 Hamburg, Germany
Telefon: 040/428 34 355
Telefax: 040/428 34 451
e-mail: hwwa@hwwa.de
Internet: http://www.hwwa.de

The HWWA is a member of:

- Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL)
- Arbeitsgemeinschaft deutscher wirtschaftswissenschaftlicher Forschungsinstitute (ARGE)
- Association d'Instituts Européens de Conjoncture Economique (AIECE)

HWWA Discussion Paper

Income Segregation and Local Progressive Taxation: Empirical Evidence from Switzerland

Kurt Schmidheiny*

HWWA Discussion Paper 248
http://www.hwwa.de

Hamburg Institute of International Economics (HWWA)
Neuer Jungfernstieg 21 - 20347 Hamburg, Germany
e-mail: hwwa@hwwa.de

* Department of Economics, University of Bern

The discussions with Klaus Neusser and Mike Gerfin 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 Kenneth Train, Dennis Epple, Marius Brühlhart and the participants of the Crans-Montana Workshop on Globalization and Economic Geography for valuable comments and suggestions. Adrian Herzog´s free MAPresso software was used to produce the informative maps. Heather Murray helped me find the right words.

This discussion paper is assigned to the HWWA's research programme "European Integration and Spatial Development".

Edited by the Department European Integration
Head: Dr. Konrad Lammers

HWWA DISCUSSION PAPER 248
September 2003

Income Segregation and Local Progressive Taxation: Empirical Evidence from Switzerland

ABSTRACT

Swiss metropolitan areas are comprised of a system of communities with considerable fiscal autonomy. This study investigates how the income tax differentials across communities in an urban area affect the households` location decisions. Data from the urban agglomeration of Basel for the year 1997 is used. This unique data set contains tax information from all households that moved either within the city center of Basel or from the city center to the outskirts. The community choice of the households is investigated within the framework of the random utility maximization model (RUM). A theoretical model with progressive income taxation is developed to identify the household preferences applied in the RUM. Different econometric spezifications 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 communities than poor households.

Keywords: Location Choice, Discrete Choice, Multinomial Probit, Mixed Logit, Spatial Autocorrelation, Income Segregation

JEL code: H71, H73, R20, R23

Kurt Schmidheiny
Department of Economics
University of Bern
Gesellschaftsstr. 49
CH-3012 Bern
Switzerland
E-mail: kurt.schmidheiny@vwi.unibe.ch

1 Introduction

Fiscal Federalism is often viewed as the natural counterpart of decentralized decision making. Oates (1972) for example 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. Or as Musgrave (1999, p. 156) pointedly remarks: “To secure an efficient outcome, the provision of public services should be determined and paid for by those who benefit.”

While the virtues of decentralized financial responsibility are uncontested, the resulting tax differentials are highly disputed. Tax differentials can be the consequence of different preferences for the level of 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 as 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 communities with extensive political and fiscal autonomy. Switzerland is also unique in that 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 communities than poor households.

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 communities 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 the households differ in incomes was studied by Ellickson (1971), Westhoff (1977) and 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 the attention to income as the main dimension of difference. Several mechanisms have been proposed that explain why rich households make different choices from poor households (see Ross and Yinger, 1999, for property tax models and Schmidheiny, 2002a, 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 schemes are progressive and that local jurisdictions can often only set the tax level within a given federal tax scheme. This mechanism explains the high priority of tax rates in rich households' decisions through the progressivity the of tax scheme.

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 of a full equilibrium model where the local income distribution and local policy variables are simultaneously determined. They show that the differing income quantiles across 92 communities 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 does not solve the problem, as the general equilibrium of tax

[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.

rates and income segregation is most likely a long-run phenomenon. Rhode and Strumpf (forthcoming) assess the importance of the segregation mechanism in Tiebout type models from a historical perspective. They collected an impressive data set with various 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 have aggravated. They conclude that their data do not support the model predictions on a national scale. For metropolitan areas, 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 communities close to San Francisco. 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 4.1). 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 communities. This variable depends on after-tax local housing prices under the ad-hoc assumption that households consume the same amount of housing in all communities. They therefore implicitly assume that the price elasticity of housing is zero. This assumption is relaxed by using another ad-hoc specification using community-specific coefficients for household income (see footnote 6). Note that the empirical approach depicted in this paragraph neglects the (long-run) reaction of aggregate community characteristics.

Bayer, McMillan and Rueben (2002) attempt a combination of the two empirical approaches. Following Berry, Levinsohn and Pakes (1995) they first estimate the households' choice of a neighborhood, using community fixed effects and a multitude of interaction effects between household and community characteristics. In a second step they explain the community fixed effects by community characteristics using instrumental variables. The estimation strategies in both steps 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 community characteristics.

This study follows the second approach but departs from the previous studies by shifting the focus to assessing the (income) segregation hypothesis. The general locational attractiveness of a community, including local public goods, is considered in community specific fixed effects, thereby avoiding the difficulty of measuring public goods provision. The identification of household effects is drawn on an explicit theoretical multi-community model. Furthermore, recent econometric developments using simulation methods are applied to consider the spatial structure in the error components.

The paper is organized as follows: Section 2 describes the institutional organization of fiscal federalism in Switzerland. A theoretical model of location choice based on progressive income taxation is proposed in Section 3. The econometric model is discussed in Section 4, while Section 5 describes the data. The empirical results and two policy experiments are presented in Section 6. Section 7 draws conclusions.

2 Fiscal Federalism in Swiss Metropolitan Areas

Switzerland is an exemplary federal fiscal system. The Swiss federation comprises 26 states, the so-called cantons. The cantons are divided into roughly 3000 communities 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 93 billion CHF in 2001, of which 46% is imposed by the federation, 32% by the cantons and 22% by the communities.[2] While the federal government is mainly financed by indirect taxes (61% of federal tax revenue) such as the VAT, the cantons and communities largely rely on direct taxes. Income taxes account for 60% of cantonal and 84% of communal tax revenue. In total, 46% of the income tax revenue go to the cantons, 38% to the communities and only 16% to the federal government. Transfers between the three levels are not

[2]All figures in this paragraph apply to 2001. Source: Swiss Federal Tax Administration (2002), Öffentliche Finanzen der Schweiz 2001, Neuchâtel: Swiss Federal Statistical Office.

a major part of the budgets of cantons (23% of total revenue) and communities (14%).

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 communities in turn can generally set a tax shifter for income and corporate taxes. The communal tax is then the cantonal tax rate multiplied by the communal tax shifter. In some cantons, for example in the Canton of Basel-Stadt before 2001, the individual communities also have some freedom in setting the tax scheme. The decisions in the cantons and communities 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 communities within the same canton to some extent, but still leave room for considerable variation.

The above outlined federal system leads to ample differences of income taxes across Swiss communities. For example, for a two-child family with a gross income of 80,000 Swiss francs (CHF) the sum of cantonal and community income tax ranged from 3,6% in the city of Zug to 11,3% 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 tax rates due to the progressive federal and cantonal tax schemes, namely ranging from 10.9% in Wollerau in the Canton of Schwyz to 28.7% in Onex in the Canton of Geneva. The federal income tax for this household was 9.4%.

The tax differences across communities within a single metropolitan area are smaller but still substantial. Figure 1 shows the community characteristics in the metropolitan area of Basel[3] (data sources in the appendix). In 1990 the Basel area was the third largest Swiss metropolitan area with a total of 406,000 inhabitants. The city of Basel with 178,000 inhabitants, hereafter called the center, is the central business district of the area. The top-left map shows the share of workers commuting to the center. The white area to the north and west of the City of Basel is French and German territory and is not included in this study. The Basel area comprises 38 communities from four cantons: Basel-Stadt, Basel-Land, Solothurn and Aargau. There is great variability in both tax levels

[3]Definition of the area according to the Swiss Federal Statistical Office based on Census 1990 data.

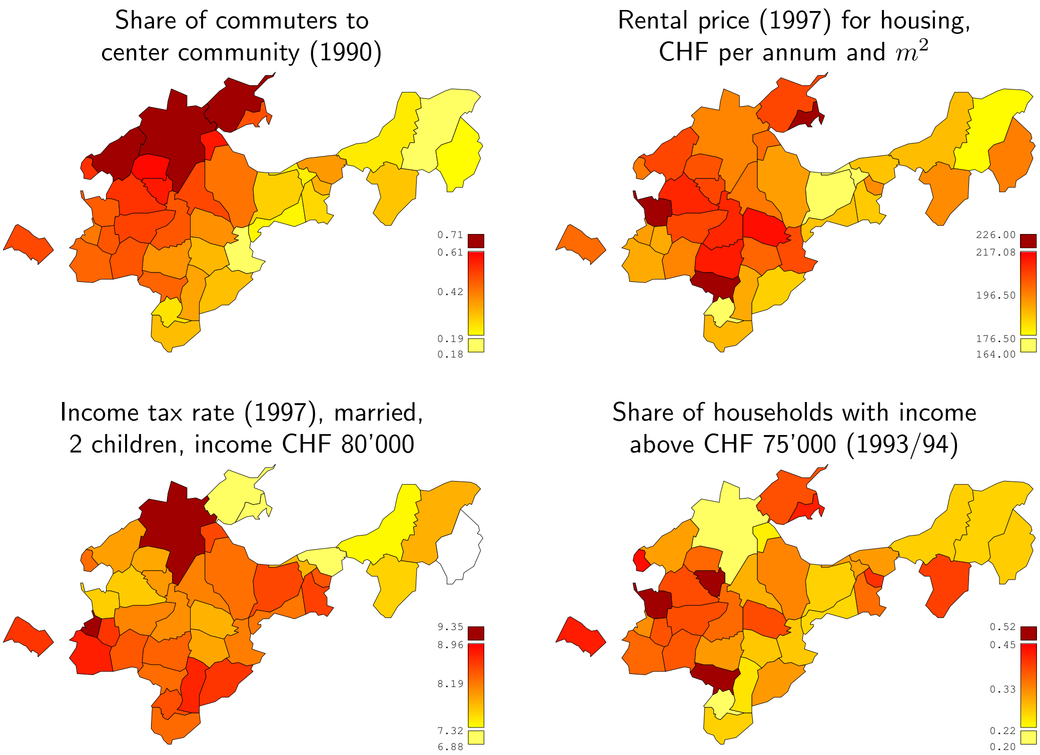

Figure 1: Community characteristics in the metropolitan area of Basel.

and tax schemes. The totalled communal and cantonal income tax rate for a two-child family is depicted in the bottom-left map. The taxes are highest in the center community (9.4 %) and up to 35% lower in the communities around the center.

It is particularly interesting to contrast the local tax rates with the income of the residents. The bottom-right map in figure 1 shows the local share of households with incomes above 75,000 CHF. The map represents to a great extent an inverted picture of the tax rates. The high-tax center community has the lowest proportion of rich households, whereas the low-tax communities close to the center are populated by many more rich households. The rental prices for housing also seem to be correlated with the tax rates. The low-tax fringe of communities around the center exhibits higher average prices than the center, although communities further away from the center are clearly the lowest-price locations.

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.

The local public expenditures are financed by a tax on the residents' income. The income tax rate $t_j(y)$ in a community j depends on the households' income y. The provision of the public good g is fixed by the state government and, hence, is identical across communities. This assumption is motivated by the observation that the autonomy of local decision making is in fact often strongly limited by state and federal laws.[4] In the case of education expenditures, for example, teachers' salaries as well as class sizes are regulated by the cantons.

The price for housing p_j in community j is determined 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 maximizing its utility as place of residence.

This paper focuses on the households' location decision and does not develop a full general equilibrium model. A complete model includes the description of the housing supply function, the production function of the public good and the equilibrium concept.

[4]The exogenous determination of public goods provision substantially simplifies the model and turns the focus to income segregation induced by local taxation. A comprehensive model of local provision of local public goods would have to endogenize the provision of local public goods. However, this greatly complicates the analysis without providing qualitatively different results for location choice and income segregation. In addition, the more general approach makes it intractable to consider progressive tax schemes, which are crucial in the empirical investigation. See Schmidheiny (2002a) for a discussion of the technical problems and Schmidheiny (2002b) for the properties of a model with endogenous local public goods provision financed by local income taxes.

3.1 Household Preferences

The preferences of the households are described by a utility function [5]

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

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. The price of the private good is set to unity. The disposable income y_d depends on the local income tax shifter $t > 0$ and the exogenous tax rate structure $r(y)$. The tax scheme $r(y) > 0$ is assumed to be continuous in y. The (average) tax rate $t_j(y) = t \cdot r(y)$ is smaller than the marginal tax rate $\partial(t\, r(y)\, y)/\partial y = t[r + y\, r'(y)]$ and both are assumed to lie in $[0, 1)$.

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

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

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

[5]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. This assumption is relaxed in Section 3.4.

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 1 states that the income tax scheme is proportional or progressive.

Property 2 (Relative preferences)
If Assumptions 1 and 2 hold and if and only if 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. 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 scheme, the income elasticity of housing below unity or a combination of both. As will become apparent in Section 4, 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.

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.

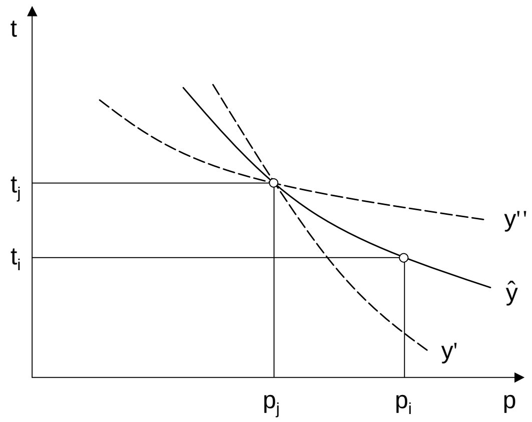

Figure 2: Indifference curves in the (t, p) space.

(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 $f(y, \alpha) > 0$.

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:

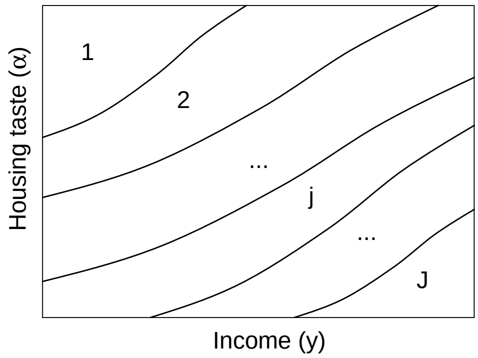

Figure 3: Simultaneous income and preference segregation. The areas denoted by $j = 1, ..., J$ show the attributes of the households that prefer community j.

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.

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 Adding Intrinsic Community Attractiveness

In reality, communities differ in much more than tax levels and housing prices. Factors such as vicinity of the central business district, cultural activities, shopping opportunities, climate and landscape can have an influence on a household's choice of residential location. In addition, the provision of local public goods such as schools, though exogenously set here, can differ across communities.

A straightforward way to incorporate these factors is to add a community specific constant k_j to the (indirect) utility function:

$$V_j = V(t_j, p_j, k_j\,; y, \alpha) = U(h_j^*, b_j^*) + k_j\,. \tag{4}$$

This extension does not fundamentally change the logic of the model. Propositions 2, 3 and 4 still hold because the additive separable specification preserves properties 1 and 2. Proposition 1, however, does not hold any more as one can imagine communities so attractive that they can attract households even when they show both higher taxes and higher housing prices. This is usually observed for the center communities in Swiss metropolitan areas.

3.5 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 scheme. The derived indirect utility function will serve as a benchmark in the empirical study.

Household preferences are described by a Cobb-Douglas utility function

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

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

$$V_j = V(t_j, p_j\,; y, \alpha) = k - \alpha\, log(p_j) + log(y) + log[1 - t_j\, r(y)]\,, \tag{5}$$

where $k = \alpha\, log(\alpha) + (1-\alpha)\, log(1-\alpha) + \gamma\, log(g)$. The intrinsic attractiveness of the place as outlined in Section 3.4 is considered by using a community specific constant k_j.

4 The Econometric Model

The empirical part of this study aims to establish factors that determine a household's choice of the residential community in a metropolitan area. The location choice in the theoretical model of the previous section leads naturally to a multinomial response model based on random utility maximization (RUM). Multinomial response models are closely connected with McFadden's (1974, 2001) seminal work on 'economic choices'. The subsequent presentation draws upon Train (2003).

The choice of one of many unordered alternatives is driven by a latent variable, often interpreted as indirect utility. The indirect utility V_{nj} of a household n in a community j is the sum of a systematic and a stochastic part

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

where n indicates the household and j the community. V_{nj} is a deterministic function of observable household and community characteristics and ε_{nj} is a household and community specific error term.

A household n chooses community j among the choice set $C = (1, ..., J)$ as its place of residence if it offers the highest value of indirect utility, i.e.

$$V_{nj}^{*} \geq V_{ni}^{*} \quad \text{for all } i \in C \ . \tag{6}$$

This is equivalent to equation (2) in the theoretical model.

4.1 Functional Form and Identification

The indirect utility function (equation 4) in the theoretical part guides the choice of systematic factors in the indirect utility function

$$V_{nj} = V(t_{nj}, p_j, k_j, y_n, a_n) , \tag{7}$$

where t_{nj} is the income tax rate of household n in community j, p_j is the housing price in community j, k_j indicates further community specific dimensions of attractiveness, y_n is household income and a_n indicate further household characteristics.

From the point of view of an individual household, the community characteristics are exogenous, although they are the (long-term) aggregate of the agents' decisions. Therefore the household's location decision is optimal given the community characteristics. In the theoretical model this is established by the assumption that there is a continuum of households, i.e. that a single household is 'small' and does not influence the equilibrium.

For the empirical implementation the functional form of the deterministic part (equation 7) of the latent variable needs to be specified. Starting point is the indirect utility function (5) from the benchmark case presented in Section 3.5,

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

where k_j and β_0 can be vectors. Note that the parameters are identical across the alternatives.

The theoretical model offers two mechanisms which explain why rich households move to different communities than poor households: progressive taxation and income elasticity of housing below unity. Two interaction terms are added to allow for the latter segregating mechanism:

$$\begin{aligned} V_{nj} = {} & \beta_0\, k_j + \beta_1\, log(1 - t_{nj}) + \beta_2\, log(p_j) + \beta_3\, log(y_n) \\ & + \beta_4\, log(1 - t_{nj})\, log(y_n) + \beta_5\, log(p_j)\, log(y_n)\,. \end{aligned} \tag{8}$$

The implied MRS between tax rate and housing price satisfies Property 2 (relative preferences) even in the case of proportional taxes

$$\left.\frac{\partial M}{\partial y}\right|_{t_{nj}=t_j} = \frac{(1 - t_j)(\beta_1\, \beta_5 - \beta_2\, \beta_4)}{p_j[\beta_1 + \beta_4\, \log(y_n)]^2\, y_n} > 0 \tag{9}$$

if $\beta_1\, \beta_5 > \beta_2\, \beta_4$. As one expects $\beta_1 > 0$ and $\beta_2 < 0$ for any household this is guaranteed by $\beta_4 > 0$ and $\beta_5 > 0$. This means that the effect of the tax rate increases with income while the effect of the housing price decreases. As is shown in the proof of Property 2, the progressive tax scheme reinforces this segregation mechanism.

The idea of heterogenous tastes for housing in Section 3.3 is applied by substituting the constant housing price effect β_2 with a household dependent effect $\beta_6 + \beta_7\, a_n$:

$$\begin{aligned} V_{nj} = {} & \beta_0\, k_j + \beta_1\, log(1 - t_{nj}) + \beta_6\, log(p_j) + \beta_3\, log(y_n) \\ & + \beta_4\, log(1 - t_{nj})\, log(y_n) + \beta_5\, log(p_j)\, log(y_n) + \beta_7\, log(p_j)\, a_n\,, \end{aligned}$$

where a_n and β_7 can be vectors.

The *level* of the indirect utility function is not identified as the agents only care about the differences between alternatives. Consequently, factors that shift the indirect utility of all alternatives in the same way are not identified; hence β_3 cannot be estimated. This leads to the identified indirect utility function:

$$\begin{aligned} V_{nj} = {} & \beta_0\, k_j + \beta_1\, log(1 - t_{nj}) + \beta_6\, log(p_j) \\ & + \beta_4\, log(1 - t_{nj}) log(y_n) + \beta_5\, log(p_j)\, log(y_n) + \beta_7\, log(p_j)\, a_n\,. \end{aligned} \tag{10}$$

Note that the *scale* of the indirect utility function will be arbitrarily set by the specification of the error term.

The community characteristics k_j may be imprecisely measured or not observable. It is therefore advantageous to include community fixed effects which capture all unobserved dimensions of intrinsic community attractiveness. However, the effect of an observed community characteristic cannot be distinguished from the fixed effect of this community and is thus not identified. The identified fixed effects specification is:

$$\begin{aligned} V_{nj} = {} & \delta_j + \beta_1\, log(1 - t_{nj}) + \beta_4\, log(1 - t_{nj}) log(y_n) \\ & + \beta_5\, log(p_j)\, log(y_n) + \beta_7\, log(p_j)\, a_n\,, \end{aligned} \tag{11}$$

where the community-specific constant δ_j is identified by setting the constant of an arbitrary community to zero. Note that the effect of the tax rate t_{nj} can still be estimated because it depends on both the community j and the household n.

4.2 Modelling the Stochastic Part

So far, the stochastic element ε_{nj} of household n's utility in community i has not been discussed. The stochastic part stands for 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 place. There is very little theoretical guidance that would help to model the stochastic term. Several specifications are therefore 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.[6] The probability that household n chooses community j is

$$P_{nj}(\theta) = \frac{e^{V_{nj}}}{\sum_{i=1}^{J} e^{V_{ni}}} , \tag{12}$$

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 across the alternatives is a strong assumption. It implies that a household's stochastic, i.e. unobserved, preference for a certain community is fully independent of its stochastic preference for other communities. 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 community specific error terms $\varepsilon_n = (\varepsilon_{n1}, ..., \varepsilon_{nJ})$ follows the generalized extreme value distribution (GEV) introduced by McFadden (1978):

$$F(\varepsilon_n) = e^{[-\sum_{k=1}^{K}\left(\sum_{i\in C_k} e^{-\varepsilon_{ni}/\lambda_k}\right)^{\lambda_k}]} .$$

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. $1 - \lambda_k$ can be used as an indication of correlation, but the link is more complicated (see McFadden, 1978). 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 community j is

$$P_{nj}(\theta) = \frac{e^{V_{nj}/\lambda_l}\left(\sum_{i\in C_l} e^{V_{ni}/\lambda_l}\right)^{\lambda_l - 1}}{\sum_{k=1}^{K}\left(\sum_{i\in C_k} e^{V_{ni}/\lambda_k}\right)^{\lambda_k}} , \tag{13}$$

[6] The conditional logit model is also called *multinomial logit* model. Modern treatises on multinomial response use the notion of multinomial logit for a specification in which the slope parameter β_j depends on the alternative j. These alternative specific parameters are difficult if not impossible to derive from economic choice behavior.

where l is the nest of community j and $\theta = (\beta, \lambda)$.[7] The nested structure of the error term can be looked at as the result of a two-stop 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 stay in the center community or to move to a community in the periphery. Households with a large unobserved preference for a community in the periphery therefore also have a higher preference for all other communities in the periphery. In this case the center community is a nest on its own, called a degenerate nest with $\lambda_k = 1$. This nested structure can be considered as a simple form of a spatial correlation pattern.

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. 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 4.3) 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. This study uses a very parsimonious specification of Ω. Following Bolduc (1992) and Bolduc, Fortin and Gordon (1997),[8] the alternative specific error terms follow a first order spatial

[7] 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, and textbooks, e.g. Greene (2003), 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. Stata offers a revised command `nlogitrum` (see Heiss, 2002) which correctly implements a nested logit model. This command is, however, not able to deal with degenerate nests and a full set of alternative fixed effects.

[8] 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.

autoregressive process (SAR)[9]

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

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

$$P_{nj}(\theta) = Prob\,[\varepsilon_1 - \varepsilon_j > V_{nj} - V_{n1}, \ldots, \varepsilon_J - \varepsilon_j > V_{nj} - V_{nJ}]\,, \qquad (14)$$

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.

4.3 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

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

where $z_{nj} = 1$ if the household n chooses community j and $z_{nj} = 0$ otherwise. The choice probabilities P_{nj} of the conditional logit and nested logit model are defined in equations (12) and (13), respectively. The maximum likelihood estimator $\hat{\theta} = (\hat{\beta}, \hat{\lambda})$ is consistent, asymptotically efficient and normally distributed.

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

The multinomial probit model is estimated with maximum simulated likelihood (MSL, see Hajivassiliou and Ruud, 1994). The 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 (14) by recursively drawing from univariate normal distributions. The number of draws R determines the quality of the approximation. 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).

All estimations are performed with the author's own programs in MATLAB.[10] 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 numerically demanding estimation of the multinomial probit model runs around 70 hours on a Sun Fire V880.

5 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 community in the periphery in the year 1997 .

The decision maker in the theoretical model is a household. Households are operationalized as all persons that 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.[11]

The choice set of these households consists of roughly 3000 Swiss communities

[10] A MATLAB toolbox with programs for conditional logit, nested logit, multinomial probit and mixed logit models is available from the author on request.

[11] Married couples that move from single households into a common flat are also treated as one household. Unmarried couples that start living together at the new address are treated as independent households.

and the communities in neighboring France and Germany and in principal the whole rest of the world. However, from both a theoretical and a practical point of view this potential choice set is not the relevant one in the analysis. Tiebout type models of location choice are only suitable for narrow metropolitan areas.[12] Moreover, the econometric methods used are numerically unfeasible for large choice sets. The analysis is therefore restricted to the city of Basel and a circle of the 16 most integrated communities around it.[13] This leaves 7,872 households with 11,540 members in the data set. The communities belong to three different cantons, Basel-Stadt (BS), Basel-Land(BL) and Solothurn (SO) and thus exhibit great variability in tax levels and tax schemes.

Some information on the communities in the choice set is also used in the analysis. As this study uses community fixed effects, little effort was devoted to finding variables describing community attractiveness. The following enumeration describes the variables used. See the appendix for a detailed description of the variables and the data sources.

- *Income* (household specific): Total gross income of all household members according to the last tax assessment before moving.[14]
- *Marital status* (household specific): Marital status of the primary earner.
- *Children* (household specific): Number of under-age children.
- *Tax rate* (household and community specific): Tax rate for totalled cantonal (state) and communal income taxes. It reflects community/state specific tax deductions, community/state specific progressive tax schemes and community specific tax shifters and thus depends on household income as well as on marital status and children. The hypothetical tax rate is computed for any household as well as any of the 17 communities in the choice set.

[12]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 community is a feasible place of residence for households whose members are working in the central business district.

[13]These communities are defined as all communities where more than 36% of the working population is commuting to the center community (Census 1990). This admittedly arbitrary cutting off point leads to a well-shaped geographic area and a tractable number of choice opportunities. The five smallest communities are omitted as they are not covered in the tax scheme data. Changing the choice set did not qualitatively change the results of the analysis.

[14]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.

Table 1: Characteristics of movers from the center community in 1997.

	households moved in	mean income	median income	chil-dren	dis-tance	rent	tax mid income†	tax high income†
Whole Area	7872	61,612	54,449	0.32		206	8.05	22.52
City of Basel (BS)	6370	59,334	52,328	0.32	0	197	9.36	26.41
Periphery	1502	71,271	61,874	0.31		207	7.97	22.30
- Binningen (BL)	165	73,405	60,106	0.19	2.5	205	7.88	21.80
- Birsfelden (BL)	98	52,351	52,033	0.23	3.2	200	8.51	23.19
- Bottmingen (BL)	43	76,376	74,131	0.37	4.1	206	7.98	22.08
- Allschwil (BL)	251	69,302	63,138	0.30	4.6	207	7.94	21.77
- Münchenstein (BL)	92	58,962	54,567	0.29	4.9	198	8.13	22.26
- Oberwil (BL)	80	77,048	64,702	0.21	5.4	211	7.66	21.05
- Riehen (BS)	280	83,950	72,428	0.39	5.6	206	6.88	21.77
- Muttenz (BL)	114	63,333	56,688	0.35	5.7	192	8.24	22.66
- Bettingen (BS)	9	69,978	67,177	0.11	6.2	220	7.20	20.86
- Reinach (BL)	151	72,242	61,992	0.25	6.5	212	8.04	22.53
- Arlesheim (BL)	56	57,601	56,688	0.21	7.2	215	7.81	21.88
- Therwil (BL)	46	91,735	79,672	0.54	7.3	207	8.11	22.73
- Biel-Benken (BL)	18	88,610	72,350	0.28	7.8	226	7.64	20.87
- Aesch (BL)	57	62,968	53,506	0.35	9.5	213	8.33	23.33
- Ettingen (BL)	24	61,541	65,999	0.38	10.1	197	8.40	23.54
- Hofst.-Flueh (SO)	18	64,902	55,863	0.61	11.5	190	8.77	24.43

† Cantonal and communal income tax rate for married couple with two children and income of CHF 80,000 and CHF 500,000 respectively.

- *Rent* (community specific): Average offer price per m^2 for a rented flat.
- *Distance* (community specific): Distance in km^2 between a community and the central business district.

Table 1 gives descriptive statistics of the household and community characteristics. From the total of 7,872 households that stayed within the choice set, 4/5 moved within the center community whereas only 1/5 moved to one of the 16 communities in the periphery. The latter were on average 20% richer than the ones remaining in the center. The tax rate of a typical two-child family with an income of CHF 80,000 is highest in the center community; this is more than 35% higher than in the neighboring community Riehen. The tax rate for an income of CHF 500,000 is about three times higher and the tax amount consequently 15 times as high, reflecting the strong progressivity of the different tax schemes. Figure 4 visualizes the association between the local tax level and the average income of households that moved in. The number of children of households in the center and in the periphery are very similar. However, there is substantial varia-

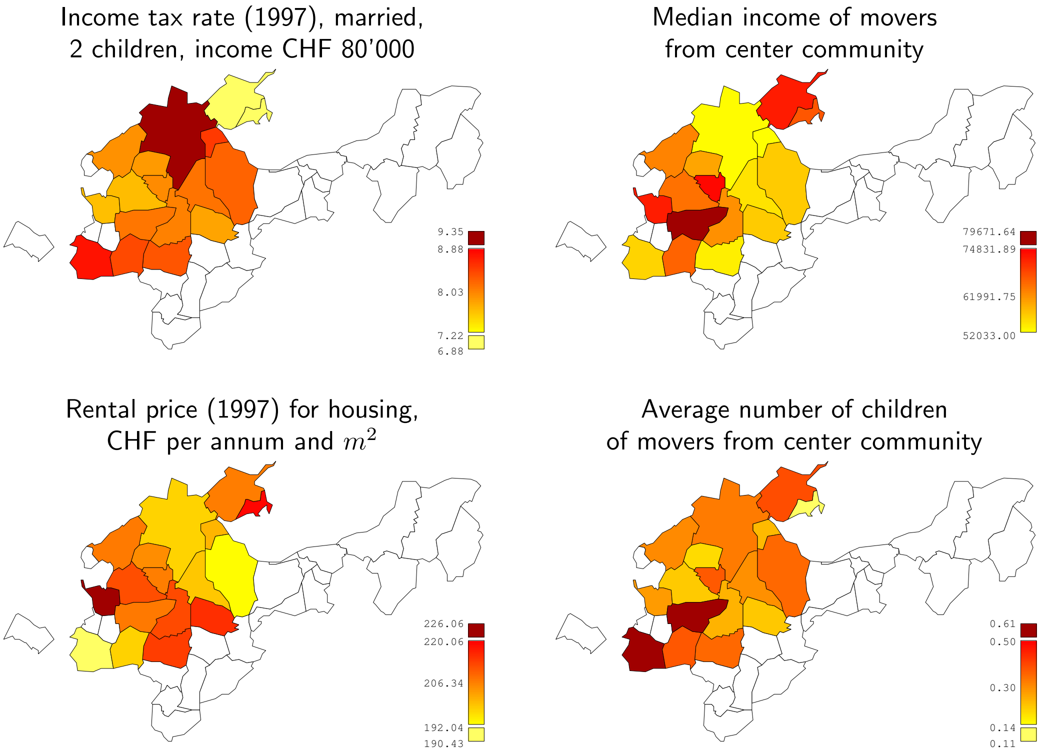

Figure 4: Characteristics of movers from the center community in 1997.

tion across the communities in the periphery. The bottom maps in Figure 4 show the local housing rent and the average number of children of the incomers. The center community is surrounded by a fringe of communities with higher rents. Families with more children tend to locate in communities with lower rents.

6 Results

The estimates of the random utility models with fixed effects are given in Table 2. Three specifications of the error term structure are reported: Column one shows the results for the conditional logit model, column 2 for the nested logit model and column three for the multinomial probit model with a spatial autoregressive process (SAR). All models are estimated with a full set of community-specific constants using the city of Basel as reference community.

The specification of the error term is discussed first. The nested logit model fits the data significantly better than the conditional logit model (likelihood ratio test statistic 5.4). The highly significant log-sum coefficient in the nested logit model clearly demonstrates the violation of the IIA assumption in the conditional logit model. The estimate $\hat{\lambda} = 0.53$ implies that the error terms across communities in the periphery are positively correlated. This means that households with an unobserved taste for a community in the periphery also prefer other communities in the periphery. The estimated spatial autocorrelation coefficient $\hat{\rho}$ in the multinomial probit model is not significantly different from zero. The spatial autocorrelation process therefore does not improve the probit model with independent error terms (log Likelihood = -7442.68), which is the analogue of the conditional logit model and also saddled with the IIA. The nested logit model is thus the preferred model. The following discussion relates to the results of the nested multinomial logit model.

While the sign and significance of the coefficients in multinomial response models are informative they cannot be directly interpreted as marginal effects. The significantly negative fixed effects of all communities in the periphery indicate that these communities are intrinsically less attractive to movers from the city of Basel. These estimates are not surprising as 4/5 of the movers decided to stay in the center. The fixed effects take account of locational factors such as housing prices, public goods provision, distance to the central business district, cultural activities and landscape but also a possible distaste for leaving the accustomed community.

The coefficient for $log(1 - tax)$ gives the effect of the tax rate on the indirect utility function for a household with an income of CHF 60,000, i.e. the average income. It is significantly positive on the 0.1% level and confirms that taxes have a negative effect on utility. The significantly positive coefficient of the interaction with $[log(inc) - log(60,000)]$ implies that the effect from taxes increases with income. For example, the effect from $log(1-tax)$ for a household with an income of CHF 500,000 is $14.55 + 11.04 \cdot [log(500,000) - log(60,000)] = 37.96$. The quantitative impact of the tax rate will be explained using an example. Consider a household with an income of CHF 500,000 that compares the city of Basel to the neighboring community of Riehen. The tax rate it faces is 26.4% in Basel and 21.8% in Riehen. The utility difference from this tax differential is $[log(1 - 0.218) - log(1 - 0.264)] \cdot 37.96 = 2.30$. Hence the negative fixed effect of Riehen (-2.72) is almost offset by its lower taxes. However, for a household with an income of CHF 80,000 the implied utility difference is only 0.5.

The coefficient of the local housing prices $log(rent)$ is not identified as its

Table 2: Multinomial response models with fixed effects.

	Conditional Logit			Nested Logit			Mult. Probit, SAR		
Slope Coefficients									
log(1-tax)†	15.40	(2.54)	***	14.55	(2.00)	***	8.06	(1.70)	***
log(1-tax)×[log(inc)-log(60k)]	11.28	(2.00)	***	11.04	(1.75)	***	7.59	(1.31)	***
log(rent)×[log(inc)-log(60k)]	0.70	(0.81)		0.56	(0.47)		0.26	(0.35)	
log(rent)× children	-1.32	(0.82)		-1.16	(0.56)	*	-0.80	(0.41)	~
Correlation Structure Coefficients									
Log-sum periphery (λ)				0.53	(0.08)	***			
Spatial autocorrelation (ρ)							0.46	(1.08)	
Fixed Effects									
Binningen (BL)	-4.14	(0.12)	***	-3.08	(0.20)	***	-2.69	(0.27)	***
Birsfelden (BL)	-4.54	(0.12)	***	-3.24	(0.23)	***	-2.86	(0.26)	***
Bottmingen (BL)	-5.45	(0.18)	***	-3.77	(0.31)	***	-3.31	(0.19)	***
Allschwil (BL)	-3.70	(0.11)	***	-2.84	(0.17)	***	-2.53	(0.17)	***
Münchenstein (BL)	-4.68	(0.13)	***	-3.35	(0.24)	***	-2.99	(0.14)	***
Oberwil (BL)	-4.90	(0.15)	***	-3.51	(0.26)	***	-3.12	(0.13)	***
Riehen (BS)	-3.52	(0.10)	***	-2.72	(0.15)	***	-2.47	(0.05)	***
Muttenz (BL)	-4.46	(0.12)	***	-3.23	(0.22)	***	-2.89	(0.14)	***
Bettingen (BS)	-6.93	(0.35)	***	-4.53	(0.45)	***	-3.87	(0.16)	***
Reinach (BL)	-4.17	(0.12)	***	-3.07	(0.20)	***	-2.79	(0.07)	***
Arlesheim (BL)	-5.21	(0.17)	***	-3.65	(0.29)	***	-3.25	(0.14)	***
Therwil (BL)	-5.35	(0.17)	***	-3.70	(0.30)	***	-3.31	(0.08)	***
Biel-Benken (BL)	-6.38	(0.26)	***	-4.29	(0.39)	***	-3.74	(0.14)	***
Aesch (BL)	-5.08	(0.16)	***	-3.53	(0.28)	***	-3.19	(0.08)	***
Ettingen (BL)	-5.97	(0.22)	***	-4.01	(0.35)	***	-3.55	(0.09)	***
Hofst.-Flueh (SO)	-6.06	(0.24)	***	-4.00	(0.37)	***	-3.53	(0.09)	***
Log likelihood	-7439.2			-7436.5			-7442.59		
Observations	7872			7872			7872		

Standard errors in brackets. ***, **, *, ~ Significance at the 0.1%, 1%, 5% and 10% level.
†The coefficient gives the effect for a household with an income of CHF 60,000 as the interaction term $log(1 - tax) \cdot [log(inc) - log(60,000)]$ becomes zero.

part is taken by the community constants. The interaction of the housing price with income can still be estimated but turns out to be insignificant. The housing rent also interacts with the number of children in the household. This effect is significantly negative at the 5% level. As one can sensibly suppose a negative effect of housing prices on utility, the negative sign of the interaction means that households with children are more concerned about housing prices than childless households.

How do these results correspond to the segregation hypotheses postulated in the theoretical model? The effect of the tax rate without interaction is identified only through the variation across households induced by the progressivity of the tax scheme. The significant positive sign of this coefficient establishes the income segregation from progressive taxation. Equation (9) depicts the conditions for segregation that are induced by mechanisms beyond the progressive tax scheme. The signs of the estimated coefficients satisfy this condition under the assumption that the unobserved price effect is negative. There is clear evidence that rich households prefer low-tax communities but to an even greater extent than is explained by the tax scheme.

The quantitative implications of the estimated nested logit model are revealed by inspecting its predictions. Given the attributes of the households and the communities, the model is able to predict the fraction of the households that move to a particular community j:

$$\frac{1}{N}\sum_{n=1}^{N} P_{nj}(\hat{\theta})\,. \tag{15}$$

Table 4 (column 3) shows the predicted migration to the communities. The actual values in the data set are given for comparison in column 1. Note the perfect forecast of migration which is an artefact of the full set of community intercepts. However, these predictions will change in the policy experiments conducted in Section 6.1. The predicted segregation pattern is more informative. The average income of the households moving to community j is predicted as

$$\sum_{n=1}^{N} y_n\, P_{nj}(\hat{\theta}) \Big/ \sum_{n=1}^{N} P_{nj}(\hat{\theta})\,. \tag{16}$$

Table 4 shows the predicted average income (column 4) of the migrants compared to the actual values (column 2). The top maps in Figure 5 visualize the actual income segregation and the segregation predicted by the nested logit model. As can be seen, the predicted pattern of income differences across communities is very similar to the observed pattern. This remarkably demonstrates the appropriateness of the econometric specification in equation (11).

The results of multinomial response models without fixed effects are given in Table 3. Remember that identifying locational factors is not the prime interest of this paper and not much effort was spent on finding proxies for community characteristics. The only additional variables are *distance* from the central business

Table 3: Multinomial response models with community characteristics.

	Conditional Logit			Nested Logit			Mult. Probit, SAR		
Slope Coefficients									
log(1-tax)†	15.89	(2.26)	***	14.10	(1.98)	***	10.11	(1.25)	***
log(1-tax)×[log(inc)-log(60k)]	12.65	(1.91)	***	12.47	(1.68)	***	8.18	(1.15)	***
log(rent)†	-2.88	(0.77)	***	-1.84	(0.48)	***	-1.81	(0.35)	***
log(rent)×[log(inc)-log(60k)]	0.15	(0.67)		0.05	(0.37)		-0.02	(0.30)	
log(rent)×children	-0.96	(0.73)		-0.83	(0.46)	~	-0.60	(0.37)	
distance	-0.18	(0.01)	***	-0.09	(0.02)	***	-0.10	(0.01)	***
periphery	-3.47	(0.10)	***	-2.67	(0.20)	***	-2.29	(0.05)	***
Correlation Structure Coefficients									
Log-sum periphery (λ)				0.51	(0.10)	***			
Spatial autocorrelation (ρ)							0.76	(0.04)	***
Standard deviation rent ($\sigma_n u$)									
Log likelihood	-7778.1			-7774.5			-7759.6		
Observations	7872			7872			7872		

Standard errors in brackets. ***, **, *, ~ Significance at the 0.1%, 1%, 5% and 10% level.
†The coefficient gives the effect for a household with an income of CHF 60,000 as the interaction term $log(1 - tax) \cdot [log(inc) - log(60,000)]$ becomes zero.

district and the now identified local housing prices $log(rent)$. A dummy variable for the periphery was also introduced to capture the high proportion of stayers. This dummy variable could as well have been labelled 'staying'.

The slope coefficients of the nested logit model are almost identical to the ones in the fixed effects model. As in the fixed effects model, the significant log-sum coefficient shows the violation of the IIA. The now identified housing price effect is significantly negative as was expected above. Not surprisingly, distance from the center has a significant negative impact.

The coefficient of spatial autocorrelation $\hat{\rho} = 0.73$ is now significantly positive. The predictions from the resulting multinomial probit model are very similar to the nested logit model. Note that the coefficients in the multinomial probit model are smaller than in the logit models because the model is scaled by the variance of the standard normal distribution $\sigma_\varepsilon = 1$ rather than that of the extreme value distribution $\sigma_\varepsilon = 1.7$.

6.1 Policy Experiments

An important feature of the estimated models is that they can be used to simulate the aggregate effects from changes in policy variables. Given the attributes of the

Table 4: Model predictions and results of policy experiments.

	Sample Values		Model		Experiment 1		Experiment 2	
	share mover	mean income	share mover	mean income	share mover	mean income	share mover	mean income
Whole area	1	61,612	1	61,612	1	61,612	1	61,612
City of Basel (BS)	0.809	59,334	0.809	59,231	0.772	56,708	0.817	59,619
Periphery	0.191	71,271	0.191	71,710	0.228	78,233	0.183	70,522
- Binningen (BL)	0.021	73,405	0.021	73,362	0.025	80,052	0.022	75,375
- Birsfelden (BL)	0.012	52,351	0.012	63,702	0.015	68,800	0.013	65,316
- Bottmingen (BL)	0.005	76,376	0.005	71,473	0.007	77,729	0.006	73,370
- Allschwil (BL)	0.032	69,302	0.032	73,655	0.038	80,444	0.034	75,695
- Münchenstein (BL)	0.012	58,962	0.012	69,193	0.014	75,276	0.012	71,056
- Oberwil (BL)	0.010	77,048	0.010	80,627	0.012	88,895	0.011	83,093
- Riehen (BS)	0.036	83,950	0.036	75,489	0.042	82,913	0.018	52,069
- Muttenz (BL)	0.014	63,333	0.014	66,291	0.017	71,905	0.015	68,047
- Bettingen (BS)	0.001	69,978	0.001	86,834	0.001	98,430	0.001	53,220
- Reinach (BL)	0.019	72,242	0.019	69,510	0.023	75,100	0.020	71,221
- Arlesheim (BL)	0.007	57,601	0.007	74,243	0.009	80,690	0.008	76,181
- Therwil (BL)	0.006	91,735	0.006	67,810	0.007	73,198	0.006	69,477
- Biel-Benken (BL)	0.002	88,610	0.002	84,549	0.003	93,477	0.002	87,255
- Aesch (BL)	0.007	62,968	0.007	65,115	0.009	69,965	0.008	66,635
- Ettingen (BL)	0.003	61,541	0.003	62,628	0.004	67,365	0.003	64,156
- Hofst.-Flueh (SO)	0.002	64,902	0.002	55,773	0.003	60,061	0.002	57,331

Notes: Predictions and results from experiments (see text) using the estimated nested logit model with fixed effects.

households and the communities after the implementation of the experiment, the models can predict the number of migrants and their average income according to equations (15) and (16). Two experiments are performed.

Experiment 1 (Tax increase in the center)

The center community increases its income tax rate by a factor of 1.1 for all household types.

Table 4 shows the predicted effects of the two experiments based on the estimated nested logit model with fixed effects. Experiment 1 means that the tax rate of a two-child family with an income of CHF 80,000 rises by almost one percentage point from 9.4% to 10.3% in the center community. The tax increase is 2.6% for an income of CHF 500,000. As a result, fewer households would choose the center community. This can be seen in column 5 of Table 4. The center

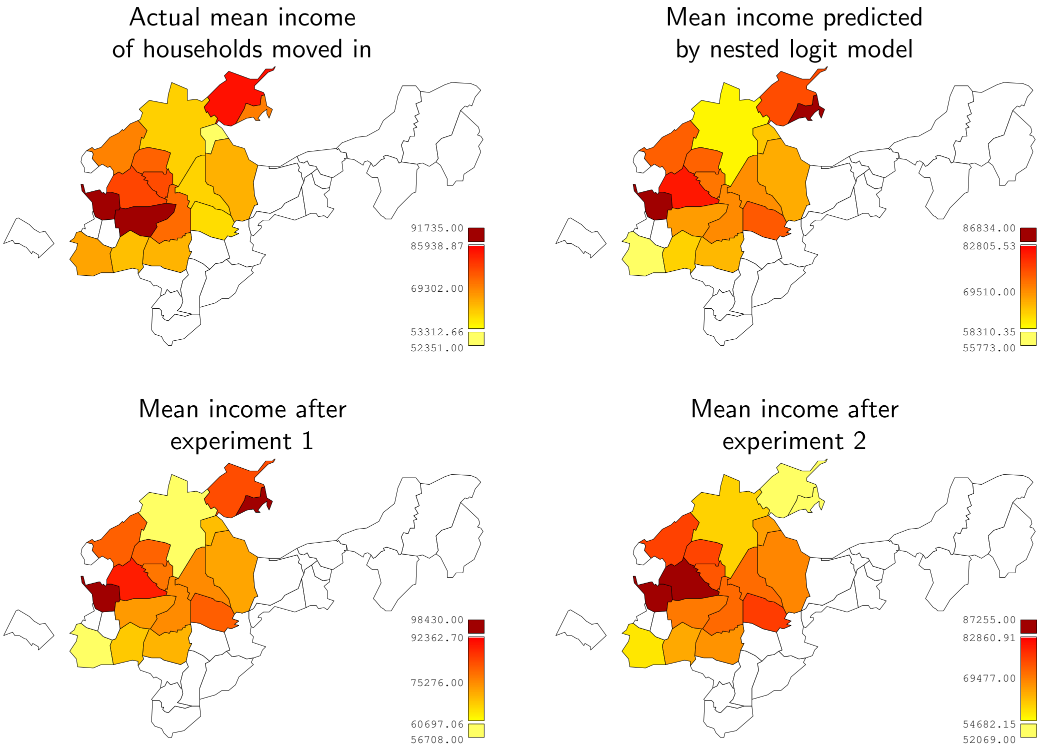

Figure 5: Model predictions and results of policy experiments.

community would lose 3.7% of the moving population. The change of the spatial distributions of incomes provided in column 6 is particularly interesting. The average income would fall by CHF 2,500 in the center community and increase by CHF 6,500 in the periphery, ranging from CHF 4,300 in Hofstetten-Flueh to 11,600 in Bettingen. These effects are depicted in the bottom-left map in Figure 5.

Experiment 2 (Tax harmonization in one canton)
The Canton of Basel-Stadt decides that its two communities in the periphery (Riehen and Bettingen) will employ the same tax scheme and rate as its capital, the city of Basel.

Experiment 2 simulates the choice of moving households if the Canton of Basel-Stadt harmonized its taxes and applied the city's tax rates in its other communities. The fairly dramatic effects are given in columns 7 and 8 in Table 4 and in the bottom-right map in Figure 5. The two peripheral communities Riehen and Bettingen would lose half of the new households. More interestingly, the average income of households that moved to Riehen and Bettingen would fall by CHF 23,400 and CHF 33,600, respectively. The average income of migrating households would increase by a meager CHF 400 in the city of Basel but by almost CHF 2000 in the other communities in the periphery.

It is important to keep in mind that the simulated effects neglect three potential sources of bias. Firstly, the estimated model neglects the possible migration to the center from households currently living in the periphery. Secondly, households that did not move under the actual situation may decide to move after the policy change and actually moving households may remain in their homes. Thirdly, the estimated model ignores that community characteristics such as housing prices and public goods provision would adjust to policy changes. However, this reaction is likely to be a long-run effect while moving households immediately adapt their community choice.

7 Conclusions and Outlook

Theoretical models of urban community systems postulate endogenous segregation of the population by incomes. The location choice of households with differing incomes is supposed to be cause and consequence of the local tax rate differentials. This paper empirically assesses the second causal connection by studying the community choice of households in a Swiss metropolitan area.

The estimation results show that rich households are substantially and significantly more likely to move to low-tax communities than poor households. This self-sorting of the migrating population perpetuates the existing income differentials across communities. The higher esteem of rich households for low taxes is partly explained by the progressivity of the local income tax. However, there is evidence that rich households prefer low-tax communities to a greater extent than is explained by the tax scheme.

The observed sorting of the population by incomes could possibly be influenced by factors not considered in this study. A promising alternative explanation is 'social interaction': If rich households preferred to live near other rich households a similar segregation pattern would emerge. Such neighborhood interaction was found by Ioannides and Zabel (2002) and Bayer, McMillan and Rueben (2002). Unfortunately, the present data do not allow to discriminate between the two explanations because the average local income levels and the local tax rates are almost multicollinear. A possible way to overcome this problem lies in inspecting the location choice below community level. The differences of average incomes across districts within the same community can be exploited to identify both the effects of the tax rates and of the neighborhood characteristics.

A possible further extension of this study is the collection of more information on potential locational factors, such as school quality. This additional information can be used to study more interaction effects with household characteristics. School quality for example is most likely an important locational factor for families with children, but not for single households. However, as argued in the beginning of Section 3, differences in the provision of public goods do not seem essential in Switzerland.

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	Swiss Federal Tax Administration, Direkte Bundessteuer 1993/94 - Gemeinden.
Commuter	Swiss Federal Statistical Office, Census 1990.

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 scheme data): The tables from the Swiss Federal Tax Administration report the totalled cantonal and communal 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 communities. The center was taken as the middle of the maximal east-west and north-south extensions.

References

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

Bayer, Patrick, Robert McMillan, and Kim Rueben (2003), An Equlibrium Model of Sorting in an Urban Housing Market: A Study of the Causes and Consequences of Residential Segregation, mimeo, Yale University.

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.

Bhat, Chandra R. (2001), Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model, Transportation Research B, 35, 677-693.

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.

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

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. (1999), Steuerwettbewerb und seine Auswirkungen auf Allokation und Distribution: Eine empirische Analyse für die Schweiz, Dissertation Nr. 2222, Universität St. Gallen.

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.

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.

Hansen, Nico A. and Anke S. Kessler (2001a), The Political Geography of Tax H(e)avens and Tax Hells, American Economic Review, 91(4), 1103-1115.

Hansen, Nico A. and Anke S. Kessler (2001b), (Non-)Existence of Equilibria in Multi-Community Models, Journal of Urban Economics, 50, 418-435.

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.

Heiss, Florian (2002), Structural choice analysis with nested logit models, The Stata Journal, 2(3), 227-252.

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

Hilber, Christian (1998), Auswirkungen staatlicher Massnahmen auf die Bodenpreise. Eine theoretische und empirische Analyse der Kapitalisierung, Chur: Rüegger.

Ioannides, Yannis M. and Jeffrey E. Zabel (2002), Interactions, Neighborhood Selection and Housing Demand, Department of Economics Discussion Paper 2002-08, Tufts University, [Online]. Available: http://ase.tufts.edu/econ/papers/200208.pdf

Judd, Kenneth L. (1998), Numerical Methods in Economics, Cambridge: MIT Press.

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.

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 (forthcoming), Assessing the Importance of Tiebout Sorting: Local Heterogeneity form 1850 to 1990, American Economic Review.

Rose-Ackerman, Susan (1979), Market Models of Local Government: Exit, Voting, and the land market, Journal of Urban Economics, 6, 319-337.

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.

Schmidheiny, Kurt (2002a), Income Stratification in Multi-Community Models, VWI Diskussionsschrift 15-02, Universität Bern. [Online]. Available: http://www.vwi.unibe.ch/publikationen/download/dp0215.pdf.

Schmidheiny, Kurt (2002b), Equilibrium and Stratification with Local Income Taxation when Households Differ in both Preferences and Incomes, VWI Diskussionsschrift 02-16, Universität Bern. [Online]. Available: http://www.vwi.unibe.ch/publikationen/download/dp0216.pdf.

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.

Walker, Joan (2002), The Mixed Logit (or Kernel Logit) Model: Dispelling Misconceptions of Identification, Transportation Research Record, 1805, 86-98.

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