Tax-induced Mobility:
Evidence from a Foreigners' Tax Scheme in Switzerland

Kurt Schmidheiny[a] and Michaela Slotwinski[b]

June 2, 2018

Abstract

We study location choice and residential mobility responses to local income taxes exploiting a special tax regime which applies to foreign employees residing in Switzerland. The institutional setting used generates a deterministic duration threshold at 5 years of stay in the country, at which the local tax rates an individual faces simultaneously change in all municipalities. We exploit this exogenous variation by applying a fuzzy regression discontinuity design to merged survey and administrative individual-level data. A dynamic location choice model allows us to derive testable hypotheses of individuals' location choices and mobility decisions. Our estimated treatment effects provide causal evidence for tax-induced residential choices and tax induced intra-national mobility.

[a]University of Basel, Faculty of Business and Economics, Peter Merian-Weg 6, 4002 Basel, Switzerland. Phone: +41 (0)61 267 33 72, email: kurt.schmidheiny@unibas.ch.
[b]University of Basel, Faculty of Business and Economics, Peter Merian-Weg 6, 4002 Basel, Switzerland. Phone: +41 (0)61 207 33 75, email: michaela.slotwinski@unibas.ch.

1 Introduction

Differences in the tax burden across local jurisdictions allow taxpayers to reduce their tax liability by moving to low-tax locations. Such tax-induced mobility creates an additional response margin and potentially leads to stronger reactions of the tax base in fiscally decentralized countries compared to centralized countries. While the basic intuition of a mobile tax base is apparent, there is still remarkably little robust evidence on the importance and magnitude of the mobility margin of the tax base except for special groups, such as soccer players, star scientists or high-income foreign workers. This paper provides causal evidence of tax-induced mobility by exploiting a special foreigners' income tax regime in Switzerland.

The outcomes and effects of income tax reforms depend fundamentally on taxpayers' responses. Individual responses affect the tax base and hence the resulting change in tax revenue. Valid estimates of the elasticity of the tax base with respect to tax rates, as well as the heterogeneous responses of different groups of taxpayers, are therefore key in designing tax policy (see, e.g., Wilson, 1980; Feldstein, 1999, 1995; Mirelees, 1971). For fiscally decentralized countries such as the United States, Canada, Spain, Sweden, Switzerland, and Denmark, the mobility margin is particularly important because mobility within the country is potentially related to the tax burden and could be obstructive to the redistribution goals of a tax system (see, e.g., Mirelees, 1982; Epple and Romer, 1991; Feld, 2000; Feldstein and Wrobel, 1998; Fernandez and Rogerson, 1996; Hodler and Schmidheiny, 2006; Roller and Schmidheiny, 2016).

The theoretical and empirical literature on behavioral reactions to taxation has almost entirely focused on responses in taxable income (see, e.g., Chetty et al., 2011; Saez, 2010; Chetty et al., 2013; Kleven and Waseem, 2013; Spencer and Selin, 2014) and labor supply responses (Saez et al., 2012). More recently, other types of behavioral responses serving to evade or avoid taxes have gained attention (Feldstein, 1999, 1995; Saez et al., 2012). So, while there is ample evidence for reactions along some margins, the literature on tax-induced mobility and location choices is more recent and the effects are less well understood. This might partly be due to the inherent reverse causality between income taxes and their tax base, which makes it difficult to causally identify tax-induced mobility in empirical settings.

Many empirical studies in this field exploit panel variation in tax changes over time or quasi-experimental variation to identify the causal impact of taxes on individual's location choices and mobility decisions (see Esteller et al., 2017, for a review). Recent contributions concentrate on tax-induced mobility of special groups, such as soccer players, star scientists or high-income foreign workers. Kleven et al. (2013) find a strong effect of tax burdens on the international mobility of top earners in the European football market. Kleven et al. (2014) evaluate the effect of a preferential tax regime for high-income foreigners, using Danish administrative tax data, and find that a preferential tax rate for foreigners substantially increased the inflow of high income individuals. Young

and Varner (2011) study the effect of an introduced millionaire tax on the migration of the affected population in New Jersey with a difference-in-differences approach. They report that there is no substantial outflow that could be associated with this change. Akcigit et al. (2016) use panel data on inventors and exploit changes in top tax rates to study the mobility reactions of "superstar" inventors. Relatedly, Moretti and Wilson (2017) exploit the variation in state personal and business taxes to investigate the migration patterns of star scientists in the US. Both document that taxes seem to matter for migration decisions. Finally, Agrawal and Foremny (2017) exploit the fact that a Spanish reform led to income tax heterogeneity among regions. They find no effect of taxes on mobility; however, they document an effect of taxes on location choices, conditional on moving.

We add to this empirical literature by studying how individuals' location choices and mobility decisions react to income taxes in a highly decentralized country. We exploit a special institutional regulation in the Swiss tax law that affects foreign employees. Ordinary income tax rates in Switzerland differ across the 26 cantons and the roughly 2500 municipalities. Foreigners whose yearly gross income is below 120,000 Swiss francs (around 130,000 US-Dollars in 2015) are, however, subject to a special tax regime (Quellenbesteuerung) until they get a permanent residence permit, for which they can apply after five years of stay in Switzerland. For taxpayers in the special foreigners' tax regime, income tax rates do not differ across municipalities within a canton. The way the special income tax rate is determined results in two types of municipalities: high-tax municipalities where the ordinary tax rate is higher than the special tax rate, and low-tax municipalities where the ordinary tax rate is lower than the special tax rate. This institutional arrangement produces a deterministic threshold at five years of stay, which leads to a shift from the special into the ordinary tax regime, leading in turn to a sudden change in the individual's local tax rates. This threshold, and the resulting local randomization, allows us to identify the causal effects of local income tax rates on individuals' location and mobility decisions, using a fuzzy regression discontinuity (RDD) design. Note that our treatment is a pure change in the local tax burden, leaving potential confounding factors such as local public goods provision, housing prices and all other municipality-specific amenities unchanged. Also note that the exploited tax regime change happens exactly once during an individual's stay in Switzerland. The tax rate changes can therefore be seen as permanent from the perspective of the treated individual. Hence, our estimates are not subject to the critique recently formulated by Hennessy and Strebulaev (2015), which states that policy parameters estimated from dynamic policy environments are plagued by biases, as economic actors in such settings form expectations that affect their responses and potentially bias the estimated parameters.

To understand the treatment induced by the two tax regimes and to derive testable hypothesis about its effect, we study a dynamic location choice model with moving costs. Incentives change fundamentally for foreigners in the special tax regime around the duration threshold of 5 years. Foreigners living in a high-tax municipality face a rise in income tax rates while foreigners in low-tax municipalities benefit from a drop in income tax rates.

But the tax regime change simultaneously affects the local tax rates in all municipalities in their choice set. This simultaneous change has to be taken into account when making predictions about how individuals should react, as not only the change in their municipality of origin matters, but also all other related tax changes. The model predicts that the effect of such a simultaneous tax rate change on the probability of choosing a particular municipality is positive for the lowest-tax municipality and negative for the highest-tax municipality. Our model also predicts that the effect on the probability of moving away from a particular municipality is negative in municipalities with above average net-of-tax rates and positive in municipalities with below average net-of-tax rates.

We document strong responses to local income taxes in high income earners' residential and mobility decisions, while lower income groups seem to be unresponsive. High income foreigners receiving the permanent residence permit at the 5 year threshold are substantially and significantly less likely to choose one of the municipalities of the highest tax group and substantially and significantly more likely to move if they have been residing in one of them, compared to their control group just below the threshold value. This constitutes causal evidence for residential choice and mobility responses to income taxation within a country. We find that these effects are strongest for households without children, which arguably have lower moving costs. However, we do not find any significant effect on the moving probability for individuals residing in one of the lowest-tax municipalities, experiencing a tax drop, which suggests an asymmetric response to tax increases and tax decreases which is not in line with our theoretical predictions.

Our theoretical framework allows us to translate the identified fuzzy RDD estimate into an estimate of the migration elasticity with respect to the net-of-tax rate. We estimate migration elasticities that are considerably higher than the values found in the literature so far. However, they are very imprecise and therefore not very informative beyond being significantly positive.

The remainder of the paper is organized as follows: Section 2 gives a detailed description of the institutional setting and Section 3 presents a dynamic multinomial location choice model. The empirical strategy applied is briefly described in Section 4. The data is described in Section 5, the estimated treatment effects are presented in Section 6 and discussed in Section 7. Section 8 concludes.

2 Institutional background

Switzerland is a natural laboratory for research on local taxation, due to the huge variation in tax rates down to the municipal level. The Swiss Confederation consists of three organizational layers: the federal level at the top, the cantonal layer of 26 cantons, and around 2500 municipalities that make up the municipal layer. All three layers have considerable autonomy in both setting tax rates and deciding on expenditures.

(a) ordinary regime   **(b)** special regime
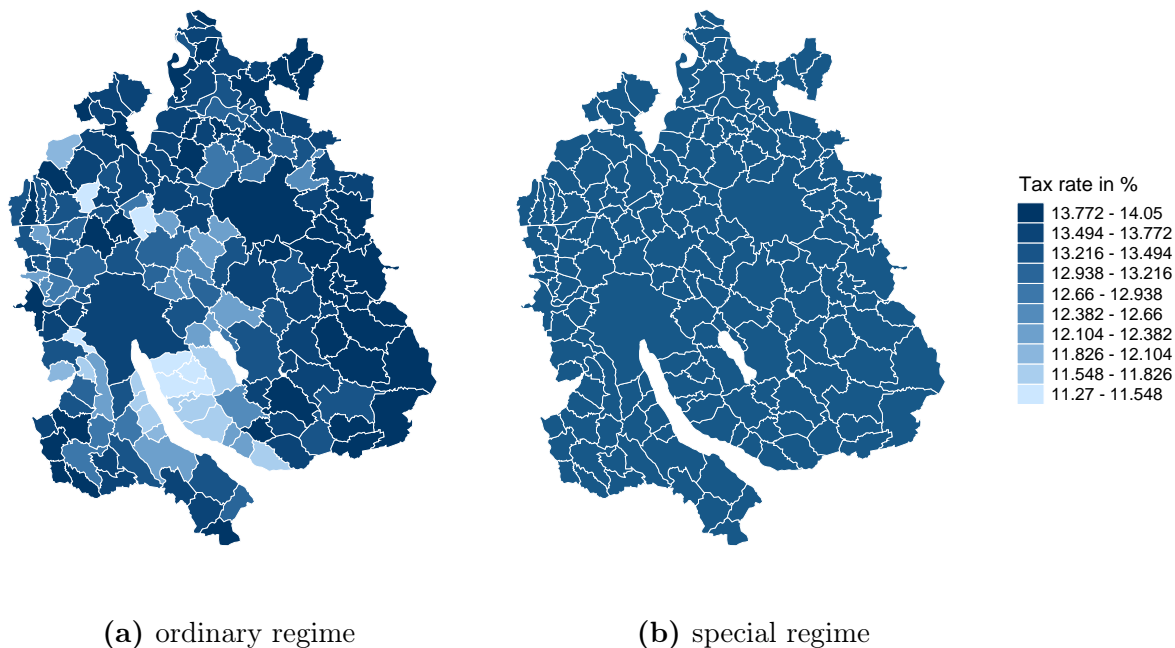
Figure 1: Average income tax rates in the canton of Zurich, for a single household without children and a gross income of 100,000 Swiss francs in the year 2010.

2.1 Ordinary tax regime

The income tax liability for a person with a tax domicile in Switzerland is composed of three parts:[1] The federal income tax, the cantonal income tax and the municipal income tax. The progressivity of tax rates is determined by the federal and the cantonal tax schedules. The municipal tax rate is calculated as the cantonal tax rate times a tax multiplier (*Steuerfuss*) set by the municipality. The total tax liability from all three parts depends on household type and income class. Taxpayers file an annual tax declaration with this information and pay their tax retrospectively. The described *ordinary* tax regime results in substantial differences in income tax rates across Swiss municipalities. The left map in Figure 1 shows the ordinary tax rates for a high-income individual in the canton of Zurich as an example. As this figure demonstrates, income taxes, even within a canton, exhibit a substantial variation. So, for example, a single individual with an income of 100,000 Swiss francs living in the city of Zurich in 2010 could reduce his tax burden from about 13.71% to 11.48% by just moving to Zollikon. Consequently, he could reduces his tax liability by about 2.23%, and thus about 2200 Swiss francs a year, by just moving about 6 km and accepting a train ride of 10 min to Zurich.

---

[1]In some cantons a church tax is also added. Depending on the canton this can be mandatory or voluntary.

2.2 Special tax regime for foreigners

Besides the ordinary tax regime, there is a *special* tax regime ("Quellensteuer") which has been applied nationwide since 2001. Income tax payments in the special regime are directly subtracted from the salary and collected by the employer (i.e. withheld at source). As a result, all labor income of foreigners within this tax regime is third-party reported. Household characteristics such as marital status and children, which affect deductibles, are taken into account such that taxpayers in the special regime do not have to, and are not meant to, file a tax declaration.

The special tax regime applies to incomes from employed *foreign* persons who have their tax domicile in Switzerland, have not yet received the Swiss permanent residence permit (type C), and are not married to a Swiss citizen or to a foreign person who holds a permanent residential permit.[2] Foreigners holding a temporary residence permit (type B) can apply for a permanent residence status (permit C) after a five-year continuous stay in Switzerland.[3] All foreigners with yearly gross income above 120,000 Swiss francs are subject to the ordinary tax regime.[4]

In the special tax regime, there is a canton-wide municipal and cantonal tax rate. This canton-wide tax rate is calculated as a weighted average of the municipal rates within the canton for a given income and household type. Foreigners subject to this special tax regime therefore face no variation in income tax rates across municipalities within a canton. The right map in Figure 1 shows the special tax rates for a high-income individual in the canton of Zurich as an example. The average income tax rate for an unmarried foreign worker without children and a gross income of about 100,000 Swiss francs amounts to 13.39 % in the special tax regime and between 11 and 14 % in the ordinary tax regime. The local ordinary tax rate will be higher than the canton-wide special tax rate in some municipalities and lower in others. We will refer to the former municipalities as *high-tax* municipalities and the latter as *low-tax* municipalities.

Figure 2 visualizes the tax rate in both tax regimes for the highest tax and the lowest tax municipality in 2010 in the canton of Zurich over the entire income distribution up to 120,000 Swiss francs. The tax rate in the at-source tax scheme lies somewhere in between for the whole income range. The ranking of municipalities within cantons by tax rates is the same across all incomes because the municipalities only set a constant tax multiplier.

---

[2]Foreigners who cross the border daily are also taxed at a special tax rate, but this is not of interest in this study, thus an explanation will be omitted here.

[3]The five year limit applies to foreigners from Belgium, Germany, Denmark, France, Liechtenstein, Greece, Italy, The Netherlands, Austria, Portugal, Spain, Finland, United Kingdom, Ireland, Iceland, Luxembourg, Norway, Sweden, USA, Canada, Andorra, Monaco, San Marino and Vatican City. For foreigners from all other countries the normal limit is ten years, but they might be also allowed to apply for a permanent residence permit after five years if they meet certain integration conditions. We restrict our sample to origin countries for which the 5 year threshold applies and further exclude citizens of the United States, as it taxes the worldwide income of its non-resident citizens. Incentives arising from the two tax regimes are consequently very different for these individuals.

[4]In the canton of Geneva the limit is 500,000 Swiss francs. Geneva is therefore ignored in the empirical analysis. We further ignore Basel-Stadt because at-source taxed persons can request to be compensated if the at-source tax rate is higher than the ordinary one.
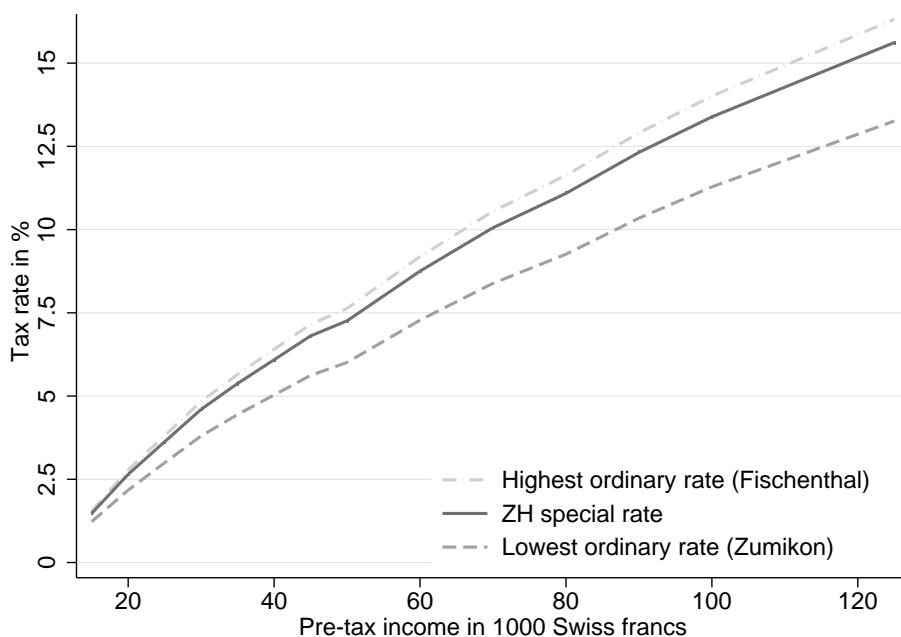
Figure 2: Combined federal, cantonal and municipal tax rate for a single household with no children over the income range in the canton of Zurich in the year 2010.

The classification of municipalities into low-tax and high-tax entities does therefore not depend on income.

2.3 Thresholds from the two tax regimes

The two tax regimes produce two thresholds: an income threshold and a duration threshold. We will exploit the latter in this paper.[5] The *duration* threshold occurs after five years of stay in Switzerland, when an individual with an income below 120,000 Swiss francs is eligible for permanent residence status (permit C). The permanent residence status leads to a shift from special into ordinary taxation.[6] Application for a permit C is a legal right after five years of residence and in principle not an obligation. The *duration* threshold therefore leads to a fuzzy regression discontinuity design (fuzzy RD or FRD) in which the treatment (ordinary tax regime) can only be taken up if the assignment variable (duration of stay) crosses the threshold (five years). The permanent residence status (permit C) is very beneficial to workers: In contrast to permit B (the temporary residence permit), permit C is unconditional, does not expire, does not depend on employment status or whether a child or person in their charge depends on welfare. Holders of permit B have to periodically ask their employers to apply for a renewal of the permit. This renewal is not automatic and might be declined. Permit C basically equates foreigners with natives

---

[5]An earlier version of this paper (Schmidheiny and Slotwinski, 2015) documents that individuals strategically adapt their income around the income threshold. However, in this study we concentrate on the second threshold and exclude the potential impact of the income threshold by excluding the bunching (or manipulation) income range ($120k \pm 6k$) from our estimation sample. We briefly discuss this income adaption in Section 7.

[6]The new tax regime is applied starting from the following month, and individuals have to file a tax declaration at the end of the year, just as Swiss residents do.

on the labor market. For example, foreigners holding the permanent residence permit are allowed to start their own business, can stay abroad for 2 years without losing their status and do not have to leave Switzerland when they become unemployed. Mobility and living conditions within the country are, however, not affected by the residence status. Many cantons automatically assign the permanent residence permit to foreigners as soon as they are eligible, and individuals cannot legally refuse to accept it. These advantages and regulations very likely prevent a strategic self selection into permit C status and thus into the special tax regime for longer. The long-run proportion of those who hold permit C is consequently very high. In our estimation sample, about 95 % of individuals residing in Switzerland for longer than 10 years report holding the permanent residence permit.[7]

The duration threshold creates a quasi-experiment in which foreign employees who crossed the threshold and applied for permit C form a treatment group that is exposed to tax rates that differ across space, while foreign employees who have not yet crossed the threshold form a control group that is exposed to a constant tax rate across space.

3 Theoretical considerations

This section studies a dynamic model of the choice between many municipalities to understand the treatment induced by the two tax regimes and to derive testable hypotheses about its effect on residential and mobility decisions.

3.1 A dynamic location choice model

Consider a household i's choice between M municipalities $m = 1, ..., M$ in two time periods $t = 1, 2$. Municipalities are characterized by the individual-specific municipal tax rate τ_{imt}, the municipal-specific housing price level p_{mt}, and other municipality-specific amenities a_{imt} which may be individual-specific.

Household i receives the utility U_{imt} when choosing municipality m in period t. The household chooses the municipality m that yields the highest utility,

$$U_{imt} > U_{ist} \text{ for all } s \neq m. \tag{1}$$

The chosen municipality of household i in period t is recorded in y_{it}. In period $t = 2$, the household's utility is given by

$$U_{im2} = V_{im2} - c \cdot \mathbb{1}[y_{i2} \neq y_{i1}] + \epsilon_{im2} \tag{2}$$

where $c > 0$ is a *moving cost* and $\mathbb{1}[y_{i2} \neq y_{i1}]$ is an indicator function that takes the value 1 if the household moves between periods 1 and 2 (see, e.g., Moretti and Wilson, 2017).

---

[7]Individuals shifting to ordinary taxation have to fill in an income tax declaration as do Swiss residents in the following year and taxes are no longer withheld by their employer. We expect this change to be quite salient, especially for those shifting due to the receipt of the permanent residence, as they have been accustomed to source taxation for five years.

The deterministic part $V_{imt} = V_{imt}(T_{imt}, p_{mt}, a_{imt}, x_{it})$ is a function of the municipal net-of-tax rate $T_{imt} = log(1 - \tau_{imt})$, the local housing price p_{mt}, other amenities a_{imt}, as well as individual characteristics in x_{it}. We assume that utility V_{imt} strictly increases with the net-of-tax rate T_{imt} and that this effect is homogeneous across municipalities m:[8]

$$\frac{\partial V_{imt}}{\partial T_{imt}} = \beta_i > 0. \tag{3}$$

The idiosyncratic error term ϵ_{imt} captures taste shocks or family events. We assume that ϵ_{imt} follows a type-1 extreme value distribution as in McFadden (1974) and that ϵ_{imt} is independent across time periods.

The choice probabilities in period $t = 2$ are conditional on the municipality chosen in period $t = 1$. The conditional probability that a household i in origin municipality m remains in the same municipality is

$$P_{im2|im1} = \frac{e^{V_{im2}}}{e^{V_{im2}} + e^{-c} \sum_{q \neq m} e^{V_{iq2}}} \tag{4}$$

while the probability of moving from another origin municipality $s \neq m$ to destination municipality m is

$$P_{im2|is1} = \frac{e^{-c} e^{V_{im2}}}{e^{V_{is2}} + e^{-c} \sum_{q \neq s} e^{V_{iq2}}}. \tag{5}$$

Note that the higher the moving costs c, the higher the probability of staying in the same municipality. In the limit as $c \to \infty$, the probability of staying is one, $P_{im2|im1} \to 1$ and $P_{im2|is1} \to 0$ for all $s \neq m$. In the absence of moving costs, the choice probabilities do not depend on the origin municipality, i.e.

$$P_{im2}^{nc} = \frac{e^{V_{im2}}}{\sum_{q=1}^{M} e^{V_{iq2}}}. \tag{6}$$

This is also the choice probability for a household that faces moving costs to all municipalities, for example, a household that newly arrives in the region.

3.2 Treatment

In our empirical strategy, we compare a *control group* that faces identical tax rates across all municipalities, $\tilde{\tau}_{i2}$ (the special tax regime) with a *treatment group* that faces varying municipal tax rates τ_{im2} (the ordinary tax regime). Our treatment is a *simultaneous* change of the local income tax rate in *all* municipalities m for individual i. The change

---

[8] The deterministic part V_{imt} could, for example, be specified as

$$V_{imt} = \alpha + \beta log(1 - \tau_{imt}) - \gamma log(p_{mt}) + \delta log(a_{imt}) + \theta log(x_{it}),$$

where x_{it} is a vector of household characteristics including income. Such a log-linear specification is directly derived from a utility maximizing household with income x_{it} and Cobb-Douglas utility over housing, another private good and a local amenity (see, e.g., Schmidheiny, 2006).

for an individual i in municipality m is $\Delta\tau_{im2} = \tau_{im2} - \tilde{\tau}_{i2}$. Without loss of generality, we order the municipalities by this change

$$\Delta\tau_{i12} \leq \Delta\tau_{i22} \leq ... \leq \Delta\tau_{iM2} \tag{7}$$

where at least one of the inequalities is strict. Equivalently,

$$\Delta T_{i12} \geq \Delta T_{i22} \geq ... \geq \Delta T_{iM2}. \tag{8}$$

Our institutional setting guarantees that this municipality order is the same for all households i. We therefore call municipalities with $\tau_{im2} < \tilde{\tau}_{i2}$ *low-tax* municipalities and municipalities with $\tau_{im2} \geq \tilde{\tau}_{i2}$ *high-tax* municipalities. The change in the net-of-tax rate is accordingly positive for low-tax municipalities, $\Delta T_{im2} > 0$, and negative for high-tax municipalities, $\Delta T_{im2} \leq 0$.

Note that our treatment is a pure change in the net-of-tax rates T_{im2}, leaving the municipality-specific housing price level p_{mt} and all other municipality-specific amenities a_{imt} unchanged. This ceteris paribus nature of our treatment is the basis of our identification strategy.

3.3 Effect on location choice

Our empirical strategy provides an estimate of the probability of an individual residing in a particular municipality for both the treatment and the control group. In the theoretical model, this is the *unconditional* probability of choosing a particular municipality m in period 2:

$$P_{im2} = \sum_{s=1}^{M} (P_{im2|is1} \cdot P_{is1}). \tag{9}$$

P_{im2} is a function of the M *conditional* probabilities $P_{im2|is1}$ of choosing municipality m and of the M *unconditional* probabilities P_{is1} of residing in municipality s in period 1.

The effect of our treatment on the marginal location choice probability is less straightforward than it may seem at first glance, because the treatment changes the tax rates simultaneously in *all* municipalities. Imagine, for example, a low-tax municipality in which the income tax rate for foreign workers is lowered by the treatment. This municipality is clearly more attractive under treatment than without treatment. However, this increase in attractiveness does not necessarily lead to an increase in the choice probability because other low-tax municipalities may have become even more attractive. Furthermore, the unconditional choice probabilities in period 2 depend on the entire spatial distribution of location choices in period 1 described by the unconditional choice probabilities in period 1.

We show in Appendix A.1 that the effect of a simultaneous change of all tax rates τ_{im2} and hence all net-of-tax rate in logs, T_{im2}, on the probability of choosing municipality m in period 2 is given by

$$\mathrm{d}P_{im2} = \beta_i \sum_{s=1}^{M} \left[P_{is1} P_{im2|is1} \left(\mathrm{d}T_{im2} - \overline{\mathrm{d}T}_{i2|is1} \right) \right] \tag{10}$$

10

where $\mathrm{d}T_{im2}$ is the marginal change in individual i's net-of-tax rate in municipality m, $\overline{\mathrm{d}T}_{i2|is1} = \sum_{k=1}^{M} P_{ik2|is1}\mathrm{d}T_{ik2}$ is the average of all net-of-tax rate changes weighted by the conditional choice probabilities of a household living in municipality of origin s. Equation (10) implies the following testable hypothesis:

Hypothesis 1 (Effect on location choice).
The effect of a simultaneous change of all tax rates on the probability of choosing a particular municipality is (a) strictly positive for the lowest-tax municipality, i.e. $\mathrm{d}P_{i12} > 0$, and (b) strictly negative for the highest-tax municipality, i.e. $\mathrm{d}P_{iM2} < 0$.

Proof follows directly from equation (10) as shown in Appendix A.2.

We show in Appendix A.3 that Hypothesis 1 generalizes to groups of municipalities ordered by tax rates. Hence, the treatment will increase the probability of choosing any group of municipalities formed by the lowest-tax municipalities, while it will decrease the probability of choosing any group of municipalities formed by the highest-tax municipalities.

3.4 Effect on mobility

Our empirical strategy also provides an estimate of the probability of moving for households living in any given municipality. In the theoretical model, this is the probability of moving from municipality m to any other municipality:

$$P_{move|im1} = 1 - P_{im2|im1} = \frac{e^{-c} \sum_{q \neq m} e^{V_{iq2}}}{e^{V_{im2}} + e^{-c} \sum_{q \neq m} e^{V_{iq2}}} \tag{11}$$

where $P_{im2|im1}$ is the probability of staying in municipality m. Note that there is a positive moving probability for any household due to shocks in the idiosyncratic error ϵ_{imt} even if the characteristics of all locations remain constant. The moving probability falls with moving costs c and converges towards 0 as moving costs $c \to \infty$.

We show in Appendix A.4 that the effect of a simultaneous change of all tax rates τ_{im2}, and hence all net-of-tax rates (in logs), T_{im2}, is given by

$$\begin{aligned}
\mathrm{d}P_{move|im1} &= -\beta_i P_{im2|im1} \left(\mathrm{d}T_{im2} - \overline{\mathrm{d}T}_{i2|im1} \right) \\
&= -\beta_i P_{im2|im1}(1 - P_{im2|im1}[1 - e^{-c}])(\mathrm{d}T_{im2} - \overline{\mathrm{d}T}_{i2})
\end{aligned} \tag{12}$$

where $\mathrm{d}T_{im2}$ is the marginal change in individual i's net-of-tax rate in municipality m, $\overline{\mathrm{d}T}_{i2|im1} = \sum_{s=1}^{M} P_{s2|im1}\mathrm{d}T_{is2}$ is the average of the net-of-tax changes weighted by the choice probabilities of a household in municipality of origin m and $\overline{\mathrm{d}T}_{i2} = \sum_{s=1}^{M} P_{is2}^{nc}\mathrm{d}T_{is2}$ is the average of the net-of-tax changes weighted by choice probabilities in the absence of moving costs.

Hypothesis 2 (Effect on moving).

The effect of a simultaneous change of all tax rates on the probability of moving away from a particular municipality is (a) strictly negative in municipalities with net-of-tax rates above average (low-tax), i.e. $\mathrm{d}P_{move|m1} < 0$ if $\mathrm{d}T_{m2} > \overline{\mathrm{d}T}_2$, and (b) strictly positive in municipalities with net-of-tax rates below average (high-tax), e.g. $\mathrm{d}P_{move|m1} > 0$ if $\mathrm{d}T_{m2} < \overline{\mathrm{d}T}_2$.

Proof follows directly from equation (12) as shown in Appendix A.5.

We show in Appendix A.6 that Hypothesis 2 generalizes to groups of municipalities ordered by tax rates. Hence, the effect of a simultaneous change of all tax rates on the probability of moving conditional on staying in any municipality of a particular group is strictly negative for groups where all municipalities have net-of-tax rates above average (low-tax) and strictly positive for groups where all municipalities have net-of-tax rates below average (high tax).

3.5 Migration elasticity

We define the *migration elasticity* as the percentage effect of a percentage change in the local net-of-tax rate $1 - \tau_{im2}$ of an individual municipality m on the unconditional probability P_{im2} of residing in that municipality:

$$\eta_{im} = \frac{\partial log(P_{im2})}{\partial log(1 - \tau_{im2})} = \beta_i \frac{1}{P_{im2}} \sum_{s=1}^{M} P_{im2|is1}(1 - P_{im2|is1})P_{is1}, \tag{13}$$

where the derivation is shown in Appendix A.7. Note that the migration elasticity depends on the entire spatial distribution of location choice probabilities P_{is1} in period 1. In the absence of moving costs, i.e. if and only if $c = 0$, the migration elasticity simplifies to the common expression $\eta_{im} = \beta_i(1 - P_{im2})$.

4 Empirical strategy

As described in Section 2, workers from many European countries and Canada can apply for a permanent residence status (permit C) after 5 years of uninterrupted stay in Switzerland.[9] Foreign residents holding the temporary residence status (permit B), which is granted for a maximum duration of 5 years, normally receive an official letter announcing its upcoming expiration about 3 months before its termination. If they intend to stay in Switzerland they need to file a form on the basis of which the cantonal migration office then decides whether and which permit is granted. Some cantons automatically grant permit C for eligible workers while other cantons only grant permit C after application. Federal regulation stipulates that workers cannot choose their type of residence status and that permit C cannot be renounced because of e.g. tax reasons.[10]

---

[9]See footnote 3 for the list of countries and why we do not include the United States.

[10]See https://www.ejpd.admin.ch/dam/data/sem/rechtsgrundlagen/weisungen/auslaender/weisungen-aug-d.pdf

For foreigners with gross incomes below 120,000 Swiss francs, the permanent residence status (permit C) leads to a switch from the special to the ordinary tax regime. Depending on the municipality of residence, these individuals experience a tax increase or decrease when granted permanent residence status (permit C). The 5-year threshold creates a deterministic threshold of treatment eligibility that allows estimating the causal effect of the change in the tax rate with a regression discontinuity design (RDD). The treatment assignment is fuzzy because not all foreign workers immediately apply for permanent residence status or are not immediately granted it. The duration threshold therefore only determines eligibility for the treatment.

We study two outcome variables motivated by Hypotheses 1 and 2: the probability of residing in a particular municipality and the probability of moving conditional on previously living in a particular municipality. Because our sample size does not allow us to study the choices among a large number of individual municipalities we categorize our outcome variables into a dichotomous choice: the probability of residing in a low-tax municipality and the probability of moving conditional on living either in a low-tax or a high-tax municipality. Hypotheses A.1 and A.2 in the Appendix generalize Hypotheses 1 and 2 to such groups of municipalities.

We argue that workers just before the 5-year threshold are very similar to those just above the threshold in terms of observable and unobservable characteristics such as income, skills and tastes. This leads to an assignment into control and treatment group which is as good as random locally around the threshold. Furthermore, the treatment is a pure change in the local tax rates, leaving all other observable and unobservable municipality characteristics such as housing prices, local public goods provision and other amenities unchanged. Our institutional setting therefore constitutes an exceptionally clean setting to identify the causal effect of local income tax changes on individual location decisions that is not confounded by other individual and municipal differences.

The following two subsections formally describe the fuzzy RDD (FRD) used for estimation and the plots used for illustration. We validate and discuss the identifying assumptions of the fuzzy RDD later in Section 6.

4.1 Fuzzy RDD

The fuzzy RDD identifies average treatment effects for compliers at the threshold. In our application these are foreign workers applying and receiving the permanent residence status (permit C) at the threshold of five years. The estimated effects are consistent under the standard assumptions of the fuzzy RDD: local randomization, a clear jump in the probability of treatment, continuity of conditional expectations, excludability and monotonicity. [11]

---

[11]These identifying assumptions are formally stated in Hahn et al. (2001) and well described in Lee and Lemieux (2010) or Imbens and Lemieux (2008).

Our estimates are standard non-parametric (local linear) fuzzy RDD estimates following Porter (2003). We include sampling weights from the survey design.[12] Under the assumptions given the local average treatment effect (τ) is consistently estimated as[13]

$$\widehat{E(\tau_i|x_i=x_c)} = \frac{\lim_{x\to x_c^+}\widehat{E(y_i|x_i=x)} - \lim_{x\to x_c^-}\widehat{E(y_i|x_i=x)}}{\lim_{x\to x_c^+}\widehat{E(w_i|x_i=x)} - \lim_{x\to x_c^-}\widehat{E(w_i|x_i=x)}} \qquad (14)$$

where y stands for the dependent variable of interest, w represents the treatment indicator, x represents the assignment variable (duration of stay in Switzerland), x_c represents the threshold value (5 years), $\lim_{x\to x_c^+}$ stands for the limit approaching the threshold from the right hand side and $\lim_{x\to x_c^-}$ that coming from the left hand side. In our application, w is set to one if an individual holds permit C and to zero if not. $E(w_i|x_i=x)$ stands for the probability of holding the permanent residence status (permit C) after five years of stay. $E(y_i|x_i=x)$ stands for the conditional expectation of the outcome variable after five years of stay. These limits are estimated by local linear regression using a triangular kernel, which is the optimal weighting function in RDD settings because of its advantageous properties at boundary points (for more details see Hahn et al. (2001) or Fan and Gijbels (1996)). There are different procedures for an automatic optimal bandwidth choice in non-parametric estimation. We choose the bandwidth to be as small as possible such that the resulting curve still looks reasonably smooth. Although the optimal bandwidth would vary in sample size, we choose to interpret the result for a fixed set of bandwidths to keep results comparable. To validate that our results are robust to the bandwidth choice, we report results using several bandwidths, where the last row, or column, in the result tables in the Appendix reports discontinuity estimates based on the optimal bandwidth as proposed in Calonico et al. (2014), and as in McCrary (2008) for the density estimates, respectively.

Fuzzy RDD is often conceptionally linked to instrumental variables (see Angrist and Imbens, 1994; Imbens and Lemieux, 2008). In the context of our application the *first stage* is the probability that an individual is granted permanent residence status (permit C) as a function of his duration of stay in Switzerland, the *second stage* is the residential choice and moving probability, respectively, as a function of the residence status, and the *reduced form* is the residential choice as a function of the duration of stay in Switzerland. We additionally report parametric fuzzy RDD estimates of our basic estimates using

---

[12]We use the results of Harms and Duchesne (2010) to incorporate and correct for sampling weights in the non-parametric LLR estimates. The correction term is estimated by

$$\widehat{\Theta+s} = \frac{N}{\sum_{i=1}^{N}\frac{1}{\Pi_i}} \cdot \frac{\sum_{i=1}^{N}\frac{1}{\Pi_i}\cdot\frac{1}{\Pi_i}}{\sum_{i=1}^{N}\frac{1}{\Pi_i}},$$

where $\frac{1}{\Pi_i}$ is the inverse probability of inclusion. It comes out as 1.624 in our application.

[13]For details of LLR estimation in the RDD framework see Porter (2003) or Fan and Gijbels (1996) for LLR estimation in general. Imbens and Lemieux (2008) and Lee and Lemieux (2010) provide a very practical guide on the estimation and validation of an RDD.

standard instrumental variable estimation incorporating survey weights.[14] We follow Lee and Lemieux (2010) and Angrist and Pischke (2009) in the specification, and use all relevant observations of individuals within a bandwidth of 4 years around the threshold, i.e. in the widest symmetric estimation window. The results of this specification are reported in parallel to the non-parametric ones for comparison. In general we find that the parametric estimates are somewhat lower, which could be attributed to the fact that in these estimates every individual is weighted equally strongly even those farther away from the threshold.[15]

Importantly, the discontinuity identifies the effect of the treatment for the respective groups around the threshold. It is thus, in our application, the change in the outcome variable resulting from the simultaneous change in the income tax rates at the threshold value.

4.2 Discontinuity plots

As graphical illustration of the reduced form relationship, we fit a local linear estimate to the residential choice of individuals (e.g., an indicator variable which takes the value 1 if the individual lives in a low-tax municipality and 0 otherwise), or their probability of taking up the treatment (permit C) respectively, separately to both sides of the duration threshold of 5 years, or 1825 days, in the country in a sharp RDD manner. Therefore, we apply the following estimation equation to every estimation point separately to both sides of the threshold x_c.

$$\min_{\alpha\beta} \sum_{i=1}^{N} (y_i - \alpha - \beta(x_i - x))^2 K_h (x_i - x) \tag{15}$$

where y is the dependent variable of interest, x_i represents the commonly labeled assignment variable (the duration of stay in our case), h represents the suitable bandwidth $h > 0$, $K_h (x_i - x)$ represents the kernel weighting function at the estimation point x and is defined as $K_h (x_i - x) = \frac{1}{h} K \left(\frac{x_i - x}{h} \right)$. We use a triangular kernel and take the sampling weights in our data into account.

4.3 Structural estimation

Our institutional setting allows us to identify the local average effect of our tax treatment. However, the magnitude of this treatment is not straightforward to interpret because our treatment is a simultaneous change in the tax rates in all municipalities. We use the theoretical model in Section 3 to estimate the structural parameter β and the implied migration elasticity η which can be related to other estimates in the literature.

---

[14]More precisely, we regress the outcome variable on duration and a treatment dummy by two-stage least squares instrumenting the treatment dummy by a dummy for being above the duration threshold and allowing for different slopes in the duration of stay on both sides of the threshold.

[15]For these estimates we follow the specification in Angrist and Pischke (2009)[p.261ff] where, we choose the polynomial order to be one or two.

Equation (10) relates the observed effect on location choice from our treatment to the fundamental parameters of the model. Solving for β yields

$$\beta_i = \frac{\mathrm{d}P_{im2}}{\sum\limits_{s=1}^{M}\left[P_{is1}P_{im2|is1}\left(\mathrm{d}T_{im2} - \overline{\mathrm{d}T}_{i2|is1}\right)\right]} \tag{16}$$

where $\overline{\mathrm{d}T}_{i2|is1} = \sum_{k=1}^{M} P_{ik2|is1}\mathrm{d}T_{ik2}$.

Taking the parameter β_i as well as the choice probabilities and the tax rate changes as constant for a specific estimation sample, we estimate β as

$$\widehat{\beta} = \frac{\widehat{\mathrm{d}P}_{m2}}{\sum\limits_{s=1}^{M}\left[\widehat{P}_{s1}\widehat{P}_{m2|s1}\left(\widehat{\mathrm{d}T}_{m2} - \widehat{\overline{\mathrm{d}T}}_{2|s1}\right)\right]} \tag{17}$$

with $\widehat{\overline{\mathrm{d}T}}_{2|s1} = \sum_{k=1}^{M} \widehat{P}_{k2|s1}\widehat{\mathrm{d}T}_{k2}$ where $\widehat{\mathrm{d}P}_{m2}$ is the fuzzy RDD estimate for the estimation sample, $\widehat{P}_{s1}$ and $\widehat{P}_{m2|s1}$ are, respectively, estimates of the unconditional and conditional location choice probabilities from observed choice frequencies, and $\widehat{\mathrm{d}T}_{m2}$ is the average individual tax change in municipality m in the estimation sample. Alternatively, we base our structural estimation of β on the effect of moving, theoretically derived in equation (12):

$$\widehat{\beta} = -\frac{\widehat{\mathrm{d}P}_{move|m1}}{\widehat{P}_{m2|m1}\left(\widehat{\mathrm{d}T}_{m2} - \widehat{\overline{\mathrm{d}T}}_{2|m1}\right)} \tag{18}$$

with $\widehat{\overline{\mathrm{d}T}}_{2|m1} = \sum_{s=1}^{M} \widehat{P}_{s2|m1}\widehat{\mathrm{d}T}_{s2}$. The estimate $\widehat{\beta}$ in equations (17) and (18) is m-specific and hence there are M possible estimates.

With a structural estimate of β at hand, we estimate the migration elasticity for each municipality m from equation (13) as

$$\widehat{\eta}_m = \widehat{\beta}\frac{1}{\widehat{P}_{m2}}\sum_{s=1}^{M}\widehat{P}_{m2|s1}(1 - \widehat{P}_{m2|s2})\widehat{P}_{s1}. \tag{19}$$

We report the average elasticity across all municipalities weighted by the population share $\widehat{P}_{m2}$ in our sample, i.e. $\widehat{\eta} = \sum_{m=1}^{M} \widehat{P}_{m2}\widehat{\eta}_m$. 95% confidence bounds for the structural estimates are computed by a standard non-parametric bootstrap, using the percentile method. Appendix B shows how we generalize the estimation of $\widehat{\beta}$, $\widehat{\eta}_m$ and $\widehat{\eta}$ to groups of municipalities and to pooling across cantons.

4.4 Choice set and spatial pooling

Our institutional setting features an ordinary tax regime with different municipal tax rates and a special tax regime with identical municipal tax rates *within a canton.* Our theoretical model describes the choice of a municipality based on local tax rates and other local and individual characteristics. Such a location choice model is best seen as a model of choosing the place of residence given a workplace in the local labor market. To tractably

match the institutional setting with the location choice model, we use the canton as the choice set when we take the model to the data. For many large Swiss cities such as Zurich and Bern, the canton indeed covers indeed most parts of the relevant commuting zone.[16]

As our sample is rather small, we pool the location choices over all cantons when we estimate the probability of choosing the group of low-tax or high-tax municipalities in the fuzzy RDD. We therefore define high- and low-tax municipalities within a canton as described in Section 5.

In the structural estimation of β and the migration elasticity η_m in equations (17) to (19), conditional and unconditional choice probabilities for individual municipalities m, $P_{m2|s1}$ and P_{m2}, respectively, are therefore also defined relative to the cantonal choice set and sum up to 1 within each canton. Choice probabilities are estimated by their sample counterparts in the control sample, i.e. individuals with a duration of stay of up to 5 years.

5   Data

We use individual data from the largest Swiss labor market survey SAKE ("Schweizerische Arbeitskräfteerhebung") augmented by individual income data from the social security administration.[17] The survey is organized as a rotating panel study. Each person is interviewed annually over a period of five years and then replaced (BFS, 2004). The linked administrative income data is essential for our analysis as we need to observe the income threshold very precisely to define who is subject to the special tax regime and to calculate the tax liability.

The sampling strategy with respect to foreign persons is not only geographically stratified but also adjusted for demographic proportions, groups of foreigners, gender, and duration of stay. Sample weights are therefore important and need to be accounted for in the estimation. The SAKE dataset provides a weight variable which is meant to correct for the stratification and complex survey design. It represents the inverse of the probability of being included in the sample ($w_i = 1/\pi_i$).[18]

We use the pooled dataset for the years 2001 to 2013[19]. The main variables in the empirical analysis are:

---

[16]For the cities of Basel and Geneva this is clearly not the case because large parts of the commuting zones belong to other cantons or even to other countries. However, Basel and Geneva are already excluded from our sample because they deviate in other details from our institutional setting (see footnote 2.2).

[17]This linked dataset is called SESAM ("Synthese Erhebung Soziale Sicherheit und Arbeitsmarkt (SESAM), BFS").

[18]For the years 2001 to 2009, we use the weighting variable *IXPXH*, and for the years after 2010 *IXPXHJ*, since the annual dataset is used. This weight not only corrects for the consequences of the stratification, but is additionally adjusted for the probability of a drop out before being sampled out and further factors not discussed in detail here (for more details with respect to the creation of this weight see BFS, 2004).

[19]We only use data up to 2013, as the practical implementation of the special tax scheme started changing in several cantons due to a 2010 court judgment that this tax practice is discriminatory. However, there is little public documentation on how the practice changed until it was changed on the federal level. Therefore, we decided to use the time span in which it was certainly in force.

- *Location*: This variable records the municipality of legal residence at the moment of the individual is interviewed.

- *Move*: This variable is an indicator variable with a value of one if an individual indicates having moved to another municipality within the last 12 months and zero otherwise.

- *Duration*: This variable contains information about how long a foreign target person has already lived in Switzerland. Respondents are first asked to indicate the year and month they arrived in Switzerland. This information is converted to days using approximations for missing information on the month (for details see BFS, 2012).

- *Permit C*: This variable is an indicator variable taking the value of one if an individual indicates holding the permanent residence permit and zero otherwise.

- *Gross income*: This variable contains the yearly gross income of an individual, i.e. income before taxes and social security contributions within the last 12 months including 13th or 14th monthly salary, premiums and bonuses.

Additionally, we use several demographic variables for the definition of the sample and the treatment groups, where needed.[20] Descriptive statistics on the main variables can be found in Table C.1.

Our basic estimation sample comprises individuals who meet all requirements for the special tax scheme, except for the income, the duration of stay, and the permit type restriction. Our sample includes foreign employees whose employment status is known, who are at least 18 years old, and whose location of residence as well as nationality is known.[21] Swiss citizens including dual citizens and self-employed individuals are excluded as this leads to ordinary taxation. The sample is further restricted to those holding either permit B or C, excluding cases like asylum seekers. We also exclude those who are married to a Swiss citizen or a dual citizen since they would be taxed ordinarily.[22] We further restrict our sample to foreigners residing in Switzerland for at least one year, so that we can observe their yearly income. Individuals who own real estate are excluded as this leads to ordinary taxation in the majority of cantons.[23] Further, we restrict our sample to those nationalities that are allowed to apply for permanent residence after 5 years of stay, as our analysis will concentrate on the duration threshold of 5 years stay. We use only observations for which the indicated duration of stay is reasonable, i.e. we exclude observations for which the duration of stay in the same municipality exceeds the duration

---

[20]These are in particular: age, employment status, nationality, permit type, marital status, nationality of spouse, information on real estate property, canton and municipality of residence.

[21]We exclude individuals who could not respond correctly when asked where their location of residence is. Further, proxi-interviews are excluded, i.e. observations of individuals who were not interviewed in person.

[22]One exception applies for the canton Ticino. Individuals married to a Swiss citizen are not excluded if they reside in Ticino, because there this does not lead to ordinary taxation.

[23]This regulation applies to the cantons Argau, Vaud, Geneva, Jura, Lucerne, Uri, Schwyz, Nidwalden, Glarus, Freiburg, Solothurn, Basel, Basel District and Schaffhausen.

of stay in Switzerland by more than half a year. We further restrict the choice set to the canton of residence and therefore only analyze individuals staying within a canton.[24]

We also exclude a range ±6000 Swiss francs around the income threshold of 120,000 Swiss francs from our analysis because we detected systematic income adaption in a small band around the income threshold as discussed in Section 7.2.[25] The *experimental* sample, i.e. the individuals that switch from being untreated to being eligible for treatment after five years, hence consists of foreign workers with incomes up to 114,000 Swiss francs.

As we would expect that reactions to income taxes are different across incomes, we construct two income groups in the experimental sample. We define the *low-income* group as having an income of below 60,000 Swiss francs (which is about the median income in the experimental sample), while the *high-income* group has an income above 60,000 and below 114,000 Swiss francs. In the validation section we additionally use a *placebo* sample of individuals with income between 126,000 and 200,000 Swiss francs. This group is already taxed ordinarily from the beginning of their stay in Switzerland and should therefore not react at the duration threshold. We exclude incomes above 200,000 Swiss francs to exclude effects from extremely high incomes.

In addition to the described individual data, we use data on the combined federal, cantonal and municipal average tax rate in the ordinary tax regime for the years 2001 to 2013. These tax rates are calculated by the Swiss Federal Tax Administration for the 813 largest Swiss municipalities. The calculation considers cantonal statutory tax rates as well as canton-specific deductions, e.g., for children. The municipal tax rate is added as the cantonal tax rate times the local tax multiplier. The federal tax is homogeneous across Switzerland and is added to the cantonal and the municipal tax. These data have been completed by Parchet (2014) to cover all municipalities in Switzerland and provide the information on the income tax rates in Swiss municipalities in the ordinary tax scheme.

The income tax in the special tax scheme is calculated by the weighted income tax rate for income groups within a canton. It is weighted by the population share of the municipality relative to the cantonal population.[26]

The limited number of observations in our sample does not allow us to estimate the choice among a large number of individual municipalities. We therefore categorize the municipalities into two and four groups of municipalities based on the local tax rate:

- *Low-tax*: This is an indicator variable set to one if the income tax rate of the municipality in the ordinary tax scheme is lower than the tax rate in the special tax scheme of the corresponding canton and year.

---

[24]About 98% of individuals in our sample stay within the same canton.

[25]We additionally exclude individuals having an income in this range in the year before, to make sure that we are not capturing strategic reactions of individuals who potentially systematically adapted their income.

[26]The original tax multipliers for the special tax rate are not available for all cantons. Based on the reported definitions of local tax authorities, we compared this measure to the original one, where available, and it seems to be a valid approximation. The correlation between our generated measure and the original data is about 0.91.

- *High-tax*: This is an indicator variable set to one if the income tax rate of the municipality in the ordinary tax scheme is higher than the tax rate in the special tax scheme of the corresponding canton and year.

- *Very low-tax (vl)*: This is an indicator variable set to one if the income tax rate of the municipality in the ordinary tax scheme is lower than or equal to the median tax rate of low-tax municipalities in a given canton and year.

- *Medium low-tax (ml)*: This is an indicator variable set to one if the income tax rate of the municipality in the ordinary tax scheme is higher than the median tax rate of low-tax municipalities and lower than the tax rate in the special tax scheme in a given canton and year.

- *Medium high-tax (mh)*: This is an indicator variable set to one if the income tax rate of the municipality in the ordinary tax scheme is lower than the median tax rate of high-tax municipalities and higher than the tax rate in the special tax scheme in a given canton and year.

- *Very high-tax* (vh): This is an indicator variable set to one if the income tax rate of the municipality in the ordinary tax scheme is higher than or equal to the median tax rate of high-tax municipalities in a given canton and year.

Our institutional setting guarantees that this classification of municipalities is the same for all individuals despite the progressive nature of local tax rates.

Descriptive statistics of the net-of-tax changes within these groups are reported in Table C.1 in Appendix C.

6   Results

This section reports the results of the individual's reactions at the duration threshold, exploiting the fuzzy RDD setting. We first test the theoretical prediction on location choice and on mobility, and then report structural estimates of the migration elasticity. Subsequently, we discuss the validity of the identifying assumptions behind the fuzzy RDD.

6.1   Treatment Probability at the Duration Threshold

We start by validating one of the major identifying assumptions of our fuzzy RDD setting, namely that the treatment probability exhibits a positive discontinuity at the duration threshold of 5 years (or 1825 days) after arriving in the country.

To see whether there is a clear jump in the probability of being treated, i.e. the probability of holding the permanent residence status (permit C) at the threshold value, we fit two local linear smooths separately to both sides of the threshold value to an indicator set to one for individuals holding permit C. The resulting pattern for the whole sample is visualized in Figure 3. We see a clear positive discontinuity at the threshold value, demonstrating that the treatment probability indeed substantially increases at the
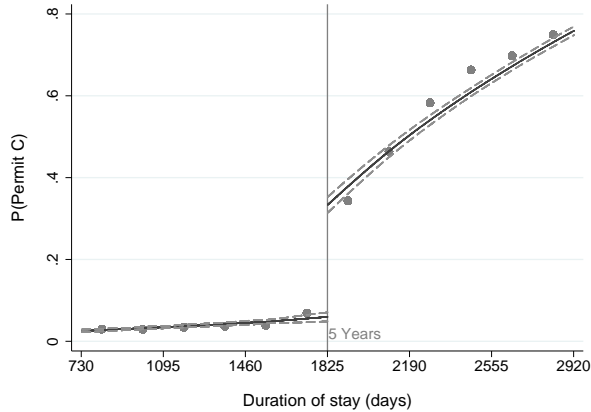
Figure 3: Probability of holding permit C - first stage of fuzzy design. This graph shows local linear smooths of the probability that an individual holds the permanent residence permit around the duration threshold for the whole sample, using a bandwidth of 1460 days. The light gray dashed lines indicate the 90% pointwise confidence bands and the gray dots raw weighted averages of the dependent variable in half yearly bins. The corresponding discontinuity estimate can be found in column (2) of Table C.2 in Appendix C.

threshold. Table C.2 in Appendix C reports the corresponding discontinuity estimate. The discontinuity amounts to about 20 percentage points and is statistically significant. We conclude that the threshold significantly and substantially changes the treatment take-up and therefore constitutes a strong instrument for the treatment status. The discontinuity in the take-up is not substantially different between individuals in high- and low-tax municipalities or between low- and high-income individuals. We also find that there is a considerable discontinuity in the probability to hold permit C at the threshold for our always-treated placebo sample. As they do not face tax related incentives with respect to permit C, this additionally validates that the treatment take-up is unlikely to be driven by tax considerations.

6.2   Effect on location choice

In a next step, we exploit the fuzzy RDD to test whether a change in the tax regime, and thus a simultaneous change in the tax rates of the whole choice set, indeed has the theoretically derived consequences for individual's residential choices (Hypothesis 1 and its generalization to groups of municipalities in Hypothesis A.1).

Given that we use observational data, it is not possible to test Hypothesis 1, and thus the change in the probability of residing in a particular municipality. However, its generalization A.1, can be tested as it makes predictions about reactions in residential choices for groups of municipalities.

We first group the municipalities into low- and high-tax municipalities. Hypothesis A.1 predicts that treated individuals are more likely to reside in low-tax municipalities com-
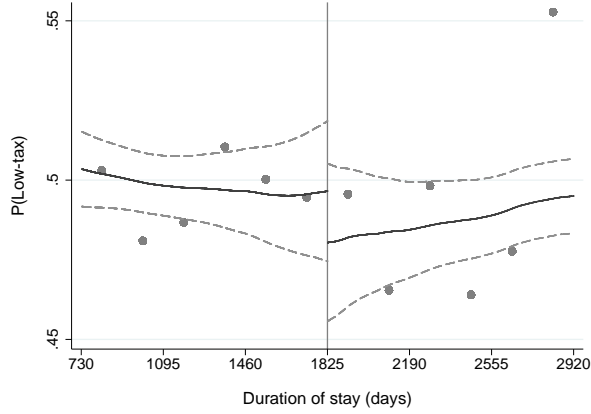
Figure 4: Probability of residing in a low-tax municipality for the experimental sample around the threshold value. The graph shows local linear smooths of the probability that an individual resides in a low-tax municipality for the experimental sample, using a bandwidth of 1460 days (4 years). The light gray dashed lines indicate the 90% pointwise confidence bands and the gray dots raw weighted averages of the dependent variable in half yearly bins.

Table 1: Probability of residing in a low-tax municipality

	All		Low income		High income	
	non-parametric	parametric	non-parametric	parametric	non-parametric	parametric
	(1)	(2)	(3)	(4)	(5)	(6)
Effect ΔP	-0.064	-0.021	-0.159	-0.026	-0.008	-0.023
	(0.096)	(0.076)	(0.180)	(0.140)	(0.112)	(0.086)
N left	7988	7988	3217	3217	4771	4771
N right	4468	4468	1775	1775	2693	2693

Notes: Fuzzy RDD estimates of the probability that an individual resides in a low-tax municipality for different income groups. Nonparametric means fuzzy RDD with local-linear regression using a triangular kernel, while parametric means a linear instrumental variable estimation. All estimates include a bandwidth of 1460 days (4years) and account for sampling weights. Standard errors in parentheses. Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

pared to untreated individuals. Individuals just above the duration threshold are more likely to be treated than individuals just before the threshold and should hence have a higher probability of residing in a low-tax municipality. Figure 4 visualizes the reduced-form relationship between the outcome variable and the eligibility of treatment, i.e. the choice probability as a function of the duration of stay in Switzerland for the experimental sample around the threshold value.[27] The probability that an individual resides in a low-tax municipality is virtually unchanged at the duration threshold of 5 years (1825 days). Table 1 reports the corresponding non-parametric and parametric fuzzy RDD estimate in columns (1) and (2). The point estimates are slightly negative but not significantly different from zero at any usual level of significance.

---

[27]Note that the magnitude and confidence bounds of the discontinuity in the reduced form figures do not equal the magnitude and confidence bounds of the discontinuity estimate in a fuzzy RDD.
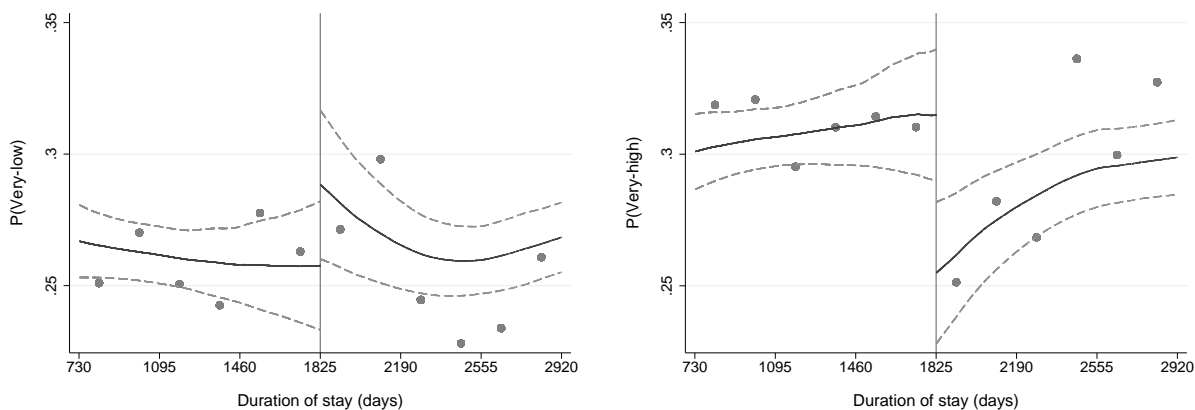
(a) Probability of residing in a
very low-tax municipality

(b) Probability of residing in a
very high-tax municipality

Figure 5: Probability that an individual resides in the respective municipality type for the high-income experimental sample around the threshold value. These graphs show local linear smooths of the probability that an individual resides in the respective municipality type for the high-income experimental sample, using a bandwidth of 1460 days. The light gray dashed lines indicate the 90% pointwise confidence bands and the gray dots raw weighted averages of the dependent variable in half yearly bins.

Income tax rates in Switzerland are highly progressive, as exemplified for the canton of Zurich in Figure 2. We expect a stronger effect for higher income individuals than for lower income individuals because the former are treated with much larger changes in tax rates. Progressive tax rates lead mechanically to stronger effects on location choice for higher income individuals in a location-choice model with homothetic preferences as in footnote 8.[28]

Columns (3) and (4) in Table 1 show the fuzzy RDD estimates for low-income households with incomes below 60,000 Swiss francs, while columns (5) and (6) show those for high-income individuals with incomes between 60 and 114 thousand Swiss francs. Again, we see no systematic effect at the duration threshold.

In many municipalities the tax changes induced by the treatment are very small. We therefore focus next on very low-tax municipalities with large reductions in tax rates and on very high-tax municipalities with large increases in tax rates. Hypothesis 1 and its generalization Hypothesis A.1 (in Appendix A.3) predict that the probability of choosing a very low-tax municipality increases from treatment while the probability of choosing a very high-tax municipality decreases.[29] Figure 5 shows the reduced form effects for these two types of municipalities for high-income individuals, i.e. individuals in the experimental sample with incomes between 60k and 114k Swiss francs. Figure 5(a) documents a small and non-significant rise in the probability of choosing a very low-tax municipality at the duration threshold and Figure 5(b) a significant and substantial drop in the

---

[28]In reality, utility over housing and other goods is typically not homothetic with decreasing budget shares of housing with increasing income. Such non-homothetic preferences are a further reason for the effect to increase with income as discussed in Schmidheiny (2006).

[29]These categories are defined in Section 5.

Table 2: Probability of residing in a very low-tax or very high-tax municipality

		All		Low income		High income	
		non-parametric	parametric	non-parametric	parametric	non-parametric	parametric
		(1)	(2)	(3)	(4)	(5)	(7)
(a)	*Probability of residing in a very low-tax municipality:*						
	Effect ΔP	0.071	0.076	0.009	0.099	0.110	0.065
		(0.085)	(0.064)	(0.153)	(0.122)	(0.101)	(0.071)
(b)	*Probability of residing in a very high-tax municipality:*						
	Effect ΔP	-0.096	-0.093	0.115	0.020	-0.215**	-0.159**
		(0.085)	(0.070)	(0.154)	(0.133)	(0.101)	(0.080)
N left		7988	7988	3217	3217	4771	4771
N right		4468	4468	1775	1775	2693	2693

Notes: Fuzzy RDD estimates of the probability that an individual resides in the respective municipality type for different income groups. Nonparametric means fuzzy RDD with local-linear regression using a triangular kernel, while parametric means a linear instrumental variable estimation. All estimates include a bandwidth of 1460 days (4years) and account for sampling weights. Standard errors in parentheses. Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

probability of choosing a very high-tax municipality. The corresponding fuzzy RDD estimates are presented in Table 2 columns (5) and (6). These estimates report a, not very precisely estimated, though significant ($p < 5\%$) and substantial drop for very high-tax municipalities of about 15 to 20 percentage points and a non-significant increase for very low-tax municipalities. High-income individuals therefore systematically shy away from municipalities in which they would experience a sharp increase in tax rates after being granted permanent residence status. However, high income individuals are, on average, not significantly more likely to choose a very low-tax municipalities which would offer a sharp decrease in tax rates after being granted permanent residence status. This suggests that they rather sort into the two middle categories, i.e. medium high- and medium low-tax municipalities, which offer more attractive tax rates and still reasonable housing prices. Table 2 also reports estimates for all incomes in the experimental group (columns 1 and 2) and for low-income households only (columns 3 and 4). We do not find any systematic reactions in the sample of low-income earners and in the combined samples with all income groups.

Table 3: Probability of moving away from a high- or low-tax municipality

	All		Low income		High income	
	non-parametric	parametric	non-parametric	parametric	non-parametric	parametric
	(1)	(2)	(3)	(4)	(5)	(6)
(a)	*Probability of moving away from a low-tax municipality:*					
Effect ΔP	0.007	0.086	0.081	0.215**	-0.034	-0.001
	(0.061)	(0.053)	(0.118)	(0.106)	(0.069)	(0.053)
N left	4061	4061	1731	1731	2330	2330
N right	2215	2215	944	944	1271	1271
(b)	*Probability of moving away from a high-tax municipality*					
Effect ΔP	0.271***	0.212***	0.144	0.135	0.339***	0.256***
	(0.082)	(0.071)	(0.150)	(0.099)	(0.098)	(0.097)
N left	3927	3927	1486	1486	2441	2441
N right	2253	2253	831	831	1422	1422

Notes: Fuzzy RDD estimates of the probability of moving given that individuals were residing in the respective municipality type for different income groups. Nonparametric means fuzzy RDD with local-linear regression using a triangular kernel, while parametric means a linear instrumental variable estimation. All estimates include a bandwidth of 1460 days (4years) and account for sampling weights. Standard errors in parentheses. Significance levels: * $0.05 < p < 0.1$, ** $0.01 < p < 0.05$, *** $p < 0.01$.

6.3 Effect on mobility

We next study directly whether the change in the tax regime, and thus a simultaneous change in all local tax rates, induces individuals to systematically move away. Hypothesis 2 and its generalization Hypothesis A.2 (in Appendix A.6) predicts that the treatment increases the mobility of individuals who live in municipalities with high ordinary tax rates (below average net-of-tax rate) while it decreases the mobility of individuals who live in municipalities with low ordinary tax rates (above average net-of-tax rate). Hence, individuals move away from places that become less attractive and stay in places that become more attractive.

We repeat the former fuzzy RDD estimates for individuals' mobility decision. The dependent variable is an indicator set to one if an individual moved to another municipality within the last 12 months. Table 3 reports the non-parametric and parametric fuzzy RDD estimates for the whole experimental sample in columns (1) and (2), respectively. Panel (a) shows the effect for individuals living in low-tax municipalities when receiving the permanent residence permit and panel (b) those for individuals living in high-tax municipalities. We find a positive and significant ($p < 0.01$) jump in the probability of moving for those residing in a high-tax municipality of between 22 and 28 percentage points but no clear reactions for individuals residing in low-tax municipalities.

Columns (3) to (6) report separate estimates for low-income and high-income individuals. As in the analysis of residential choices in Section 6.2, we expect a stronger reaction for high income households. We indeed observe a stronger positive jump in the
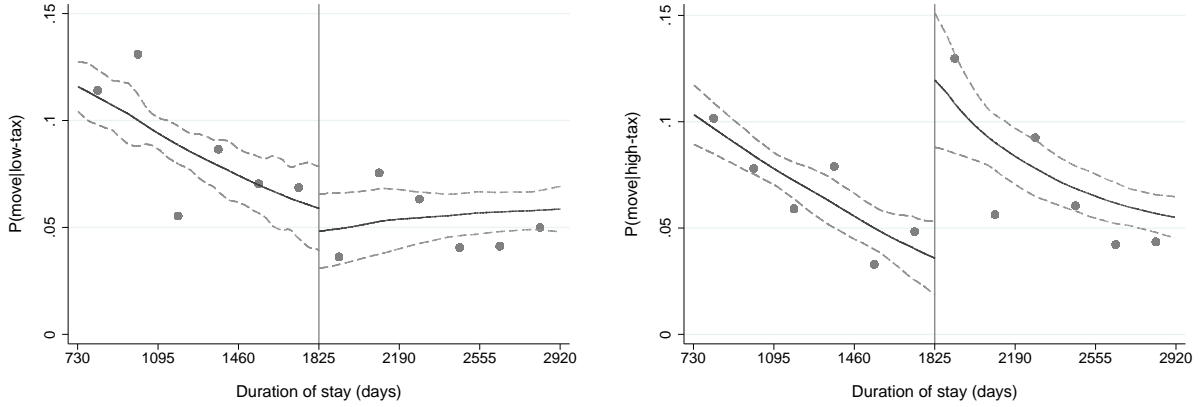
(a) Probability of moving away from a
low-tax municipality

(b) Probability of moving away from a
high-tax municipality

Figure 6: Probability of moving given the individual was residing in the respective municipality type for the high-income experimental sample around the threshold value. These graphs show local linear smooths of the probability of moving conditioning on the previous municipality type for the high-income experimental sample using a bandwidth of 1460 days (4 years). The light gray dashed lines indicate the 90% pointwise confidence bands and the gray dots raw weighted averages of the dependent variable in half yearly bins.

moving probability of up to 34 percentage points for high-income individuals in high-tax municipalities, but again no significant effect for low-income individuals. The graphs in Figure 6 show the reduced form relationship for high-income individuals. We see a steep and significant increase in the moving probability for high-income individuals residing in high-tax municipalities and a very small not significant decrease for high income individuals in low-tax municipalities. Note that we would expect the probability of moving to react only locally around the threshold, as the incentives to move arise just after receiving the permanent residence permit and the moving probability should return to its old level once individuals have adapted to the new circumstances and thus for higher durations of stay.

We next divide the municipalities into four municipality groups: very low-tax, medium low-tax, medium high-tax and very high-tax municipalities. Hypothesis 2 (Appendix A.2) predicts that the probability of moving should fall in both low-tax groups and increase in both high-tax groups. We expect that the effect is stronger in the extreme groups, where the tax rate changes are strongest. Table 4 reports the fuzzy RDD estimates for these four groups in panels (a) to (d) separately for all incomes in columns (1) and (2), for low-income individuals in columns (3) and (4) and for high-income individuals in columns (5) and (6). We see a very strong and significant ($p < 0.01$) positive effect for individuals in the highest-tax municipalities with most of the effect stemming from high-income individuals. According to the non-parametric fuzzy RDD estimate, high-income individuals living in very high-tax municipalities are 35 to 52 percentage points more likely to move when treated after 5 years, with a 95% confidence interval from 23 to 80 percentage points, for the local linear estimates. In line with the theoretical prediction,
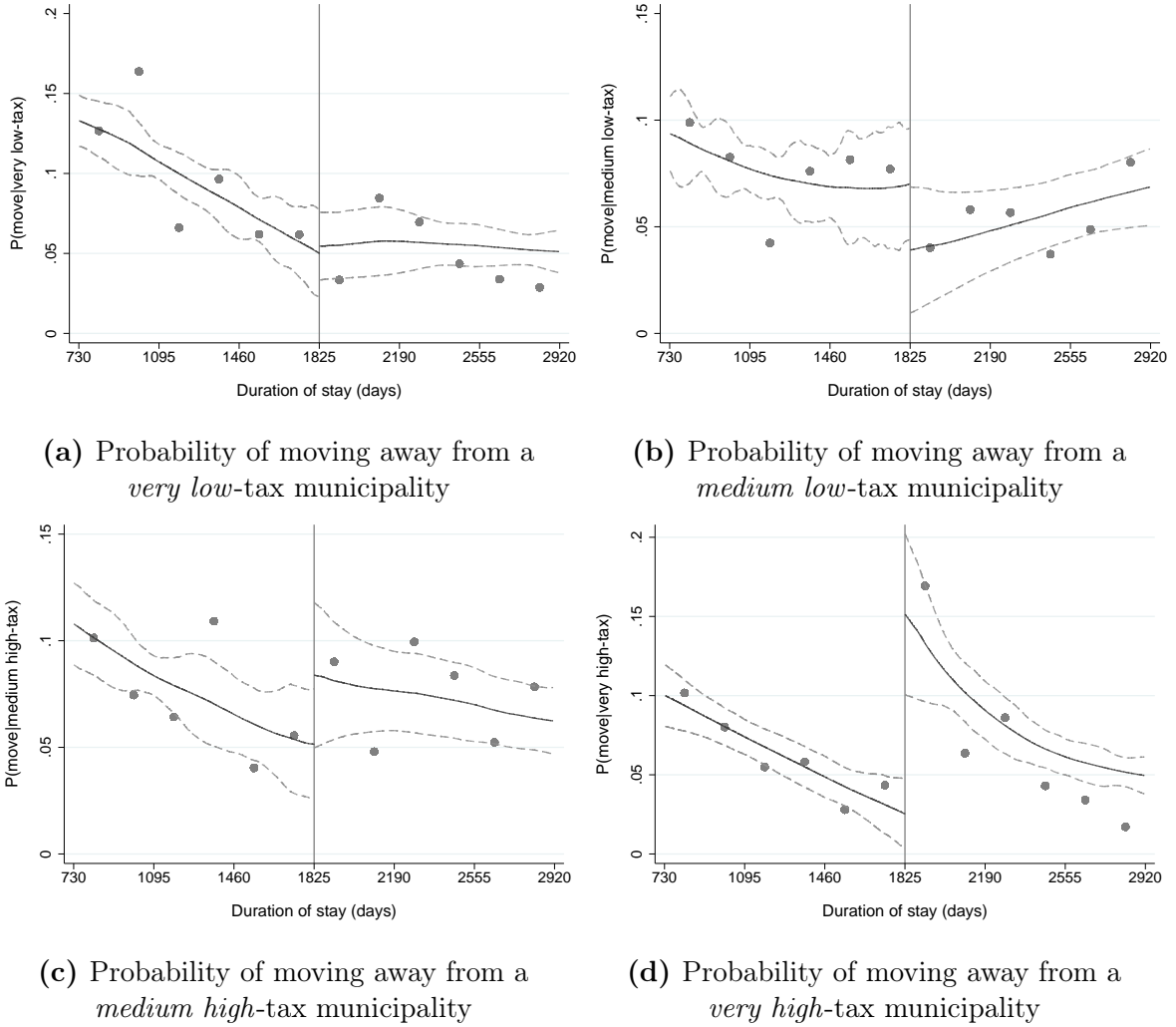
26

Table 4: Probability of moving away from a particular municipality for four groups of municipalities

		All		Low income		High income	
		non-parametric	parametric	non-parametric	parametric	non-parametric	parametric
		(1)	(2)	(3)	(4)	(5)	(7)
(a)	*Probability of moving away from a very low-tax municipality:*						
	Effect ΔP	-0.001	0.094	-0.036	0.234	0.015	0.023
		(0 .098)	(0.080)	(0.271)	(0.184)	(0.097)	(0.078)
	N left	2108	2108	787	787	1321	1321
	N right	1154	1154	431	431	723	723
(b)	*Probability of moving away from a medium low-tax municipality:*						
	Effect ΔP	0.012	0.072	0.133	0.192	-0.090	-0.033
		(0.074)	(0.069)	(0.115)	(0.126)	(0.097)	(0.072)
	N left	1953	1953	944	944	1009	1009
	N right	1061	1061	513	513	548	548
(c)	*Probability of moving away from a medium high-tax municipality:*						
	Effect ΔP	0.113	0.123	0.075	0.068	0.130	0.152*
		(0.119)	(0.076)	(0.233)	(0.141)	(0.138)	(0.089)
	N left	1781	1781	728	728	1053	1053
	N right	1065	1065	386	386	679	679
(d)	*Probability of moving away from a very high-tax municipality:*						
	Effect ΔP	0.401***	0.292**	0.194	0.190	0.517***	0.356**
		(0.116)	(0.119)	(0.199)	(0.136)	(0.145)	(0.178)
	N left	2146	2146	758	758	1388	1388
	N right	1188	1188	445	445	743	743

Notes: Fuzzy RDD estimates of the probability of moving given the individuals were residing in the respective municipality type for different income groups. Nonparametric means fuzzy RDD with local-linear regression using a triangular kernel, while parametric means a linear instrumental variable estimation. All estimates include a bandwidth of 1460 days (4years) and account for sampling weights. Standard errors in parentheses. Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

the estimates for individuals living in medium high-tax municipalities are also positive though smaller and not significant at the 5% level. For individuals in medium low-tax and very low-tax municipalities, the effects tend to be negative but not systematically and not significant even at the 10% level. Figure 7 visualizes these findings for high-income individuals.

In sum, we find a sharp increase in mobility of households that experience a strong increase in their income tax rates, i.e. high-income households in very high-tax municipalities and a weak increase for households that experience a mild increase in tax rates, low-income households in medium high-tax municipality. This is exactly what our theory predicts about residents of high-tax municipalities. However, we do not find strong effects for residents who experience decreases in their income tax rates, not even for those

(a) Probability of moving away from a
very low-tax municipality

(b) Probability of moving away from a
medium low-tax municipality

(c) Probability of moving away from a
medium high-tax municipality

(d) Probability of moving away from a
very high-tax municipality

Figure 7: Probability of moving given the individuals were residing in the respective
municipality type for the high-income experimental sample around the threshold value. These
graphs show local linear smooths of the probability of moving conditioning on the previous
municipality type for the high-income experimental sample using a bandwidth of 1460 days.
The light gray dashed lines indicate the 90% pointwise confidence bands and the gray dots raw
weighted averages of the dependent variable in half yearly bins.

individuals who experience the strongest decline in tax rates, i.e. high-income individuals
in the lowest-tax municipalities. There seems to be an asymmetric mobility response to
tax increases and decreases which is not in line with our theoretical prediction.

6.4 More effect heterogeneity

Our theoretical model would further predict that the mobility response is smaller for
individuals with higher moving costs. In this section, we test this hypothesis by comparing
individuals with and without children living in the household. Households with children
are arguably less spatially mobile because of the social costs involved in changing schools
and the social fabric in the neighborhood (see e.g. Mincer, 1978; Long, 1972).
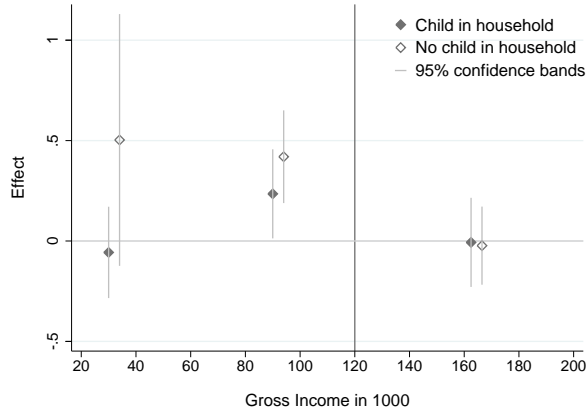
Figure 8: Probability of moving away from a high-tax municipality for individuals with and without children in the household and for three income groups. Estimation results can be found in Table C.3 in Appendix C.

Figure 8 summarizes the fuzzy RDD estimates reported in Table C.3 in Appendix C. We find that the effect is indeed stronger for treated individuals in households without children. For individuals with income above 120k Swiss francs, i.e. the placebo sample, there is no effect, either with or without children in the household.

6.5 Robustness

First, to check that our results are not driven by the chosen bandwidth of the non-parametric estimates, we report the main mobility estimates for all income groups and all municipality types for diverse bandwidths in Tables C.6, C.7, and C.8 in Appendix C. Of these, the last bandwidth is the optimal one, following the procedure of Calonico et al. (2014). We do not find our estimates to be particularly sensitive to bandwidth choice.

Second, to check whether the parametric fuzzy RDD estimates are not driven by functional form assumptions, we additionally control for second order polynomials of the forcing variable as well as for individual characteristics in Table C.9 in Appendix C. In general, we find that the parametric estimates tend to be lower. However, the main pattern persists and is rather similar in magnitude and significance.

6.6 Validation of RDD

This section discusses the validity of the fuzzy RDD as causal estimates by inspecting the main assumptions.

First, we address the validity of the *local randomization* around the duration threshold. Every individual in our sample will automatically cross the threshold after five years of stay if she/he intends to stay in Switzerland. The only possibility of manipulation would be to emigrate from Switzerland shortly before crossing the threshold and then to immigrate into Switzerland again. We consider this manipulation as rather unrealistic because the individual would have to wait another 5 years before being granted permanent

29
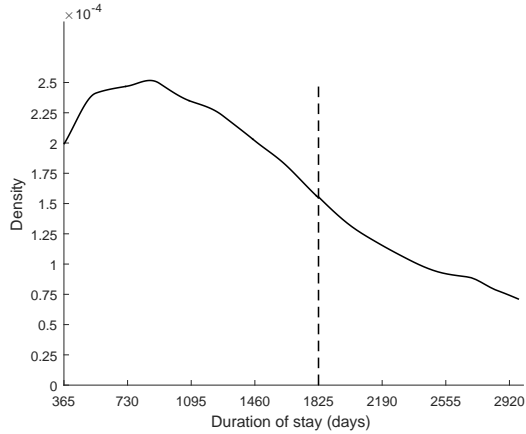
(a) Distribution of duration of stay in Switzerland
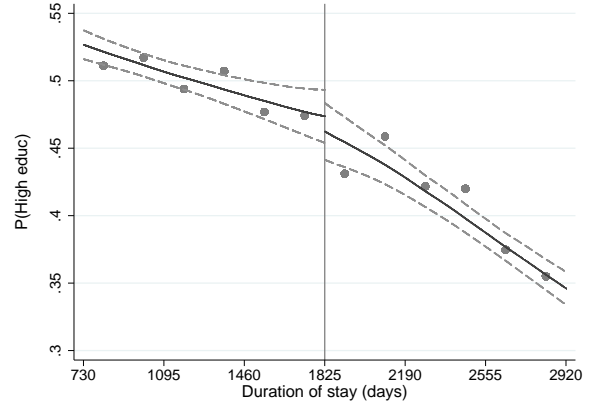
(b) Fraction of individuals with higher education

Figure 9: Validation of local randomization around the threshold. Figure 9a shows local likelihood density estimates of the duration of stay using $bandwidth = 360$, accounting for survey weights and, separately, from both sides of the duration threshold of 1825 days, or 5 years. The corresponding density discontinuity estimates can be found in row (b) of Table C.14 in Appendix C. Figure 9b presents local linear smooths of the probability of individuals around the threshold being highly educated using the whole sample and a bandwidth of 1460 days. The light gray dashed lines indicate the 90% pointwise confidence bands and the gray dots raw weighted averages of the dependent variable in half yearly bins. The corresponding discontinuity estimate can be found in column (12) of Table C.5 in Appendix C.

residence status (permit C) which comes with much more important benefits than the tax treatment, as discussed above. We nevertheless validate the local randomization assumption empirically by testing for a discontinuity in the distribution of the duration of stay at the threshold value. Any discontinuity at the threshold value could indicate systematic sorting and a violation of the local randomization assumption (see, e.g, McCrary, 2008).[30] Figure 9(a) visually confirms that there is no discontinuity in the duration density at the threshold (detailed results are reported in Table C.14 in Appendix C). Individuals could still be systematically sorted around the threshold. We therefore check whether individuals' observable characteristics are continuous around the duration threshold. Table C.5 and Figure C.1 in Appendix C show sharp RDD estimates for age, gender, household size, education, and contract type. Figure 9(b) visualizes this test for education. We do not find any substantial and significant discontinuities in the tested covariates. We conclude that the treatment group just after the threshold is comparable to the control group just before the threshold, when it comes to observable characteristics.

Second, we check the validity of the continuity and excludability assumptions. In our application this means checking whether there are any other discontinuities after 5 years

---

[30]Our estimates for the discontinuity in the density are based on the empirical likelihood approach derived in Otsu et al. (2013). In order to adjust the estimates for the sampling weights in our data, we multiply the kernel weights by the sampling weights in the summation part of the FOC. For the graphic evidence we apply the local likelihood implementation (locfit) in the Chronux software package for Matlab, including the weighting (Bokil et al., 2010).
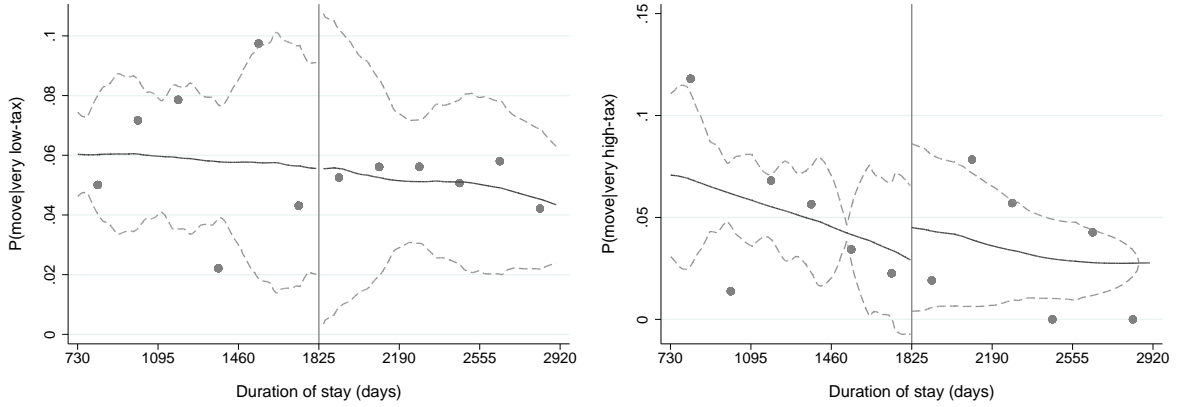
(a) Probability of moving away from a
very low-tax municipality

(b) Probability of moving away from a
very high-tax municipality

Figure 10: Probability of moving given the individual was residing in the respective municipality type for the placebo sample around the threshold value. These graphs show local linear smooths of the probability of moving conditioning on the previous municipality type for the high-income experimental sample using a bandwidth of 1460 days (4 years). The light gray dashed lines indicate the 90% pointwise confidence bands and the gray dots raw weighted averages of the dependent variable in half yearly bins.

of stay, and whether the effect from being granted permanent resident status (permit C) is really due to changes in tax rates and not confounded with other changes that take place at the same time. There may by systematic changes in the preferences of individuals after 5 years in the country or when they receive permanent resident status in the country. For example, an individual may decide to start a family and have children or buy a home. We check this by performing a fuzzy RDD in a placebo sample of always treated individuals, i.e. foreigners with incomes above 126,000 Swiss francs but otherwise the same sample restrictions as our experimental sample of individuals with incomes below 114,000 Swiss francs. Foreigners with incomes above 126,000 Swiss francs are eligible for permanent resident status (permit C) in the exact same way as individuals with incomes below 114,000 Swiss francs. Hence they experience exactly the same changes at the 5-year threshold except for the change in the tax treatment because they are already in the ordinary tax regime. Fuzzy RDD estimates using foreigners with income above 126,000 Swiss francs can therefore serve as a placebo test. Table 5 shows the results of this placebo test and Figure 10 visualizes the reduced form relationship for the very high and very low-tax municipalities. We do not find any significant discontinuity in the placebo group, which suggests that there is no other discontinuity after 5 years that explains the estimated effect in the treated group, nor can this effect be confounded with other benefits of the permanent residence status.

Third, we argue that the monotonicity assumption is evidently given in our setting. The incentive to hold permit C increases with the time spent in the country. In our view it is not conceivable that the crossing of the threshold reduces these incentives for a particular group. Therefore, the incentives to accept the treatment seem to work in one

31

Table 5: Probability of moving away from a particular municipality for 4 municipality groups

	Probability of moving from a very high-tax		Probability of moving from a medium high-tax		Probability of moving from a medium low-tax		Probability of moving from a very low-tax	
	non-parametric	parametric	non-parametric	parametric	non-parametric	parametric	non-parametric	parametric
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect ΔP	0.052	0.034	-0.079	-0.013	-0.023	-0.005	-0.000	0.014
	(0.147)	(0.091)	(0.106)	(0.078)	(0.153)	(0.101)	(0.124)	(0.076)
N left	303	303	170	170	216	216	373	373
N right	184	184	132	132	139	139	212	212

Notes: Fuzzy RDD estimates of the probability of moving given individuals were residing in the corresponding municipality type for the quasi-placebo group, i.e. individuals with an income above 126k. Nonparametric means fuzzy RDD with local-linear regression using a triangular kernel, while parametric means a linear instrumental variable estimation. All estimates include a bandwidth of 1460 days (4years) and account for sampling weights. Standard errors in parentheses. Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

direction. Consistently, we observe in Figure 3 that the probability of holding permit C increases, the longer individuals stay in Switzerland.

Finally, we check whether the estimated discontinuities at the duration threshold of five years are not some artefact of our estimation procedure. We therefore repeat the main estimates with four types of municipalities for placebo-thresholds at four and six years as a falsification test. Results for the high income group are reported in Tables C.10 and C.11 for the mobility estimates and in Tables C.12 and C.13 for the residential choice estimates. We find no significant jumps in the probability at these placebo-thresholds.

7 Discussion

This section addresses the interpretation of the magnitude of the identified effects and further challenges for the causal interpretation of our estimated effects.

7.1 Implied migration elasticity

Our quasi-experimental institutional setting allows us to estimate migration elasticities from the causal treatment effects estimated in section 6 as described in Section 4.3.

Table C.4 in the Appendix reports elasticity estimates based on two different estimated treatment effects: Based on the estimated effect on the probability of residing in a very *high*-tax municipality, we estimate a value of 22.3 for the implied average migration elasticity. This migration elasticity means that an increase in the local net-of-tax rate by 1% from, e.g., 86 to 87 percentage points leads to an increase in the foreign population by 22.3%. This point estimate is, however, very imprecise with a 95% confidence interval ranging from 1.98 to 45.86. Alternatively, based on the estimated effect on the probability

of moving away from a very *high*-tax municipality, we estimate an elasticity of 26.9 with a 95% confidence interval ranging from 7.2 to 65.9. The elasticity estimates based on location choice and mobility decisions in very *low*-tax municipalities are not significantly different from 0 as are the underlying estimated treatment effects.

Our point estimates of the migration elasticity are larger by an order of magnitude compared to the estimates in the literature. Elasticities for the mobility of domestic individuals within countries range between 0.03 and 0.4 (see Kleven et al., 2013; Akcigit et al., 2016; Moretti and Wilson, 2017; Agrawal and Foremny, 2017) while elasticities for the mobility of foreign individuals within countries or for international mobility tend to be higher with a range between 1 and 2 (see, e.g., Kleven et al., 2013, 2014; Akcigit et al., 2016). Our large point estimates are very imprecisely estimated and therefore not very informative beyond being significantly positive. We still see plausible arguments why migration elasticities in our context may be higher than the ones in the previous literature: First, we are estimating mobility reactions between municipalities. Moving costs are therefore arguably lower and migration responses likely stronger than across states or countries. Second, tax rate differentials across Swiss municipalities have been remarkably stable over decades. The shift from special to ordinary taxation, implies therefore a known *permanent* change in the tax rates which has likely stronger effects than temporary or volatile changes. Third, our treatment is a very salient change in the tax regime because previously withheld income taxes are now invoiced after filing taxes. Fourth, we are investigating the mobility reactions of recently immigrated foreigners which are likely to have fewer social ties and hence lower migration costs than the general population. It is also important to keep in mind that our RDD estimates are inherently local and apply to foreigners in Switzerland after 5 years of stay and thus might be strongly affected by the investigated institutional setting. Consequently, any extrapolation has to be taken with caution.

7.2 Income response at the income threshold

As described in Section 2, there are two thresholds that determine whether foreign workers in Switzerland are subject to the ordinary or the special tax regime: a duration threshold after 5 years of stay and an income threshold at 120,000 Swiss francs. We show in this subsection that the income threshold creates incentives that lead workers to strategically adjust earnings in a narrow band around the threshold. Individuals with observed incomes just below or just above the threshold are therefore potentially individuals who systematically adapted their income and might be systematically different from non-adapting individuals. To avoid a bias from sample selection in our fuzzy RDD estimates, we exclude a range of ±6000 of incomes around the threshold of 120,000 Swiss francs from both our experiment sample and our placebo sample.

Foreigners with a given income below the income threshold have an incentive to choose a high-tax municipalities during the first five years of stay because they are not subject
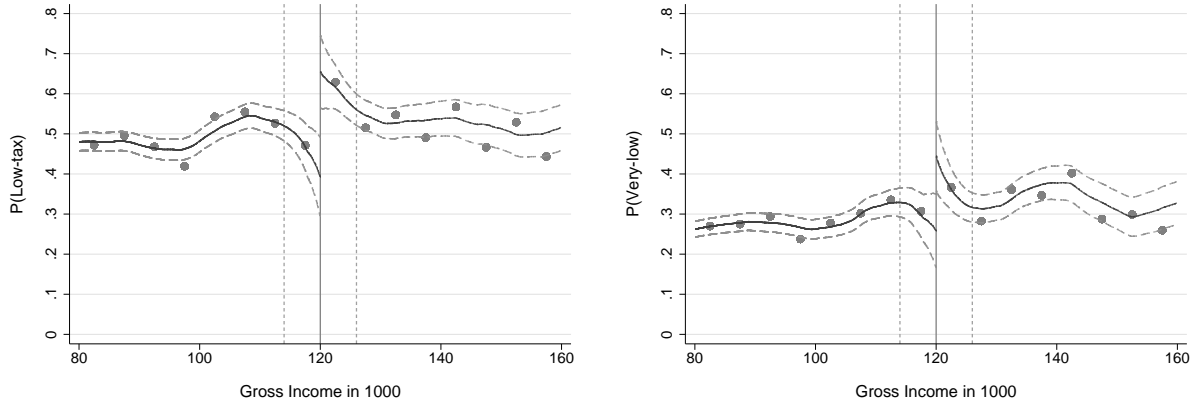
(a) Probability of residing in a low-tax municipality

(b) Probability of residing in a very low-tax municipality

Figure 11: Probability that an individual resides in a low-tax municipality. Foreigners not holding the permanent residence status (permit C) during their first five years in Switzerland. Local linear regression separately at both sides of the income threshold using a bandwidth $10k$ and taking sampling weights into account. The light gray dashed lines indicate the 90% pointwise confidence bands and the gray dots raw weighted averages of the dependent variable in bins of $5k$.

to the high tax rate but benefit from the lower housing prices in these municipalities. A sharp RDD around the income threshold should therefore reveal a positive jump in the probability of living in a low-tax municipality. We inspect this visually in Figure 11(a), where we indeed find a positive discontinuity, reported in the estimates in Table C.16 in Appendix C. The observed discontinuity is, however, not the result of a level-shift but of a sharp decline in the probability from the left of the threshold and a sharp increase from the right. Outside of a narrow income band of about $\pm 6k$ around the threshold, the probability of choosing a low-tax municipality seems unaffected. This impression is confirmed when a range of $\pm 6k$ is excluded around the threshold in Table C.16 in Appendix C. After the exclusion the probabilities on both sides do not systematically differ anymore. This conclusion also holds when analyzing the probability of residing in a very low-tax municipality, where the incentives are even stronger. The results can be found in Figure 11(b) and Table C.16 in Appendix C. The observed pattern is consistent with systematic income adaption at the income threshold: foreigners in low-tax municipalities seek to push their income above the threshold to profit from the lower ordinary tax rates; foreigners in high-tax municipalities seek to keep their income below the threshold to benefit from the lower special tax rate. This is also what we find when testing for a discontinuity in the income distribution at the income threshold (see Figure C.2): the income distributions feature excess mass at the preferential side, i.e., just below the income threshold in high-tax municipalities and just above it in low-tax municipalities.

7.3 Anticipated moving

Foreigners in our experimental sample could in principle already move before the 5-year threshold in anticipation of the change in the tax regime after 5 years which is known to the informed migrant. In the presence of anticipation, our estimated effect at the threshold would be biased. In this section, we argue theoretically that our estimated effect would be biased towards zero.

In the absence of moving costs, the location decision at any point in time, only depends on the tax rates of that point in time and individuals do not have an incentive to anticipate moving to a municipality which will be optimal in later periods. In the presence of moving costs, however, future tax rates may affect today's location decision. For simplicity, consider a deterministic version of our theoretical model in Section 3 with a pre-treatment period $t = 1$ and a post-treatment period $t = 2$ and two municipalities $m = \ell, h$, where municipality ℓ has low ordinary tax rates and municipality h high ordinary tax rates. Consider now an individual who would prefer living in the high-tax municipality in period 1, i.e $V_{h1} > V_{\ell 1}$, and living in a low-tax municipality in period 2, i.e $V_{h2} < V_{\ell 2}$. Choosing the low-tax municipality in period 1 would incur a static loss of $\Delta V_1 = V_{h1} - V_{\ell 1}$ and choosing the low-tax municipality in period 2 would realise a static gain of $\Delta V_2 = V_{\ell 2} - V_{h2}$. Considering moving costs c, this individual would already initially choose a low-tax municipality and stay there if the loss ΔV_1 from living in a sub-optimal location in period 1 is smaller than both the discounted moving cost $\rho \cdot c$ and the discounted gain $\rho \cdot \Delta V_2$ from living in the optimally location in period 2, where ρ is the discount factor. A forward looking individual is therefore more likely to anticipate if moving costs are high and if the gain in period 2 is high.

In our results on location choices visualized in Figure 5, initial anticipation would show up as a parallel upward shift to the left of the threshold in Figure (a) and a parallel downward shift to the left of the threshold in Figure (b). The existence of a clear jump in both figures rules out full anticipation in particular in Figure (b) where the jump is statistically significant. However, *partial* anticipation could be present and would bias our estimated effects towards zero.

An individual could in principle also anticipate at some later point during the first five years in Switzerland. However, such anticipation is theoretically not beneficial because it incurs both a static loss and moving costs. In this case, it would be best to stay in the statically optimal location for the whole pre-treatment period and move just after 5 years. Later anticipation would show up in Figure 5 as an increase of the slope to the left of the threshold in Figure(a) and a decrease of the slope in Figure (b). Later *partial* anticipation would also bias our estimated effects towards zero.

Note that initial anticipation requires foreign workers to understand the geography of municipal income tax differentials already when they arrive in Switzerland. As most foreigners have never experienced such decentralized taxation, they likely learn about the Swiss tax regime only later, at the latest with their first tax invoice when switching to

ordinary taxation after five years in the country. Such information frictions also limit the scope for a bias from anticipation.

7.4 Initial selection into municipalities

Our fuzzy RDD results at the duration threshold on mobility in Section 6.3 show that the effect of the tax treatment is different for foreigners in low-tax municipalities than for those in high-tax municipalities. We explain the opposing effect with the opposing incentives in the two types of municipalities. However, the observed difference could in principle also stem from systematic initial selection into low-tax municipalities. In combination with heterogeneous effects, such initial selection could also explain the estimated differences between the effects in low-tax vs. high-tax municipalities.

First, we directly test for initial selection on observables by estimating a linear probability model for the probability of residing in a low-tax municipality or in a very low-tax municipality, respectively, as a function of demographic variables (age, marital status, children), income, educational attainment, type of work contract, occupational characteristics, and industry sector. We restrict this estimation to foreign employees with incomes below 114,000 Swiss francs, i.e. foreign workers that are subject to the special foreigners tax regime below the 5-year duration threshold and to the ordinary tax regime after the duration threshold.[31] The results are reported in Table C.15 in the Appendix. Columns (1) and (2) show the estimated models for foreign employees that stayed in Switzerland for less than 2 years. The initial location choice is not significantly related to education, occupation or industry sector and only weakly related to demographic variables. Columns (3) and (4) estimate the same models for foreign employees that stayed in Switzerland for more than 8 years. This later location choice is significantly related to occupation, industry sector and education. The demographic controls also have more predictive power in the later group. A test on the joint hypothesis that all demographic coefficients are zero yields an F-statistic of 1.36 (p=0.088) and 1.44 (p=0.0547) in columns (1) and (2), respectively, and an F-statistic of 5.46 (p=0.000) and 2.16 (p=0.0002) in columns (3) and (4). Hence, foreign employees do only weakly sort into municipalities based on the level of ordinary local tax rates when they arrive in Switzerland and when they are not subject to the ordinary tax rates under the special tax regime. However, foreign employees systematically sort into municipalities based on the ordinary local tax rates after they passed the duration threshold and when they are subject to the ordinary tax rates. We conclude that we cannot find evidence of strong systematic selection into a particular municipality type that could explain the difference between the effects in low-tax and high-tax municipalities.

Second, our estimated effects allow us to test against the null hypothesis that tax rate changes have no effect on the mobility of individuals even when initial sorting were

---

[31]We also restrict the group to individuals with incomes above 10,000 Swiss francs because local tax rates are virtually zero for the ones below.

present. Under the null hypothesis, we should not see any systematic effect on mobility at the threshold neither in low-tax nor in high-tax municipalities. Even for a selective local population, the only difference before and after the duration threshold is the change in local tax rates. In the presence of initial sorting, our estimated causal effects would, however, need to be interpreted as *local effects* for the respective populations.

8 Conclusion

We study residential choice and residential mobility responses to local income taxes in a quasi-experimental setting generated by a special foreigners' tax regime in Switzerland.

Ordinary income tax rates in Switzerland differ across the 26 cantons and the roughly 2500 municipalities. Foreigners whose yearly gross income is below 120,000 Swiss francs (around 130,000 US-Dollars in 2015) are, however, subject to a special foreigners' tax regime until they obtain a permanent residence permit, for which they can apply after five years of stay in Switzerland. Taxpayers in the special tax regime face an income tax rate that does not differ across the municipalities within a canton. This institutional setting generates a deterministic duration threshold at 5 years of stay in the country when foreigners in the special tax regime with identical tax rates across space become suddenly eligible for the ordinary tax scheme with tax rates that differ across space. We exploit this threshold by applying a fuzzy regression discontinuity design (RDD) to estimate the causal effect from a simultaneous change of the local income tax rates in all municipalities.

We document strong responses in both location choices and moving decisions for high-income earners, but no effects for low-income earners. High income foreigners receiving the permanent residence permit at the 5-year threshold are systematically less likely to reside in one of the municipalities with the highest local tax rates and are systematically more likely to move away from a very high-tax municipality when compared to their control group just below the threshold value. Furthermore, we find that this effect is strongest for households without children, which arguably have lower moving costs. Different from the theoretical prediction, however, we do not find a significant effect on the moving probability for individuals residing in one of the lowest-tax municipalities suggesting an asymmetric response to tax increases and tax decreases. Our fuzzy RDD estimates are robust to a series of robustness tests using different bandwidths and functional forms, pass placebo tests with counterfactual threshold values and with an always treated population.

Our theoretical framework also allows us to estimate the migration elasticity with respect to the net-of-tax rate from the fuzzy RDD estimates. The resulting elasticities are considerably higher than the values found in the literature so far. However, our estimates are very imprecise and therefore not very informative beyond being significantly positive.

Acknowledgement

Data Statement

The confidential individual-level data (SESAM) were obtained under contract Nr. 130249 from the Swiss Statistical Office. Access is granted for scientific research projects after review of a detailed application.[32]

References

Agrawal, D. R. and D. Foremny (2017). Relocation of the Rich: Migration in Response to Top Tax Rate Changes from Spanish Reforms. *Mimeo, Department of Economics, University of Kentucky*.

Akcigit, U., S. Baslandze, and S. Stantcheva (2016). Taxation and the International Mobility of Inventors. *American Economic Review 106*(10), 2930–2981.

Angrist, J. D. and G. W. Imbens (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–475.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton NJ: Princeton University Press.

BFS (2004). Die schweizerische Arbeitskräfteerhegung (SAKE). Konzepte-Methodische Grundlagen-Praktische Ausfürung. *Bundesamt für Statistik (BFS). Neuchâtel*.

BFS (2012). SAKE-Daten Benutzerhandbuch. *Bundesamt für Statistik (BFS). Neuchâtel*.

---

[32]See https://www.bfs.admin.ch/bfs/de/home/statistiken/arbeit-erwerb/erhebungen/sesam.html

Bokil, H., Andrews Peter, J. Kulkarni, S. Mehta, and P. P. Mitra (2010). Chronux: A Platform for Analyzing Neural Signals. *Journal of Neuroscience Methods 192*(1), 146–151.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica 82*(6), 2295–2326.

Chetty, R., J. N. Friedman, T. V. Olsen, and L. Pistaferri (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *The Quarterly Journal of Economics 126*(2), 749–804.

Chetty, R., J. N. Friedman, and E. Saez (2013). Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings. *American Economic Review 103*(7), 2683–2721.

Epple, D. and T. Romer (1991). Mobility and Redistribution. *Journal of Political Economy 99*(4), 828–858.

Esteller, A., A. Piolatto, and M. D. Rablen (2017). Taxing High-Income Earners: Tax Avoidance and Mobility. *IEB Working Paper 2017/06*.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications*, Volume 66 of *Monographs on Statistics and Applied Probability*. London: Chapman and Hall.

Feld, L. P. (2000). Tax Competition and Income Redistribution: An Empirical Analysis for Switzerland. *Public Choice 105*(1), 125–164.

Feldstein, M. (1995). The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act. *Journal of Political Economy 103*(3), 551.

Feldstein, M. (1999). Tax Avoidance and the Deadweight Loss of the Income Tax. *Review of Economics and Statistics 81*(4), 674–680.

Feldstein, M. and M. V. Wrobel (1998). Can State Taxes Redistribute Income? *Journal of Public Economics 68*(3), 369–396.

Fernandez, R. and R. Rogerson (1996). Income Distribution, Communities, and the Quality of Public Education. *Quarterly Journal of Economics 111*(1), 135–164.

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

Harms, T. and P. Duchesne (2010). On Kernel Nonparametric Regression Designed for Complex Survey Data. *Metrika 72*(1), 111–138.

Hennessy, C. and I. A. Strebulaev (2015). Natural Experiment Policy Evaluation: A Critique. NBER Working Paper No. 20978.

Hodler, R. and K. Schmidheiny (2006). How Fiscal Decentralization Flattens Progressive Taxes. *FinanzArchiv: Public Finance Analysis 62*(2), 281–304.

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

Kleven, H. J., C. Landais, and E. Saez (2013). Taxation and International Migration of Superstars: Evidence from the European Football Market. *American Economic Review 103*(5), 1892–1924.

Kleven, H. J., C. Landais, E. Saez, and E. Schultz (2014). Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark. *Quarterly Journal of Economics 129*(1), 333–378.

Kleven, H. J. and M. Waseem (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *Quarterly Journal of Economics 128*(2), 669–723.

Lee, D. and T. Lemieux (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature 48(2)*, 281–355.

Long, L. H. (1972). The Influence of Number and Ages of Children on Residential Mobility. *Demography 9*(3), 371–382.

McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics 142*(2), 698–714.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In E. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. Academic Press, New York.

Mincer, J. (1978). Family Migration Decisions. *Journal of Political Economy 86*(5), 749–773.

Mirelees, J. A. (1971). An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies 38*(2), 175–208.

Mirelees, J. A. (1982). Migration and Optimal Income Taxes. *Journal of Public Economics 18*(3), 319–341.

Moretti, E. and D. Wilson (2017). The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists. *American Economic Review 107*(7), 1858–1903.

Otsu, T., K.-l. Xu, and Y. Matsushita (2013). Estimation and Inference of Discontinuity in Density. *Journal of Business & Economic Statistics 31*(4), 507–524.

Parchet, R. (2014). Are Local Tax Rates Strategic Complements or Strategic Substitutes? *IdEP Economic Papers, University of Lugano* (2014/07).

Porter, J. (2003). Estimation in the Regression Discontinuity Model. *Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison*.

Roller, M. and K. Schmidheiny (2016). Effective Tax Rates and Effective Progressivity in a Fiscally DecentralizedCountry. CEPR Discussion Paper No. DP 11152, London.

Saez, E. (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Saez, E., J. Slemrod, and S. H. Giertz (2012). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature 50*(1), 3–50.

Schmidheiny, K. (2006). Income Segregation and Local Progressive Taxation: Empirical Evidence from Switzerland. *Journal of Public Economics 90*(3), 429–458.

Schmidheiny, K. and M. Slotwinski (2015). Behavioral Responses to Local Tax Rates: Quasi-Experimental Evidence from a Foreigners' Tax Scheme in Switzerland. CESifo Working Paper No. 5518, Munich.

Spencer, B. and H. Selin (2014). Bunching and Non-Bunching at Kink Points of the Swedish Tax Schedule. *Journal of Public Economics 109*, 36–49.

Wilson, J. D. (1980). The Effect of Potential Emigration on the Optimal Linear Income Tax. *Journal of Public Economics 14*(3), 339–353.

Young, C. and C. Varner (2011). Millionaire Migration and State Taxation of Top Incomes: Evidence from a Natural Experiment. *National Tax Journal 64*(2), 255–284.

A Appendix Theory

This appendix derives the main equations of our theoretical model and generalizes the results to groups of municipalities. We will ignore the individual identifier i for ease of exposition in the appendix.

We partition the ordered municipalities into G groups such that all municipalities in group g have (weakly) lower tax rates than all municipalities in group $g + 1$. The set $\mathbb{G}_g$ contains all municipalities that belong to group g. Consequently,

$$\overline{\Delta T}_{\mathbb{G}_1 2} \geq ... \geq \overline{\Delta T}_{\mathbb{G}_g 2} \geq ... \geq \overline{\Delta T}_{\mathbb{G}_G 2} \tag{20}$$

where $\overline{\Delta T}_{\mathbb{G}_g 2}$ is any (weighted) average over all net-of-tax rate changes ΔT_{m2} of the municipalities within group g.

The effect of a change in the net-of-tax rate T_{m2} of a particular municipality m on the conditional probability of choosing this destination municipality m is positive

$$\frac{\partial P_{m2|s1}}{\partial T_{m2}} = \beta \cdot P_{m2|s1} \cdot (1 - P_{m2|s1}) > 0 \tag{21}$$

while the effect of a change in the net-of-tax rate T_{k2} in another municipality $k \neq m$ on choosing municipality m is negative

$$\frac{\partial P_{m2|s1}}{\partial T_{k2}} = -\beta \cdot P_{m2|s1} \cdot P_{k2|s1} < 0. \tag{22}$$

A.1 Derivation of Eq. (10)

The unconditional probability of choosing a particular destination municipality m is

$$P_{m2} = \sum_{s=1}^{M}(P_{m2|s1} \cdot P_{s1}). \tag{23}$$

The effect of an increase in the net-of-tax rate T_{m2} of a particular municipality m on the probability of choosing this municipality m is positive

$$\frac{\partial P_{m2}}{\partial T_{m2}} = \sum_{s=1}^{M}\left(\frac{\partial P_{m2|s1}}{\partial T_{m2}}P_{s1}\right) = \beta \sum_{s=1}^{M}\left[P_{m2|s1}(1 - P_{m2|s1})P_{s1}\right] > 0 \tag{24}$$

and negative when the net-of-tax rate in another municipality $k \neq m$ is raised.

$$\frac{\partial P_{m2}}{\partial T_{k2}} = \sum_{s=1}^{M}\left(\frac{\partial P_{m2|s1}}{\partial T_{k2}}P_{s1}\right) = -\beta \sum_{s=1}^{M}\left(P_{m2|s1}P_{k2|s1}P_{s1}\right) < 0. \tag{25}$$

The effect of the simultaneous change of all tax rates τ_{k2} and hence all net-of-tax rates in logs, $T_{k2} = \log(1 - \tau_{k2})$, is given by the total derivative

$$
\begin{aligned}
\mathrm{d}P_{m2} &= \sum_{k=1}^{M} \frac{\partial P_{m2}}{\partial T_{k2}} \mathrm{d}T_{k2} = \frac{\partial P_{m2}}{\partial T_{m2}} \mathrm{d}T_{m2} + \sum_{k\neq m} \frac{\partial P_{m2}}{\partial T_{k2}} \mathrm{d}T_{k2} \\
&= \beta \sum_{s=1}^{M} \left[P_{m2|s1}(1\text{-}P_{m2|s1})P_{s1}\mathrm{d}T_{m2} \right] - \beta \sum_{k\neq m}\sum_{s=1}^{M} \left[P_{m2|s1}P_{k2|s1}P_{s1}\mathrm{d}T_{k2} \right] \\
&= \beta \sum_{s=1}^{M} \left[P_{s1}P_{m2|s1}\left((1 - P_{m2|s1})\mathrm{d}T_{m2} - \sum_{k\neq m} P_{k2|s1}\mathrm{d}T_{k2} \right) \right] \\
&= \beta \sum_{s=1}^{M} \left[P_{s1}P_{m2|s1}\left(\mathrm{d}T_{m2} - P_{m2|s1}\mathrm{d}T_{m2} - \sum_{k\neq m} P_{k2|s1}\mathrm{d}T_{k2} \right) \right] \\
&= \beta \sum_{s=1}^{M} \left[P_{s1}P_{m2|s1}\left(\mathrm{d}T_{m2} - \sum_{k=1}^{M} P_{k2|s1}\mathrm{d}T_{k2} \right) \right] \\
&= \beta \sum_{s=1}^{M} \left[P_{s1}P_{m2|s1}\left(\mathrm{d}T_{m2} - \overline{\mathrm{d}T}_{2|s1} \right) \right] \quad (26)
\end{aligned}
$$

where $\overline{\mathrm{d}T}_{2|s1} = \sum_{k=1}^{M} P_{k2|s1}\mathrm{d}T_{k2}$ is the average of all net-of-tax rate changes weighted by the choice probabilities of a household in the municipality of origin s.

A.2   Proof of Hypothesis 1

(a) $(\mathrm{d}T_{m2} - \overline{T}_{2|s1}) > 0$ for the lowest-tax municipality $m = 1$ as any weighted average $\overline{T}_{2|s1}$ with strictly positive weights $P_{k2|s1}$ is smaller than the change in the net-of-tax rate $\mathrm{d}T_{m2}$ of the lowest-tax municipality. The weighted sum over all municipalities of origin s is also positive as $P_{s1} > 0$ and $P_{m2|s1} > 0$. Hence, $\mathrm{d}P_{12} > 0$ as $\beta > 0$. (b) Analogously. $\square$

A.3   Generalization of Hypothesis 1

The probability of choosing any destination municipality of group g is

$$
P_{\mathbb{G}_g 2} = \sum_{m \in \mathbb{G}_g} P_{m2} = \sum_{m \in \mathbb{G}_g}\sum_{s=1}^{M} (P_{m2|s1}P_{s1}). \quad (27)
$$

The effect of the simultaneous change of all net-of-tax rates, $T_{k2} = \log(1 - \tau_{k2})$, is the sum of the effects on the individual municipalities in the group

$$
\begin{aligned}
\mathrm{d}P_{\mathbb{G}_g 2} &= \sum_{k=1}^{M} \frac{\partial P_{\mathbb{G}_g 2}}{\partial T_{k2}} \mathrm{d}T_{k2} = \sum_{k=1}^{M} \sum_{m \in \mathbb{G}_g} \frac{\partial P_{m2}}{\partial T_{k2}} \mathrm{d}T_{k2} = \sum_{m \in \mathbb{G}_g} \mathrm{d}P_{m2} \\
&= \sum_{m \in \mathbb{G}_g} \left[\beta \sum_{s=1}^{M} \left[P_{s1} P_{m2|s1} \left(\mathrm{d}T_{m2} - \overline{\mathrm{d}T}_{2|s1} \right) \right] \right] \\
&= \beta \sum_{s=1}^{M} \left[P_{s1} \sum_{m \in \mathbb{G}_g} P_{m2|s1} \left(\mathrm{d}T_{m2} - \overline{\mathrm{d}T}_{2|s1} \right) \right] \\
&= \beta \sum_{s=1}^{M} \left[P_{s1} P_{\mathbb{G}_g 2|s1} \sum_{m \in \mathbb{G}_g} \frac{P_{m2|s1}}{P_{\mathbb{G}_g 2|s1}} \left(\mathrm{d}T_{m2} - \overline{\mathrm{d}T}_{2|s1} \right) \right] \\
&= \beta \sum_{s=1}^{M} \left[P_{s1} P_{\mathbb{G}_g 2|s1} \left(\overline{\mathrm{d}T}_{\mathbb{G}_g 2|s1} - \overline{\mathrm{d}T}_{2|s1} \right) \right] \quad (28)
\end{aligned}
$$

where $\overline{\mathrm{d}T}_{\mathbb{G}_g 2|s1} = \sum_{m \in \mathbb{G}_g} \frac{P_{m2|s1}}{P_{\mathbb{G}_g 2|s1}} \mathrm{d}T_{m2}$ is the weighted average over all net-of-tax rate changes in group g, $P_{\mathbb{G}_g 2|s1}$ is the probability of choosing a destination municipality in group g, and $\overline{\mathrm{d}T}_{2|s1} = \sum_{m=1}^{M} P_{m2|s1} \mathrm{d}T_{m2}$ as in Eq.(26).

Hypothesis A.1 (Effect on choosing a group of municipalities).
The effect of a simultaneous change of all tax rates on the probability of choosing a group of municipalities is

a) *strictly positive for the lowest-tax group of municipalities, i.e.*

$$\mathrm{d}P_{\mathbb{G}_1 2} > 0,$$

b) *and strictly negative for the highest-tax group of municipalities, i.e.*

$$\mathrm{d}P_{\mathbb{G}_G 2} < 0.$$

Proof: (a) $\left(\overline{\mathrm{d}T}_{\mathbb{G}_1 2} - \overline{\mathrm{d}T}_{2|s1} \right) = \sum_{m \in \mathbb{G}_1} \frac{P_{m2|s1}}{P_{\mathbb{G}_g 2|s1}} \mathrm{d}T_{m2} - \sum_{m=1}^{M} P_{m2|s1} \mathrm{d}T_{m2} > 0$ for the lowest-tax group of municipalities $g = 1$ as the average $\overline{\mathrm{d}T}_{\mathbb{G}_g}$ of all changes in the lowest-tax group is strictly bigger than the average of all changes $\overline{\mathrm{d}T}_{2|s1}$ because both averages are weighted by the same relative weights $P_{m2|s1}$. The weighted sum over all municipalities of origin m is also positive as $P_{s1} > 0$ and $P_{\mathbb{G}_g 2|s1} > 0$. Hence, $\mathrm{d}P_{\mathbb{G}_1 2} > 0$ as $\beta > 0$. (b) Analogously. $\square$

A.4 Derivation of Eq. (12)

The probability of moving out of municipality m strictly decreases with the net-of-tax rate T_{m2} in this municipality m

$$
\frac{\partial P_{move|m1}}{\partial T_{m2}} = -\frac{\partial P_{m2|m1}}{\partial T_{m2}} = -\beta P_{m2|m1} \cdot (1 - P_{m2|m1}) < 0 \quad (29)
$$

and increases with the net-of-tax rate of any other municipality $s \neq m$

$$\frac{\partial P_{move|m1}}{\partial T_{s2}} = \beta P_{m2|m1} P_{s2|m1} > 0. \tag{30}$$

The effect of the simultaneous change of all tax rates τ_{m2}, and hence all net-of-tax rates (in logs), $T_{m2} = \log(1 - \tau_{m2})$, is given by the total derivative

$$
\begin{aligned}
\mathrm{d}P_{move|m1} &= \sum_{s=1}^{M} \frac{\partial P_{move|m1}}{\partial T_{s2}} \mathrm{d}T_{s2} = \frac{\partial P_{move|m1}}{\partial T_{m2}} \mathrm{d}T_{m2} + \sum_{s \neq m} \frac{\partial P_{move|m1}}{\partial T_{s2}} \mathrm{d}T_{s2} \\
&= -\beta P_{m2|m1} \cdot (1 - P_{m2|m1}) \mathrm{d}T_{m2} + \sum_{s \neq m} \beta P_{m2|m1} P_{s2|m1} \mathrm{d}T_{s2} \\
&= -\beta P_{m2|m1} \left((1 - P_{m2|m1}) \mathrm{d}T_{m2} - \sum_{s \neq m} P_{s2|m1} \mathrm{d}T_{s2} \right) \\
&= -\beta P_{m2|m1} \left(\mathrm{d}T_{m2} - \sum_{s=1}^{M} P_{s2|m1} \mathrm{d}T_{s2} \right) \\
&= -\beta P_{m2|m1} \left(\mathrm{d}T_{m2} - \overline{\mathrm{d}T}_{2|m1} \right) \\
&= -\beta P_{m2|m1} (1 - P_{m2|m1}[1 - e^{-c}])(\mathrm{d}T_{m2} - \overline{\mathrm{d}T}_2) \tag{31}
\end{aligned}
$$

where $\overline{\mathrm{d}T}_{2|m1} = \sum_{s=1}^{M} P_{s2|m1} \mathrm{d}T_{s2}$ is the average of the net-of-tax changes weighted by the choice probabilities of a household in municipality of origin m and $\overline{\mathrm{d}T}_2 = \sum_{s=1}^{M} P_{s2}^{nc} \mathrm{d}T_{s2}$ is the average weighted by the choice probabilities in the absence of moving costs. Note that $(1 - P_{m2|m1}[1 - e^{-c}]) \in (0, 1]$ equals one in the absence of moving costs, $c = 0$, and zero for infinite moving costs $c \to \infty$.

The last step in Eq. (12) was derived as

$$
\begin{aligned}
\mathrm{d}T_{m2} - \overline{\mathrm{d}T}_{2|m1} &= \mathrm{d}T_{m2} - \sum_{s=1}^{M} P_{s2|m1}\mathrm{d}T_{s2} \\
&= (1 - P_{m2|m1})\mathrm{d}T_{m2} - \sum_{s \neq m} P_{s2|m1}\mathrm{d}T_{s2} \\
&= \left(1 - \frac{e^{V_{m2}}}{e^{V_{m2}} + \sum_{q \neq m} e^{-c}e^{V_{q2}}}\right)\mathrm{d}T_{m2} - \sum_{s \neq m} \frac{e^{-c}e^{V_{s2}}}{e^{V_{m2}} + \sum_{q \neq m} e^{-c}e^{V_{q2}}}\mathrm{d}T_{s2} \\
&= \frac{1}{e^{V_{m2}} + \sum_{q \neq m} e^{-c}e^{V_{q2}}}\left(\sum_{q \neq m} e^{-c}e^{V_{q2}}\mathrm{d}T_{m2} - \sum_{s \neq m} e^{-c}e^{V_{s2}}\mathrm{d}T_{s2}\right) \\
&= \frac{1}{e^{V_{m2}} + \sum_{q \neq m} e^{-c}e^{V_{q2}}}\left(\sum_{q=1}^{M} e^{-c}e^{V_{q2}}\mathrm{d}T_{m2} - \sum_{s=1}^{M} e^{-c}e^{V_{s2}}\mathrm{d}T_{s2}\right) \\
&= \frac{\sum_{q=1}^{M} e^{-c}e^{V_{q2}}}{e^{V_{m2}} + \sum_{q \neq m} e^{-c}e^{V_{q2}}}\left(\mathrm{d}T_{m2} - \sum_{s=1}^{M} \frac{e^{-c}e^{V_{s2}}}{\sum_{q=1}^{M} e^{-c}e^{V_{q2}}}\mathrm{d}T_{s2}\right) \\
&= \frac{e^{-c}e^{V_{m2}} + \sum_{q \neq m} e^{-c}e^{V_{q2}}}{e^{V_{m2}} + \sum_{q \neq m} e^{-c}e^{V_{q2}}}\left(\mathrm{d}T_{m2} - \sum_{s=1}^{M} \frac{e^{-c}e^{V_{s2}}}{\sum_{q=1}^{M} e^{-c}e^{V_{q2}}}\mathrm{d}T_{s2}\right) \\
&= \frac{e^{-c}e^{V_{m2}} + \sum_{q \neq m} e^{-c}e^{V_{q2}}}{e^{V_{m2}} + \sum_{q \neq m} e^{-c}e^{V_{q2}}}\left(\mathrm{d}T_{m2} - \sum_{s=1}^{M} \frac{e^{V_{s2}}}{\sum_{q=1}^{M} e^{V_{q2}}}\mathrm{d}T_{s2}\right) \\
&= (1 - P_{m2|m1}[1 - e^{-c}])\left(\mathrm{d}T_{m2} - \sum_{s=1}^{M} P_{s2}^{nc}\mathrm{d}T_{s2}\right) \\
&= (1 - P_{m2|m1}[1 - e^{-c}])(\mathrm{d}T_{m2} - \overline{\mathrm{d}T}_{2}).
\end{aligned}
$$

A.5 Proof of Hypothesis 2

(a) $\mathrm{d}P_{move|m1} = -\beta P_{m2|m1}(1 - P_{m2|m1}[1 - e^{-c}])(\mathrm{d}T_{m2} - \overline{\mathrm{d}T}_{2}) < 0$ if $\mathrm{d}T_{m2} > \overline{\mathrm{d}T}_{2}$ (low-tax) as $P_{m2|m1} \in [0, 1]$ and $(1 - P_{m2|m1}[1 - e^{-c}]) \in (0, 1]$. (b) Analogously. $\square$

A.6 Generalization of Hypothesis 2

The probability of moving from any municipality of group g to any other municipality is

$$
P_{move|\mathbb{G}_g1} = \sum_{m \in \mathbb{G}_g} \frac{P_{m1}}{P_{\mathbb{G}_g1}} P_{move|m1} \tag{32}
$$

The effect of the simultaneous change of all tax rates τ_{m2} and hence all net-of-tax rates (in logs), $T_{m2} = \log(1 - \tau_{m2})$, is given by the total derivative

$$
\begin{aligned}
\mathrm{d}P_{move|\mathbb{G}_g1} &= \sum_{m \in \mathbb{G}_g} \frac{P_{m1}}{P_{\mathbb{G}_g1}} \mathrm{d}P_{move|m1} \qquad\qquad\qquad (33)\\
&= -\beta \sum_{m \in \mathbb{G}_g} \left[\frac{P_{m1}}{P_{\mathbb{G}_g1}} P_{m2|m1} \left(\mathrm{d}T_{m2} - \overline{\mathrm{d}T}_{2|m1} \right) \right] \\
&= -\beta \sum_{m \in \mathbb{G}_g} \left[\frac{P_{m1}}{P_{\mathbb{G}_g1}} P_{m2|m1}(1 - P_{m2|m1}[1 - e^{-c}])(\mathrm{d}T_{m2} - \overline{\mathrm{d}T}_2) \right]
\end{aligned}
$$

where P_{m1} is the probability of choosing municipality m in period 1 and $P_{\mathbb{G}_g1}$ is the probability of choosing any municipality of group g in period 1.

Hypothesis A.2 (Effect on moving, group version).
The effect of a simultaneous change of all tax rates on the probability of moving, conditional on staying in any municipality of a particular group of municipalities is

a) *strictly negative for groups where all municipalities have net-of-tax rates above average (low-tax), i.e.*

$$
\mathrm{d}P_{move|\mathbb{G}_g1} < 0 \;\; if \;\; \mathrm{d}T_{m2} > \overline{\mathrm{d}T}_2 \;\; for \; all \;\; m \in \mathbb{G}_g,
$$

b) *strictly positive for groups where all municipalities have net-of-tax rates below average (high tax), i.e.*

$$
\mathrm{d}P_{move|\mathbb{G}_g1} > 0 \;\; if \;\; \mathrm{d}T_{m2} < \overline{\mathrm{d}T}_2 \;\; for \; all \;\; m \in \mathbb{G}_g.
$$

Proof: (a) $\mathrm{d}P_{move|\mathbb{G}_g1}$ is a weighted average over $\mathrm{d}P_{move|m1}$. From Hypothesis 2, $\mathrm{d}P_{move|m1} < 0$ if $\mathrm{d}T_{m2} > \overline{\mathrm{d}T}_2$. Hence $\mathrm{d}P_{move|\mathbb{G}_g1} < 0$ as $\mathrm{d}T_{m2} > \overline{\mathrm{d}T}_2$ for all $m \in \mathbb{G}_g$. (b) Analogously. $\square$

A.7 Derivation of the migration elasticity

The elasticity of the unconditional probability P_{m2} of choosing a municipality m w.r.t. the net-of-tax rate in that municipality is

$$
\begin{aligned}
\eta_m &= \frac{\partial P_{m2}}{\partial(1 - \tau_{m2})} \frac{(1 - \tau_{m2})}{P_{m2}} = \frac{\partial \log(P_{m2})}{\partial \log(1 - \tau_{m2})} = \frac{\partial P_{m2}}{\partial T_{m2}} \frac{1}{P_{m2}} \\
&= \frac{1}{P_{m2}} \sum_{s=1}^{M} \frac{\partial P_{m2|s1}}{\partial T_{m2}} P_{s1} \\
&= \beta \frac{1}{P_{m2}} \sum_{s=1}^{M} P_{m2|s1}(1 - P_{m2|s2})P_{s1} \qquad\qquad (34)
\end{aligned}
$$

where $P_{m2} = \sum_{s=1}^{M} P_{m2|s1} P_{s1}$.

The in-migration elasticity is given by

$$
\begin{aligned}
\eta_m^{in} &= \frac{\partial \sum\limits_{s\neq m} P_{m2|s1}P_{s1}}{\partial(1-\tau_{m2})}\frac{(1-\tau_{m2})}{\sum\limits_{s\neq m} P_{m2|s1}P_{s1}} = \frac{\partial \sum\limits_{s\neq m} P_{m2|s1}P_{s1}}{\partial log(1-\tau_{m2})}\frac{1}{\sum\limits_{s\neq m} P_{m2|s1}P_{s1}} \\
&= \frac{1}{\sum\limits_{s\neq m} P_{m2|s1}P_{s1}}\sum_{s\neq m}\frac{\partial P_{m2|s1}}{\partial \log(1-\tau_{m2})}P_{s1} = \frac{1}{(P_{m2}-P_{m2|m1}\,P_{m1})}\sum_{s\neq m}\frac{\partial P_{m2|s1}}{\partial T_{m2}}P_{s1} \\
&= \beta\frac{1}{(P_{m2}-P_{m2|m1}\,P_{m1})}\sum_{s\neq m} P_{m2|s1}\left(1-P_{m2|s1}\right)P_{s1} \qquad\qquad (35)
\end{aligned}
$$

while the out-migration elasticity is

$$
\begin{aligned}
\eta_m^{out} &= \frac{\partial log[(1-P_{m2|m1})P_{m1}]}{\partial \log(1-\tau_{m2})} = \frac{\partial(1-P_{m2|m1})P_{m1}}{\partial T_{m2}}\frac{1}{(1-P_{m2|m1})P_{m1}} \\
&= -\frac{1}{1-P_{m2|m1}}\frac{\partial P_{m2|m1}}{\partial T_{m2}} = -\frac{1}{1-P_{m2|m1}}\beta P_{m2|m1}(1-P_{m2|m1}) \\
&= -\beta\,P_{m2|m1}. \qquad\qquad (36)
\end{aligned}
$$

B Appendix: Empirical strategy

B.1 Structural estimation with groups of municipalities

To obtain an explicit structural expression for estimation, we solve Eq. (28) for β. This yields an estimate for β of the following form:

$$\widehat{\beta} \;=\; \frac{\widehat{\mathrm{d}P}_{\mathbb{G}_g 2}}{\sum\limits_{s=1}^{M}\left[\widehat{P}_{s1}\widehat{P}_{\mathbb{G}_g 2|s1}\left(\widehat{\overline{\mathrm{d}T}}_{\mathbb{G}_g 2|s1} - \widehat{\overline{\mathrm{d}T}}_{2|s1}\right)\right].} \tag{37}$$

Here, $\widehat{\mathrm{d}P}_{\mathbb{G}_g 2}$ is the discontinuity estimate from the fuzzy regression discontinuity design, $\widehat{P}_{s1}$ is the share of households in municipality s, $\widehat{P}_{\mathbb{G}_g 2|s1} = \sum_{m\in\mathbb{G}_g}\widehat{P}_{m2|s1}$, $\widehat{\overline{\mathrm{d}T}}_{2|s1} = \sum_{m=1}^{M}\widehat{P}_{m2|s1}\widehat{\mathrm{d}T}_{m2}$, and $\widehat{\overline{\mathrm{d}T}}_{\mathbb{G}_g 2|s1} = \sum_{m\in\mathbb{G}_g}\frac{\widehat{P}_{m2|s1}}{\widehat{P}_{\mathbb{G}_g 2|s1}}\widehat{\mathrm{d}T}_{m2}$. The net-of-tax rate change $\widehat{\mathrm{d}T}_{m2}$ is the average over the individual net-of-tax rate changes $\mathrm{d}T_{im}$ of all households residing in municipality m in period 1.

Alternatively, we can also estimate β from $\mathrm{d}P_{move|\mathbb{G}_g 1}$ by solving Eq. (33)

$$\widehat{\beta} \;=\; -\frac{\widehat{\mathrm{d}P}_{move|\mathbb{G}_g 1}}{\sum\limits_{m\in\mathbb{G}_g}\left[\frac{\widehat{P}_{m1}}{\widehat{P}_{\mathbb{G}_g 1}}\widehat{P}_{m2|m1}\left(\widehat{\mathrm{d}T}_{m2} - \widehat{\overline{\mathrm{d}T}}_{2|m1}\right)\right]}$$

where $\widehat{\overline{\mathrm{d}T}}_{2|s1} = \sum_{m=1}^{M}\widehat{P}_{m2|s1}\mathrm{d}T_{m2}$ and $\widehat{\mathrm{d}P}_{move|\mathbb{G}_g 1}$ is again the discontinuity estimate from the corresponding fuzzy RDD.

The migration elasticity defined in Section 3.5 is

$$\eta_m \;=\; \frac{\partial\log(P_{m2})}{\partial\log(1-\tau_{m2})} = \beta\frac{1}{P_{m2}}\sum_{s=1}^{M}\left[P_{m2|s1}(1-P_{m2|s1})P_{s1}\right]$$

We estimate the migration elasticity as

$$\widehat{\eta}_m = \widehat{\beta}\frac{1}{\widehat{P}_{m2}}\sum_{s=1}^{M}\left[\widehat{P}_{m2|s1}(1-\widehat{P}_{m2|s1})\widehat{P}_{s1}\right]. \tag{38}$$

The migration elasticity varies by municipality m. We summarize these municipality-specific parameters by reporting the average elasticity over all municipalities weighted by the municipal population shares P_{m2}:

$$\widehat{\eta} = \sum_{m=1}^{M}\widehat{P}_{m2}\widehat{\eta}_m = \widehat{\beta}\sum_{m=1}^{M}\sum_{s=1}^{M}\left[\widehat{P}_{m2|s1}(1-\widehat{P}_{m2|s1})\widehat{P}_{s1}\right]. \tag{39}$$

B.2 Structural estimation with spatial pooling

The theoretical model outlined above describes the choice of a group of municipalities within a canton where the control group faces identical tax rates across all municipalities. As discussed in Section 4.4, we pool the location choices over all cantons considering all

municipalities of a canton as the choice set for the household. Therefore, we model the observed treatment effect on location choice or mobility in the pooled data as the average effect over all cantons k, weighted by the population share P_{k1} of canton k in period 1. Consequently, we estimate β as

$$\widehat{\beta} = \frac{\widehat{\mathrm{d}P}_{\mathbb{G}_g 2}}{\sum_{k=1}^{K} \widehat{P}_{k1} \sum_{s=1}^{M_k} \left[\widehat{P}_{s1} \widehat{P}_{\mathbb{G}_g 2|s1} \left(\widehat{\overline{\mathrm{d}T}}_{\mathbb{G}_g 2|s1} - \widehat{\overline{\mathrm{d}T}}_{2|s1} \right) \right]}, \tag{40}$$

and

$$\widehat{\beta} = -\frac{\widehat{\mathrm{d}P}_{move|\mathbb{G}_g 1}}{\sum_{k=1}^{K} \widehat{P}_{k1} \sum_{m \in \mathbb{G}_g} \left[\frac{\widehat{P}_{m1}}{\widehat{P}_{\mathbb{G}_g 1}} \widehat{P}_{m2|m1} (\widehat{\mathrm{d}T}_{m2} - \widehat{\overline{\mathrm{d}T}}_{2|m1}) \right]}, \tag{41}$$

where all population shares are within cantons such that $\sum_{m=1}^{M_k} \widehat{P}_{m2|s1} = 1$ and $\sum_{m=1}^{M_k} \widehat{P}_{m2} = 1$. M_k is the number of municipalities in canton k. The migration elasticity is then also based on the M_k municipalities of the respective canton k:

$$\widehat{\eta}_m = \widehat{\beta} \frac{1}{\widehat{P}_{m2}} \sum_{s=1}^{M_k} \left[\widehat{P}_{m2|s1} (1 - \widehat{P}_{m2|s1}) \widehat{P}_{s1} \right]. \tag{42}$$

We report the weighted average over all M Swiss municipalities in all K cantons:

$$\widehat{\eta} = \sum_{m=1}^{M} \widehat{P}_{k2} \widehat{P}_{m2} \widehat{\eta}_m = \sum_{k=1}^{K} \widehat{P}_{k2} \sum_{m=1}^{M_k} \widehat{P}_{m2} \widehat{\eta}_m = \widehat{\beta} \sum_{k=1}^{K} \widehat{P}_{k2} \sum_{m=1}^{M_k} \sum_{s=1}^{M_k} \left[\widehat{P}_{m2|s1} (1 - \widehat{P}_{m2|s1}) \widehat{P}_{s1} \right], \tag{43}$$

where $\widehat{P}_{m2}$ is the population share of municipality m within the respective canton, $\widehat{P}_{k2}$ is the population share of canton k within Switzerland and $\widehat{P}_{k2} \widehat{P}_{m2}$ is the population share of municipality m within Switzerland.

C Appendix: Additional tables and figures

Table C.1: Descriptive Statistics for Experimental Sample

Variable	Mean	St. Dev.	Median	Min	Max	Obs
Individual characteristics						
Gross income in 1000	61.975	26.683	63.700	0.000	114.000	19,313
Duration in 100	28.804	19.838	22.240	3.650	73.000	19,313
Permit C	0.497			0	1	19,313
Age	36.500	8.756	36	18	78	19,313
Type of residential municipality and moving						
Low-tax	0.515			0	1	19,313
Very low-tax	0.267			0	1	19,313
Medium low-tax	0.249			0	1	19,313
Medium high-tax	0.230			0	1	19,313
Very high-tax	0.255			0	1	19,313
High-tax	0.485			0	1	19,313
Move	0.061			0	1	19,313
Difference between ordinary and special income tax rate in %, all incomes (0 – 114k)						
All observations	−0.065	0.575	0.000	−4.332	2.424	19,313
Obs. in low-tax municipalities	−0.461	0.496	−0.328	−4.332	0.000	9,954
Obs. in high-tax municipalities	0.356	0.276	0.309	0.000	2.424	9,359
Difference between ordinary and special income tax rate in %, high incomes (60k – 114k)						
All observations	−0.089	0.687	0.038	−4.332	2.406	10,716
Obs. in low-tax municipalities	−0.635	0.549	−0.495	−4.332	0.000	5,208
Obs. in very low-tax municipalities	−0.924	0.563	−0.760	−4.332	0.000	2,950
Obs. in medium low-tax municipal.	−0.258	0.177	−0.238	−1.078	0.000	2,258
Obs. in medium high-tax municipal.	0.288	0.192	0.264	0.000	1.512	2,508
Obs. in very high-tax municipalities	0.545	0.303	0.449	0.168	2.406	3,000
Obs. in high-tax municipalities	0.428	0.288	0.370	0.000	2.406	5,508

Notes: Summary statistics for the main variables in the experimental sample for observations with durations of stay up to 20 years in Switzerland.
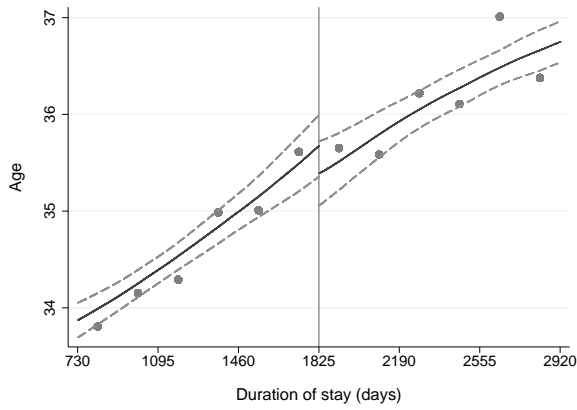
(a) Age
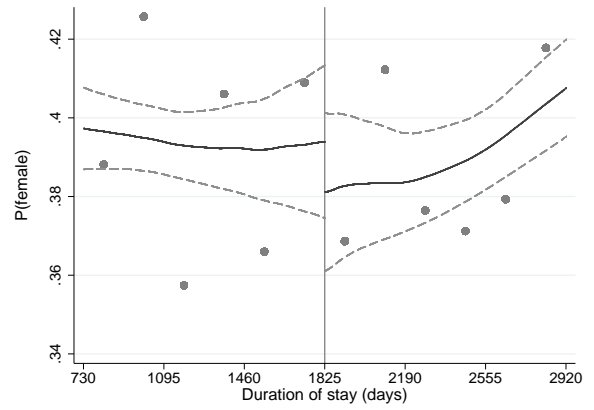
(b) Gender
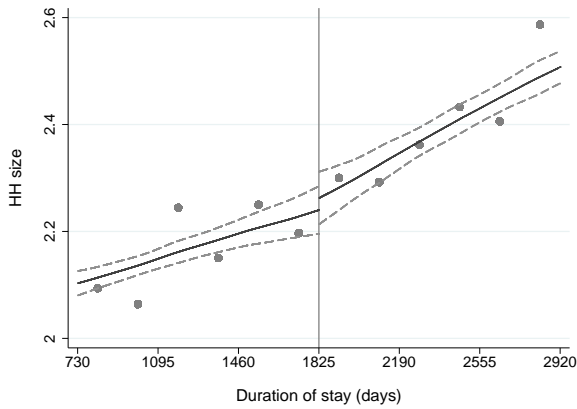
(c) HHsize
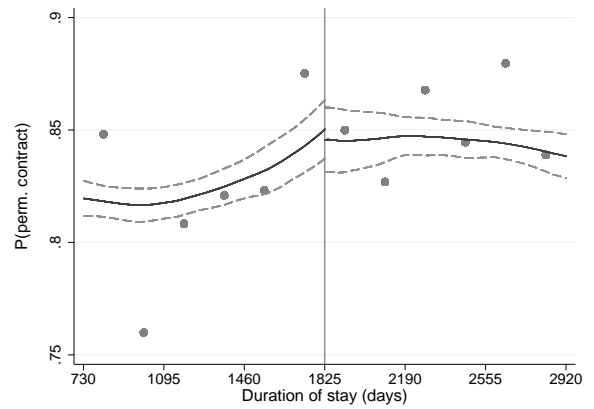
(d) Permanent contract

Figure C.1: Individuals' characteristics around the threshold of 5 years of stay. These graphs show local linear smooths of the covariate values of individuals around the threshold, using the whole sample and a bandwidth of 1460 days. The light gray dashed lines indicate the 90% pointwise confidence bands and the gray dots raw weighted averages of the dependent variable in half yearly bins. The corresponding discontinuity estimates can be found in Table C.5

Table C.2: Probability of holding permit C

	Full sample		Low-tax municipalities		High-tax municipalities		Very high-tax municipalities	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect ΔP	0.223***	0.274***	0.247***	0.279***	0.179***	0.231***	0.169***	0.234***
	(0.023)	(0.015)	(0.035)	(0.023)	(0.036)	(0.023)	(0.049)	(0.032)
Bandwidth	730	1460	730	1460	730	1460	730	1460
N left	4787	9760	1978	4061	1913	3927	1048	2146
N right	3426	5543	1338	2215	1390	2253	746	1188

	Medium high-tax municipalities		Medium low-tax municipalities		Very low-tax municipalities	
	(9)	(10)	(11)	(12)	(13)	(14)
Effect ΔP	0.191***	0.228***	0.255***	0.319***	0.245***	0.248***
	(0.053)	(0.034)	(0.052)	(0.033)	(0.048)	(0.032)
Bandwidth	730	1460	730	1460	730	1460
N left	865	1781	949	1953	1029	2108
N right	644	1065	627	1061	711	1154

	Low-income group		High-income group		Placebo sample	
	(15)	(16)	(17)	(18)	(19)	(20)
Effect ΔP	0.169***	0.218***	0.239***	0.279***	0.309***	0.389***
	(0.040)	(0.026)	(0.032)	(0.021)	(0.066)	(0.044)
Bandwidth	730	1460	730	1460	730	1460
N left	1496	3217	2395	4771	553	1062
N right	1041	1775	1687	2693	439	667

Notes: Local linear SRD estimates of the discontinuity in the probability of treatment, i.e. holding permit C, for different municipality types, income groups, and bandwidths. Estimates account for sampling weights. Standard errors in parentheses. Sample specification: 'Full sample' is based on the whole defined sample, 'Placebo' is based on the placebo sample, and the remaining estimates are based on the respective groups within the experimental sample. Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.3: Probability of moving away from a particular *high*-tax municipality for households with and without children

		Low − income		High − income		Placebo − income	
		non-parametric	parametric	non-parametric	parametric	non-parametric	parametric
		(1)	(2)	(3)	(4)	(5)	(6)
(a)	*Child in Household*						
	Effect ΔP	-0.057	-0.028	0.235*	0.221**	-0.007	0.02
		(0.138)	(0.100)	(0.135)	(0.108)	(0.134)	(0.102)
	N left	453	453	449	449	147	147
	N right	409	409	401	401	85	85
(b)	No-Child in Household						
	Effect ΔP	0.503	0.322*	0.420***	0.294**	-0.023	-0.008
		(0.380)	(0.195)	(0.140)	(0.141)	(0.118)	(0.075)
	N left	1033	1033	1992	1992	326	326
	N right	422	422	1021	1021	231	231

Notes: Fuzzy RDD estimates of the probability of moving given the individual was residing in a high-tax municipality for different income groups, separately for households with and without children. Nonparametric means fuzzy RDD with local-linear regression using a triangular kernel, while parametric means a linear instrumental variable estimation. All estimates include a bandwidth of 1460 days (4years) and account for sampling weights. Standard errors in parentheses. Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.4: Implied Migration Elasticity

		Based on estimated effect on very high tax municipalities		Based on estimated effect on very low tax municipalities	
		Point Estimate	95%-Confidence Interval	Point Estimate	95%-Confidence Interval
		(1)	(2)	(3)	(4)
(a)	*Elasticity estimates based on effects on residential choices (Table 2)*				
	Effect ΔP	-0.159	[-0.323, -0.013]	0.065	[-0.076, 0.207]
	Parameter β	208.8	[18.39, 484.27]	80.39	[-107.9, 297.0]
	Elasticity $\overline{\eta}$	22.34	[1.98, 45.86]	8.597	[-10.22, 27.74]
(b)	*Elasticity estimates based on effects on mobility decisions (Table 4)*				
	Effect ΔP	0.356	[0.090, 0.786]	0.023	[-0.128, 0.177]
	Parameter β	251.4	[77.81, 702.1]	-19.66	[-182.8, 130.3]
	Elasticity $\overline{\eta}$	26.89	[7.173, 65.89]	-2.103	[-17.48 , 12.26]

Notes: Estimates of the average migration elasticity with respect to the net-of-tax rate based on the parametric fuzzy RDD estimates on residential choices (Table 2) and mobility decisions (Table 4) for the high income experimental sample. Confidence bounds are determined by a non-parametric bootstrap with 2000 replications and the percentile method.

Table C.5: Covariate balance around the threshold

	Age (years)			P(Female)			HH size		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Effect Δ	-0.480	-0.405	-0.287	-0.088**	-0.036	-0.013	0.107	0.067	0.022
	(0.712)	(0.509)	(0.367)	(0.042)	(0.030)	(0.022)	(0.099)	(0.070)	(0.050)
Mean dep. var.	36.35	36.35	36.35	0.41	0.41	0.41	2.13	2.13	2.13
Bandwidth	365	730	1460	365	730	1460	365	730	1460
N left	2210	4787	9760	2210	4787	9760	2210	4787	9760
N right	1930	3426	5543	1930	3426	5543	1930	3426	5543

	P(High educ)			P(perm. contract)		
	(10)	(11)	(12)	(13)	(14)	(15)
Effect Δ	-0.048	-0.034	-0.011	-0.022	-0.026	-0.005
	(0.043)	(0.031)	(0.022)	(0.027)	(0.020)	(0.014)
Mean dep. var.	0.51	0.51	0.51	0.89	0.89	0.89
Bandwidth	365	730	1460	365	730	1460
N left	2210	4787	9760	2210	4787	9760
N right	1930	3426	5543	1930	3426	5543

Notes: Local linear SRD estimates of the discontinuity in individuals' characteristics for the sample of individuals around the duration threshold in the full sample. Estimates account for sampling weights. Standard errors in parentheses. Mean dep. var. stands for the mean of the dependent variable for individuals between 4 to 5 years of stay. Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.6: Probability of moving away from a particular municipality for four groups of municipalities and several bandwidths: Low-income group

	Probability of moving from a very high-tax				Probability of moving from a medium high-tax			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect ΔP	12.878	0.496	0.194	0.360	0.074	0.153	0.075	0.130
	(440.361)	(0.796)	(0.199)	(0.598)	(0.635)	(0.440)	(0.233)	(0.397)
Bandwidth	365	730	1460	842	365	730	1460	806
N left	155	353	758	422	151	341	728	388
N right	160	279	445	313	117	213	386	244

	Probability of moving from a medium low-tax				Probability of moving from a very low-tax			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Effect ΔP	0.843	0.045	0.133	0.042	-0.322	-0.379	-0.036	-0.354
	(1.270)	(0.188)	(0.115)	(0.172)	(0.359)	(0.342)	(0.271)	(0.319)
Bandwidth	365	730	1460	846	365	730	1460	822
N left	204	441	944	520	155	361	787	409
N right	158	294	513	329	135	255	431	279

Notes: Local linear FRD estimates of the probability of moving conditional on the previous municipality type for the low-income experiment sample and different bandwidths. Estimates account for sampling weights. Standard errors in parentheses. Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.7: Probability of moving away from a particular municipality for four groups of municipalities and several bandwidths: High-income group

	Probability moving from a very high-tax				Probability of moving from a medium high-tax			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect ΔP	1.261*	0.756***	0.517***	0.749***	0.300	0.224	0.130	0.219
	(0.736)	(0.278)	(0.145)	(0.274)	(0.297)	(0.224)	(0.138)	(0.223)
Bandwidth	365	730	1460	738	365	730	1460	740
N left	333	695	1388	699	240	524	1053	529
N right	268	467	743	472	246	431	679	440

	Probability moving from a medium low-tax				Probability of moving from a very low-tax			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Effect ΔP	-0.102	-0.124	-0.090	-0.130	-0.252	-0.079	0.015	-0.079
	(0.287)	(0.170)	(0.097)	(0.156)	(0.245)	(0.148)	(0.097)	(0.148)
Bandwidth	365	730	1460	808	365	730	1460	724
N left	237	508	1009	562	313	668	1321	664
N right	175	333	548	368	277	456	723	454

Notes: Local linear FRD estimates of the probability of moving conditional on the previous municipality type for the high-income experiment sample and different bandwidths. Estimates account for sampling weights. Standard errors in parentheses. Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.8: Probability of moving away from a particular municipality for four groups of municipalities and several bandwidths: Quasi-placebo group

	Probability of moving from a very high-tax				Probability of moving from a medium high-tax			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect ΔP	-0.100	0.015	0.053	-0.033	-0.110	-0.141	-0.079	-0.160
	(0.319)	(0.248)	(0.147)	(0.290)	(0.259)	(0.165)	(0.106)	(0.197)
Bandwidth	365	730	1460	522	365	730	1460	594
N left	89	189	303	134	37	83	170	68
N right	72	122	184	88	51	88	132	76

	Probability of moving from a medium low-tax				Probability of moving from a very low-tax			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Effect ΔP	-11.511	-0.189	-0.023	-13.656	0.084	0.0255	-0.000	0.068
	(272.528)	(0.57)	(0.153)	(377.917)	(0.215)	(0.179)	(0.124)	(0.208)
Bandwidth	365	730	1460	324	365	730	1460	535
N left	47	109	216	39	84	172	373	134
N right	43	81	139	38	76	148	212	106

Notes: Local linear FRD estimates of the probability of moving conditional on the previous municipality type for the placebo sample and different bandwidths. Estimates account for sampling weights. Standard errors in parentheses. Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.9: Probability of moving away from a particular municipality for four groups of municipalities and different income groups

	1st order polynomial				2nd order polynomial			
	$P(m\|vh)$	$P(m\|mh)$	$P(m\|ml)$	$P(m\|vl)$	$P(m\|vh)$	$P(m\|mh)$	$P(m\|ml)$	$P(m\|vh)$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(a) *Low-income*								
Effect ΔP	0.074	0.071	0.195**	0.218	−0.088	0.125	0.052	−0.482*
	(0.075)	(0.194)	(0.086)	(0.163)	(0.286)	(0.601)	(0.182)	(0.281)
No. obs.	1196	1111	1450	1213	1196	1111	1450	1213
No. of clusters	23	22	23	23	23	22	23	23
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(b) *High-income*								
Effect ΔP	0.171***	0.156*	−0.010	0.030	0.312**	0.079	−0.198	−0.077
	(0.063)	(0.092)	(0.036)	(0.072)	(0.153)	(0.229)	(0.176)	(0.118)
No. obs.	2123	1727	1554	2037	2123	1727	1554	2037
No. of clusters	23	23	23	23	23	23	23	23
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(c) *Placebo-income*								
Effect ΔP	0.040	−0.032	0.019	−0.005	0.097	−0.294*	0.122	−0.022
	(0.063)	(0.055)	(0.095)	(0.059)	(0.069)	(0.171)	(0.177)	(0.137)
No. obs.	487	302	354	583	487	302	354	583
No. of clusters	17	20	20	21	17	20	20	21
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard parametric FRD (IV) estimates using a linear probability model for foreigners whose duration of stay in Switzerland was between 1 and 8 years. The estimates are reported for all four municipality types, the low and high income group, as well as for the quasi-placebo group with an income of above 126,000 Swiss francs. Estimates account for sampling weights and the standard errors are clustered at the cantonal level. The controls encompass: age, age squared, sex, an indicator for a child in the household, the household size, income, civil status, education, occupation, and cantonal fixed effects. Standard errors in parentheses. *Significance levels:* * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.10: Probability of moving away from a particular municipality for four groups of municipalities at a placebo threshold of 4 years' stay

	Probability of moving from a very high-tax				Probability of moving from a medium high-tax			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect ΔP	-1.328	0.416	-0.035	0.382	0.867	0.611	0.342	0.686
	(3.853)	(0.427)	(0.326)	(0.409)	(0.928)	(0.492)	(0.416)	(0.541)
Bandwidth	365	730	1460	751	365	730	1460	672
N left	361	765	1054	787	284	587	813	543
N right	334	602	962	612	240	486	819	451

	Probability of moving from a medium low-tax				Probability of moving from a very low-tax			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Effect ΔP	0.309	-1.1630	-0.866	-0.865	-0.888	0.077	0.312	0.084
	(0.853)	(2.144)	(1.185)	(1.338)	(1.501)	(0.688)	(0.440)	(0.670)
Bandwidth	365	730	1460	808	365	730	1460	737
N left	271	552	772	616	355	724	1008	731
N right	237	412	695	446	313	590	926	597

Notes: Local linear FRD estimates at the placebo threshold of four years of the probability of moving conditional on the previous municipality type for the high-income experimental sample and different bandwidths. Estimates account for sampling weights. Standard errors in parentheses.
Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

(a) High-tax municipalities   **(b)** Low-tax municipalities

Figure C.2: Income distribution for foreigners not holding the permanent residence permit with a duration of stay below 5 years. Local likelihood density estimates using $bandwidth = 5k$ and taking the sampling weights into account, based on the local likelihood implementation (locfit) in the Chronux software package for Matlab (Bokil et al., 2010). The corresponding density discontinuity estimates can be found in row (a) of Table C.17.
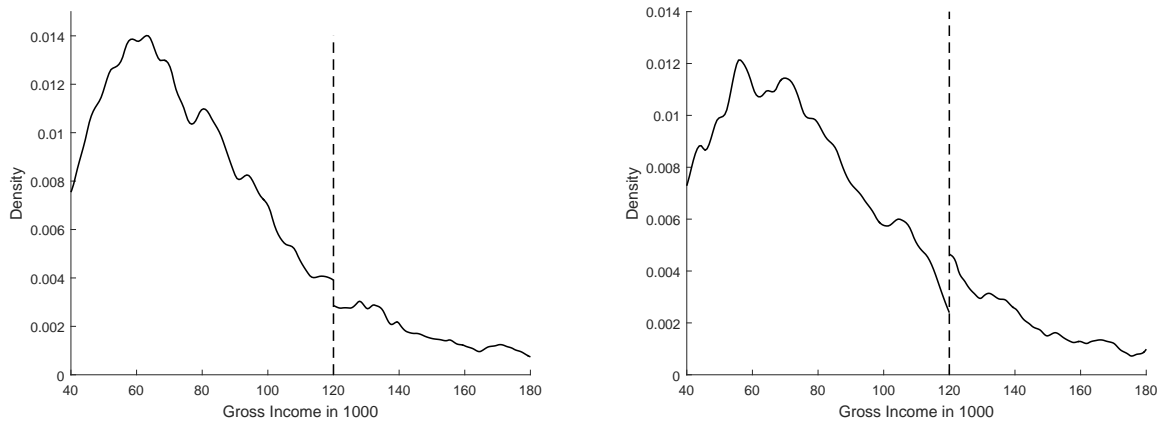
Table C.11: Probability of moving away from a particular municipality for four groups of municipalities at a placebo threshold of 6 years' stay

	Probability of moving from a very high-tax				Probability of moving from a medium high-tax			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect ΔP	0.272	-0.520	-0.103	-0.524	0.410	0.987	0.213	1.215
	(0.894)	(1.852)	(0.167)	(1.007)	(1.036)	(2.585)	(0.210)	(3.806)
Bandwidth	365	730	1460	900	365	730	1460	670
N left	267	601	1367	779	244	486	1073	445
N right	199	360	571	425	185	333	512	319

	Probability of moving from a medium low-tax				Probability of moving from a very low-tax			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Effect ΔP	-4.744	0.010	-0.120	0.008	-0.309	0.872	0.035	0.986
	(196.059)	(0.800)	(0.215)	(0.755)	(0.992)	(5.715)	(0.215)	(7.428)
Bandwidth	365	730	1460	719	365	730	1460	742
N left	175	412	964	401	277	590	1314	598
N right	158	283	443	281	179	336	546	345

Notes: Local linear FRD estimates at the placebo threshold of six years of the probability of moving conditional on the previous municipality type for the high-income experimental sample and different bandwidths. Estimates account for sampling weights. Standard errors in parentheses.
Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.12: Probability of residing in a particular municipality type at a placebo threshold of 4 years' stay

	Probability of residing in a very high-tax				Probability of residing in a very low-tax			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect ΔP	-2.539	-0.05	0.157	4.558	4.780	-0.742	-0.658	-1.16
	(3.916)	(0.503)	(0.376)	(17.516)	(6.005)	(0.544)	(0.406)	(0.844)
Bandwidth	365	730	1460	465	365	730	1460	625
N left	1271	2628	3647	1627	1271	2628	3647	2201
N right	1124	2090	3402	1425	1124	2090	3402	1842

Notes: Local linear FRD estimates at the placebo threshold of four years of the probability that an individual is residing in the respective municipality type for the high-income experimental sample and different bandwidths. Estimates account for sampling weights. Standard errors in parentheses.
Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.13: Probability of residing in a particular municipality type at a placebo threshold of 6 years' stay

	Probability of residing in a very high-tax				Probability of residing in a very low-tax			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect ΔP	-0.539	1.182	0.110	0.662	0.0514	-3.010	-0.287	-4.649
	(0.895)	(6.003)	(0.181)	(1.613)	(0.795)	(12.08)	(0.185)	(33.713)
Bandwidth	365	730	1460	836	365	730	1460	657
N left	963	2089	4718	2451	963	2089	4718	1847
N right	721	1312	2072	1487	721	1312	2072	1212

Notes: Local linear FRD estimates at the placebo threshold of six years of the probability that an individual resides in the respective municipality type for the high-income experimental sample and different bandwidths. Estimates account for sampling weights. Standard errors in parentheses.
Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.14: Density discontinuity estimates for the duration of stay

	c	h in 100	$\hat{f}_l$	$\hat{f}_r$	$\hat{\theta}$	$\hat{l}$	$\hat{f}_r/\hat{f}_l$	p-value	N
(a)	18.25	1.80	0.015	0.017	0.002	0.727	1.102	0.394	33,109
(b)	18.25	3.65	0.016	0.016	−0.000	0.003	0.995	0.955	33,109
(c)	18.25	7.30	0.016	0.015	−0.001	0.520	0.957	0.471	33,109
(d)	18.25	4.96	0.016	0.016	−0.000	0.001	0.998	0.979	33,109

Notes: Local likelihood ratio results for the discontinuity in the distribution of the duration of stay, for several bandwidths. $\hat{\theta}$ stands for the discontinuity estimate ($\hat{\theta} = \hat{f}_r - \hat{f}_l$), $\hat{l}r$ is the value of the empirical likelihood ratio, $\hat{f}_r/\hat{f}_l$ is the relative size of the discontinuity, and N stands for the number of observations with regard to the observations available to estimate the whole density in the sample.

The estimates for the discontinuity in the density are based on the approach in Otsu et al. (2013) which was adjusted for the sampling weights in our data. We multiply the kernel weights by the sampling weights in the summation part of the FOC.

Table C.15: Determinants of the probability of residing in a low-tax municipality

	Duration ≤ 2 years		Duration ≥ 8 years	
	Probability of residing in		Probability of residing in	
	low-tax	very low-tax	low-tax	very low-tax
	(1)	(2)	(3)	(4)
Demographic:				
Age	0.001 (0.002)	0.003 (0.002) *	0.000 (0.002)	0.002 (0.001)
Married	0.066 (0.037) *	−0.027 (0.034)	0.038 (0.036)	0.013 (0.029)
Female	0.068 (0.032) **	0.027 (0.029)	−0.063 (0.031) **	−0.037 (0.029)
Child	0.012 (0.044)	0.070 (0.042) *	−0.000 (0.035)	0.016 (0.030)
Income:				
1000 CHF	0.000 (0.001)	0.000 (0.001)	0.001 (0.001)	0.001 (0.001)
Highest educational attainment:				
High educ.	−0.024 (0.039)	−0.014 (0.037)	−0.052 (0.040)	−0.062 (0.035) *
Medium (ref.)				
Low educ.	0.005 (0.056)	−0.001 (0.051)	−0.063 (0.037) *	−0.050 (0.030) *
In educ.	0.083 (0.148)	−0.066 (0.161)	−0.157 (0.085) *	−0.118 (0.071) *
Educ NA	−0.106 (0.234)	0.147 (0.213)	0.086 (0.184)	−0.044 (0.174)
Work contract:				
Perm. contract	0.057 (0.044)	0.038 (0.037)	0.011 (0.054)	−0.039 (0.047)
Current occupation:				
Craftsmen (ref.)				
Manager	0.050 (0.069)	0.057 (0.066)	0.138 (0.066) **	0.130 (0.060) **
Academic profession	0.015 (0.055)	0.037 (0.050)	0.088 (0.059)	0.093 (0.053) *
Engeneers/Technician	−0.016 (0.055)	0.025 (0.050)	0.130 (0.051) **	0.109 (0.049) **
Service/Salesman	0.025 (0.065)	0.081 (0.063)	0.126 (0.046) ***	0.076 (0.037) **
Forestry	0.244 (0.186)	0.019 (0.188)	0.387 (0.119) ***	0.331 (0.137) **
Machine operator	−0.025 (0.085)	0.035 (0.083)	0.055 (0.054)	0.014 (0.044)
Elementary occupation	−0.077 (0.108)	−0.080 (0.082)	0.157 (0.071) **	0.023 (0.042)
Occupation NA	−0.167 (0.101) *	−0.058 (0.080)	0.033 (0.154)	0.019 (0.133)
Industry sector:				
Construction (ref.)				
Land & Forestry	0.195 (0.213)	−0.110 (0.185)	−0.073 (0.151)	−0.074 (0.154)
Industry	0.106 (0.087)	0.016 (0.080)	0.034 (0.082)	0.158 (0.063) **
Energy/Water supply	−0.202 (0.314)	−0.104 (0.306)	0.414 (0.076) ***	0.459 (0.224) **
Trade	0.068 (0.096)	−0.041 (0.088)	0.098 (0.086)	0.188 (0.072) ***
Catering industry	−0.027 (0.113)	−0.056 (0.100)	−0.160 (0.086) *	−0.045 (0.056)
Traffic/News	0.039 (0.120)	−0.125 (0.094)	−0.090 (0.119)	0.014 (0.104)
Financial services	−0.080 (0.112)	−0.070 (0.098)	0.112 (0.143)	0.333 (0.137) **
Real estate/Informatics	0.090 (0.093)	0.069 (0.085)	−0.014 (0.102)	0.067 (0.070)
Public services	0.258 (0.156) *	0.097 (0.178)	−0.018 (0.153)	0.014 (0.148)
Education	0.023 (0.098)	−0.061 (0.086)	−0.148 (0.105)	−0.032 (0.067)
Health & social services	0.070 (0.092)	0.042 (0.087)	−0.124 (0.084)	0.010 (0.061)
Other services	0.123 (0.116)	−0.144 (0.091)	0.030 (0.124)	0.067 (0.080)
Private households	0.087 (0.187)	−0.029 (0.148)	0.072 (0.165)	0.323 (0.139) **
Other/ Sector NA	−0.076 (0.207)	0.121 (0.177)	−0.207 (0.137)	−0.066 (0.104)
FE year	Yes	Yes	Yes	Yes
Observations	1669	1669	2234	2234
R^2	0.03	0.03	0.05	0.06
Adj. R^2	0.01	0.01	0.03	0.04

Notes: Correlation between individual characteristics and the probability that an individual resides in a low-tax or a very low-tax municipality. Sample specification: Specifications (1) and (2) use the experimental sample of individuals with an income above 10,000 and below 114,000 Swiss francs and not holding the permanent residence permit in their first 2 years of stay in the country. Specifications (3) and (4) use the sample of individuals in the experimental sample having an income of above 10,000 and below 114,000 Swiss francs but already holding the permanent residence permit in their 8 to 12 years of stay in the country. Estimates account for sampling weights, standard errors are reported in brackets to the right of the point estimates. *Significance levels:* * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.16: Probability of residing in a low-tax or very-low tax municipality

		Basic sample			Excluding manipulation range ($\pm 6k$)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(a) *Probability of residing in a low-tax municipality*								
Effect ΔP	0.221	0.298**	0.263**	0.148**	-0.081	-0.125	-0.083	-0.090
	(0.207)	(0.145)	(0.102)	(0.067)	(0.199)	(0.142)	(0.100)	(0.069)
Bandwidth	2.5k	5k	10k	21.918k	2.5k	5k	10k	20.132k
N left	80	180	407	1111	109	225	488	1181
N right	93	186	335	669	74	151	297	555
(b) *Probability of residing in a very low-tax municipality*								
Effect ΔP	0.195	0.257*	0.186*	0.041	-0.207	-0.122	-0.136	-0.136*
	(0.204)	(0.141)	(0.097)	(0.067)	(0.174)	(0.127)	(0.092)	(0.073)
Bandwidth	2.5k	5k	10k	19.840k	2.5k	5k	10k	15.600k
N left	80	180	407	963	109	225	488	836
N right	93	186	335	628	74	151	297	462

Notes: Local linear sharp RDD estimates for the probability that an individual resides in a low or a very low-tax municipality at the income threshold of 120,000 Swiss francs for the sample of foreigners not holding the permanent residence status (permit C) during their first five years in Switzerland. Estimates account for sampling weights. Standard errors in parentheses.
Significance levels: * $.05 < p < .1$, ** $.01 < p < .05$, *** $p < .01$.

Table C.17: Density discontinuity estimates for income

	c	h	$\hat{f}_l$	$\hat{f}_r$	$\hat{\theta}$	$\hat{l}r$	$\hat{f}_r/\hat{f}_l$	p-value	N
High-tax municipalities									
(a)	120k	5k	0.003	0.003	-0.000	0.054	0.905	0.816	4,562
(b)	120k	6k	0.003	0.003	-0.001	0.185	0.824	0.667	4,562
(c)	120k	7k	0.004	0.003	-0.001	0.365	0.759	0.546	4,562
(d)	120k	24.62k	0.004	0.003	-0.001	0.587	0.819	0.444	4,562
Low-tax municipalities									
(a)	120k	5k	0.002	0.005	0.003	9.029	2.760	0.003	4,827
(b)	120k	6k	0.002	0.005	0.003	9.080	2.463	0.003	4,827
(c)	120k	7k	0.002	0.005	0.003	9.290	2.302	0.002	4,827
(d)	120k	26.04k	0.003	0.004	0.001	2.505	1.291	0.114	4,827

Notes: Local likelihood ratio results for the discontinuity in the income distribution of foreigners not holding the permanent residence permit with a duration of stay below 5 years at the 120k income threshold for several bandwidths. $\hat{\theta}$ stands for the discontinuity estimate ($\hat{\theta} = \hat{f}_r - \hat{f}_l$), $\hat{l}r$ is the value of the empirical likelihood ratio, $\hat{f}_r/\hat{f}_l$ is the relative size of the discontinuity, and N stands for the number of observations with regard to the observations available to estimate the whole density in the sample.
The estimates for the discontinuity in the density are based on the approach in Otsu et al. (2013), which was adjusted for the sampling weights in our data. We multiply the kernel weights by the sampling weights in the summation part of the FOC.