Online Appendix to Immigrant Representation and Local Ethnic Concentration

Karl-Oskar Lindgren, Michael D. Nicholson, and Sven Oskarsson

A. Data Availability

In this paper, we use individual level information obtained from various administrative registers. The data are stored on an encrypted server and all our analysis have been conducted through a remote desktop application. We are under contractual obligation not to disseminate these data to other individuals. Interested researchers can however order the data directly from Statistics Sweden. Currently, Statistics Sweden requires that researchers obtain permission from a Swedish Ethical Review Board before data can be ordered (a description, in Swedish, of how to order data from Statistics Sweden is available at: http://www.scb.se/sv_/Vara-tjanster/Bestalla-mikrodata). We will also make available a complete list of the variables that we ordered from Statistics Sweden for this project.

B. Variables and Data Sources

Nominated – Equal to 1 if the individual ran for office at the municipal level in the six elections held in 1991, 1994, 1998, 2002, 2006, and 2010. Information is retrieved from the Register of Nominated and Elected Candidates held at Statistics Sweden.

Elected – Equal to 1 if the individual was elected for office at the municipal level in the six elections held in 1991, 1994, 1998, 2002, 2006, and 2010. Information is retrieved from the Register of Nominated and Elected Candidates held at Statistics Sweden.

Female – Equal to 1 if female. Information is retrieved from the Swedish Population Register.

Birth Year – Information is retrieved from the Swedish Population Register.

Immigration Year – Information is retrieved from the Swedish Population Register.

Married – Equal to 1 if married at the time of arrival in Sweden. Information is retrieved from the Longitudinal integration database for health insurance and labour market studies (LISA by Swedish acronym).

Native Partner – Equal to 1 if married to or cohabiting with a swede at the time of each of the six elections held between 1991 and 2010. Information is retrieved from the 1991-2010 waves of the Longitudinal integration database for health insurance and labour market studies (LISA by Swedish acronym).

Years of Schooling – Educational attainment at the time of arrival in Sweden or, if information is missing, at the first available year according to the three-digit Swedish standard classification of education (SUN 2000). Following the manual for classifying educational programmes in OECD countries (ISCED-97), we assigned the following years of schooling to each category: (old) primary school (7); (new) compulsory school (9); (old) junior secondary education (9.5); high school (10-12 depending on the program); short university (13); longer university (14-16 depending on the program); short postgraduate (17); long post-graduate (19). The information on educational attainment is retrieved from the 1990-2010 waves of the Longitudinal integration database for health insurance and labour market studies (LISA by Swedish acronym).

Years of Schooling at Election Years – Educational attainment at the time of each of the six elections held between 1991 and 2010 according to the three-digit Swedish standard classification of education (SUN 2000). Following the manual for classifying educational programmes in OECD countries (ISCED-97), we assigned the following years of schooling to each category: (old) primary school (7); (new) compulsory school (9); (old) junior secondary education (9.5); high school (10-12 depending on the program); short university (13); longer university (14-16 depending on the program); short postgraduate (17); long post-graduate (19). The information on educational attainment is retrieved from the 1991-2010 waves of the Longitudinal integration database for health insurance and labour market studies (LISA by Swedish acronym).

Labor Income – Individual monthly labor income at the time of each of the six elections held between 1991 and 2010 (in 1,000 SEK). The variable is retrieved from the 1991-2010 waves of the Longitudinal integration database for health insurance and labour market studies (LISA by Swedish acronym).

Children – Number of children at the time of arrival in Sweden. Information is retrieved from the Longitudinal integration database for health insurance and labour market studies (LISA by Swedish acronym).

Assigned Municipality – Code for the municipality to which the refugee was assigned after receiving a residence permit. Information is retrieved from the Swedish Population Register.

Ethnic Density – The share (in percentage points) of the total population in the assigned municipality belonging to the same country group of origin (see below) as the refugee at the time of arrival. Annual information on the size of the stock of co-ethnics is only available from 1990 and onwards. For the years 1990-1994 we use information from the Swedish Population Register, and the Longitudinal integration database for health insurance and labour market studies (LISA by Swedish acronym). For the years 1987-1989 we instead use information on the size of the stock of co-ethnics from the census in 1985 in combination with information on the inflow of immigrants in the years 1986-1989 from the Swedish Population Register. Thus, in order to estimate the annual stock of co-ethnics across municipalities for the early years we need to rely on the assumption that immigrants residing in Sweden at the time of the census in 1985 did not move or die between 1985 and 1990.

Share Native Colleagues – the share of Swedish born individuals employed at the same plant as the immigrant at the time of each of the six elections held between 1991 and 2010. For individuals without employment this share is set to 0. Information is retrieved from the Longitudinal integration database for health insurance and labour market studies (LISA by Swedish acronym).

Seats-to-Voters Ratio – The ratio of the number of municipality council seats to the number of eligible voters in the municipality within which the individual resides at the time of each of the six elections between 1991 and 2010. The data is retrieved from the Swedish Election Authority (www.val.se).

Country of Origin – Country of birth. Information is retrieved from the Swedish Population Register. For reasons of confidentiality, the country of birth variable has been grouped into 27 distinct groups as described in the table below. For immigrants from significant sending countries (e.g., Iran, Iraq, and Turkey) the region code is that of the country, but for those from other countries the code also includes neighboring countries.

C. Additional Descriptive Statistics

As described in the main text, the information on country of birth has been aggregated into 27 groups for confidentiality reasons. Table A1 provides information on the country groups.

Table A1: The classification of country groups

Code	Country of origin
26	Finland
27	Denmark
28	Norway, Iceland
29	Bosnia-Herzegovina,
30	Yugoslavia, Croatia, Macedonia, Slovenia
31	Poland
32	Ireland, Great Britain
33	Germany, West Germany, East Germany
34	Greece, Italy, Malta, Monaco, Portugal, San Marino, Spain, Vatican City
35	Estonia, Latvia, Lithuania
36	Albania, Armenia, Azerbaijan, Bulgaria, Georgia, Kazakhstan, Kyrgyzstan, Moldavia, Romania, Russia, Soviet Union, Tajikistan, Turkmenistan, Ukraine, Uzbekistan, Belarus
37	Czech Republic, Slovakia, Czechoslovakia, Hungary
38	Andorra, Belgium, France, Liechtenstein, Luxembourg, Netherlands, Switzerland, Austria

39	Canada, USA
40	Antigua, Bahamas, Barbados, Belize, Costa Rica, Cuba, Dominica, the Dominican Republic, El Salvador, Grenada, Guatemala, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, S:t Lucia, St. Vincent, St. Kitts-Nevis
41	Chile
42	Bolivia, Brazil, Colombia, Ecuador, Guayana, Paraguay, Peru, Surinam, Uruguay, Venezuela
43	Djibouti, Eritrea, Ethiopia, Somalia, Sudan
44	Algeria, Bahrain, Cyprus, Egypt, French protectorate in Morocco, United Arab Emirates, Gaza Strip, Israel, Yemen, Jordan, Kuwait, Lebanon, Libya, Morocco, Palestine, Qatar, Saudi Arabia, South Yemen, Syria, Tunisia
45	Angola, United Arab Republic, Benin, Botswana, Burkina Faso, Burundi, Central African Republic, Comoros, Equatorial Guinea, Ivory Coast, Gabon, Ghana, Guinea, Guinea-Bissau, Cameroon, Cape Verde, Kenya, Democratic Republic of the Congo, Lesotho, Liberia, Madagascar, Malawi, Mauretania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, Sao Tomé and Príncipe, Senegal, Seychelles, Sierra Leone, Swaziland, South Africa, Tanzania, Chad, Togo, Uganda, Zaire, Zambia,

	Zanzibar, Zimbabwe
46	Iran
47	Iraq
48	Turkey
49	Hong Kong, Japan, China, Taiwan, North Korea, South Korea
50	Burma, Philippines, Indonesia, Laos, Federation of Malaya, Malaysia, Singapore, Thailand, Vietnam
51	Afghanistan, Bangladesh, Bhutan, Brunei, India, Democratic Kampuchea, Maldives, Mongolia, Nepal, Oman, Pakistan, Sikkim, Sri Lanka
52	Australia, Fiji, Kiribati, Micronesia, Nauru, Palau, Papua New Guinea, Solomon Islands, Tonga, Vanuatu, Samoan Islands
53	Unknown

Table A2 provides information on the number of individuals in our main sample that are born in each of these country groups. Our estimation sample includes 62,230 unique individuals.

Table A2: Individuals by country groups in the main sample

Country group	Individuals	Share
30	2453	3.94
31	2014	3.24
35	192	0.31
36	4175	6.71
37	1431	2.30
40	959	1.54
41	6326	10.17
42	1605	2.58
43	6475	10.40
44	9296	14.94
45	1325	2.13
46	15241	24.49
47	4222	6.78
48	2044	3.28
50	2805	4.51
51	1667	2.68
Total	62230	100

Note: The frequencies refer to the individuals included in the main estimation sample.

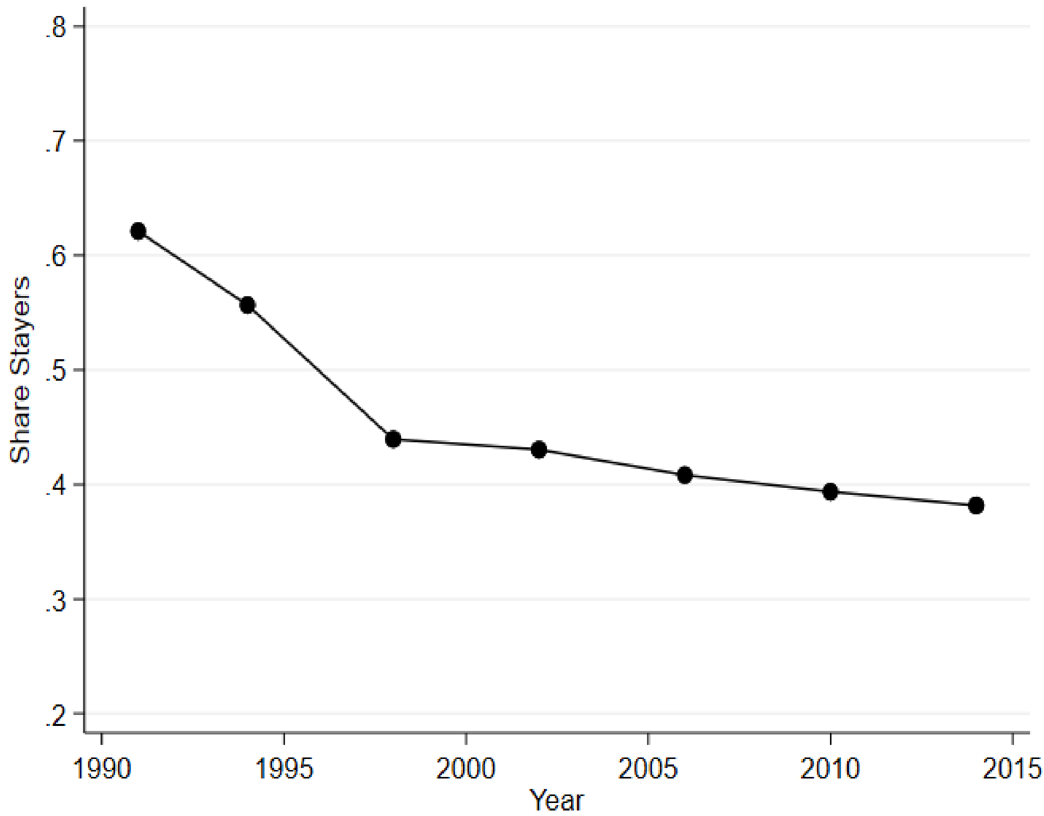

Figure A1: Share stayers by election year

In Figure A1 we display the share of the individuals who remained in their assigned municipality in a specific year. That is, among the individuals who were placed in a municipality between 1987 and 1991 about 60 percent were still living in that municipality in 1991. In 2014 the corresponding figure is just below 40 percent, i.e., after 25 years as many as 4 out of 10 immigrants still lived in their municipality of assignment.

However, another interesting question is to what extent those moving out of their assigned municipality differs systematically from those remaining in their initial municipality. In Table A3 we provide separate descriptive statistics for the movers and stayers in our sample for the year 2014. In this table all individuals who remain in their municipality of assignment in 2014 are classified as stayers, whereas all those who live in a different municipality in 2014 are classified as movers. Although there are some differences between the two groups, most of the differences appear rather marginal compared to the standard deviations of these variables (but gender is a notable exception).

Table A3: Descriptive statistics for Movers and Stayers in 2014

	All	Stayers	Movers
Age	55.51	56.58	54.85
	(8.01)	(8.48)	(7.64)
Female	0.43	0.48	0.40
	(0.50)	(0.50)	(0.49)
Married	0.59	0.61	0.58
	(0.49)	(0.49)	(0.49)
Nr. of children under 16	1.07	1.10	1.05
	(1.34)	(1.31)	(1.35)
Years of education	10.79	10.59	10.91
	(2.73)	(2.77)	(2.70)
Ethnic density (log)	-5.72	-5.50	-5.85
	(0.95)	(0.90)	(0.95)
Nominated (%)	0.79	0.80	0.76
	(8.77)	(8.92)	(8.68)
Elected (%)	0.19	0.18	0.19
	(4.31)	(4.27)	(4.33)
Age at immigration	30.31	31.36	29.67
	(7.93)	(8.36)	(7.57)
Immigration year	1989.07	1989.05	1989.08
	(1.35)	(1.35)	(1.35)
Observations	51562	19683	31879

Note: All data in this table refers to the year 2014 and movers are those for who the municipality of residence in 2014 is not the same as the municipality of placement.

D. Additional Analyses

D.1. Examining Linearity

In our main analysis we have used 2SLS with a continuous instrument, which means that we have implicitly invoked the assumptions that the treatment effect of interest is (conditionally) linear. In this section we will focus on the assumption of linearity.

The best way to assess the assumption of linearity is often by means of graphical inspection. We will therefore provide simple (partial) regression plots that can be used to determine whether the first-stage and the reduced form relationships appear approximately linear. All graphs are based on a second degree polynomial regression, and to ease visualization the graphs have been trimmed at 1 and 99 percentiles of the ethnic concentration variable.

Starting with the first-stage relationship, the leftmost graph in Figure A2 displays the unconditional bivariate relationship between (logged) ethnic concentration in the year of immigration and (logged) ethnic concentration in the later election years, whereas the rightmost graph show the residualized instrument-treatment relationship. The latter graph thus shows the strength of the first-stage relationship once all covariates and fixed effects included in our main specification have been incorporated into the model.

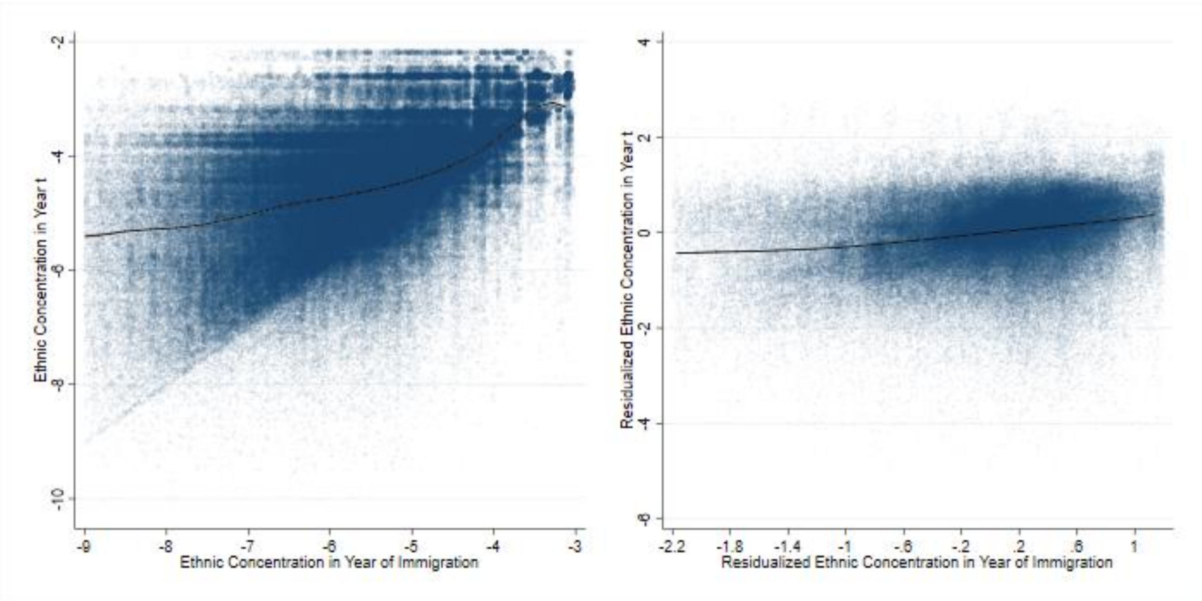

Figure A2: Local polynomial graph for the first-stage relationship, pooled model.

As can be seen, there are no strong signs of non-linearities in these graphs. It is particularly comforting to note that the residualized relationship appears to be fairly linear since it is on this model our main analysis is based. The reasonableness of the linear approximation for the first-stage equation is further corroborated by figures A3 and A4, which display the bivariate and residualized first-stage relationships for each election year separately.

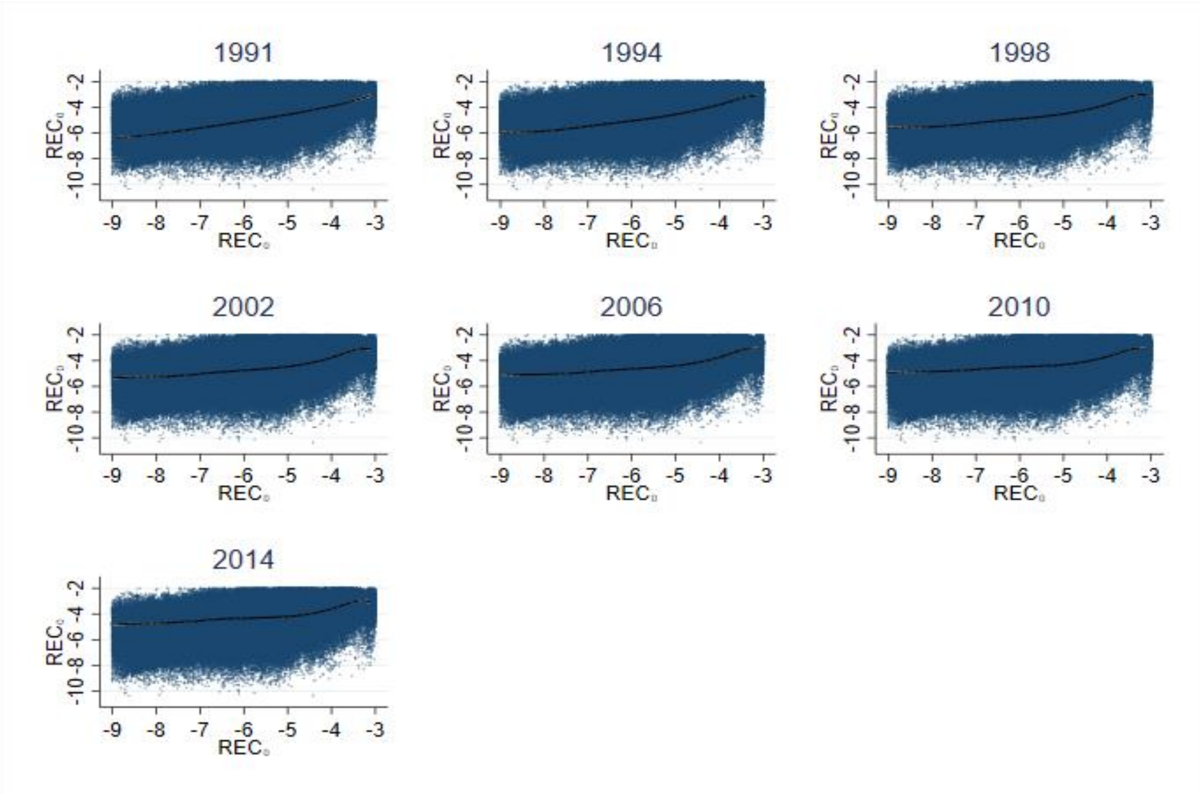

Figure A3: Local polynomial graph for the bivariate first-stage relationship by year.

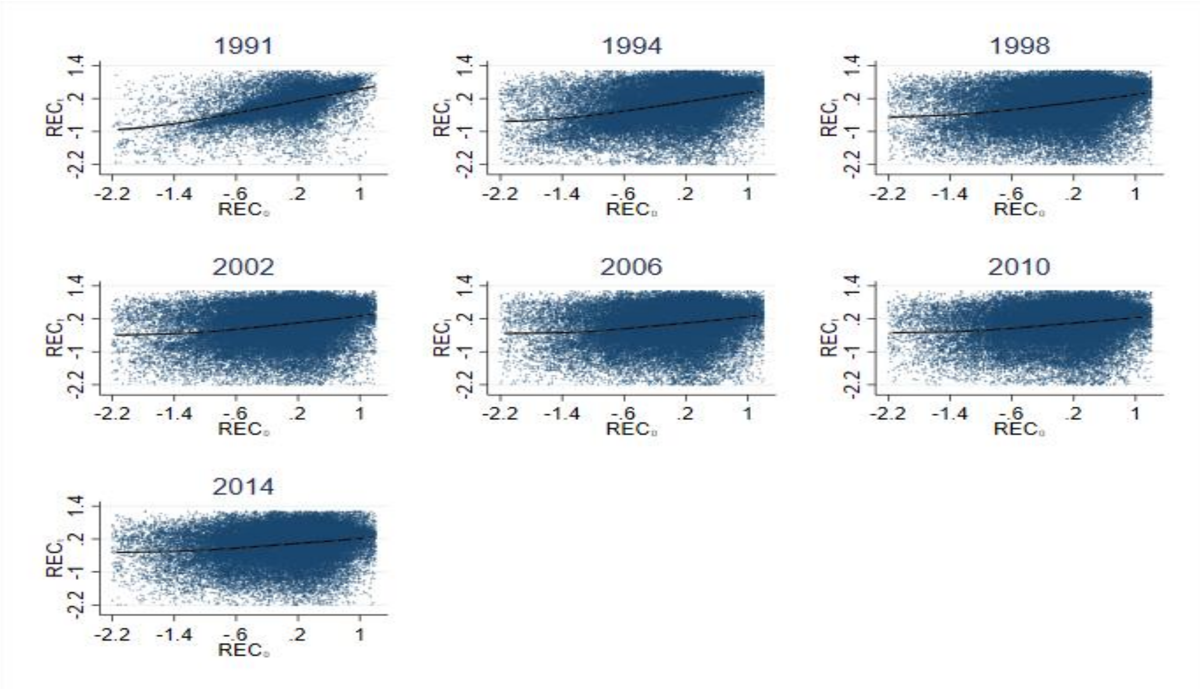

Figure A4: Local polynomial graph for the residualized first-stage relationship by year.

Turning to the reduced form relationship, the leftmost graph in Figure A5 displays the bivariate reduced form relationship and the rightmost graph the corresponding residualized relationship.

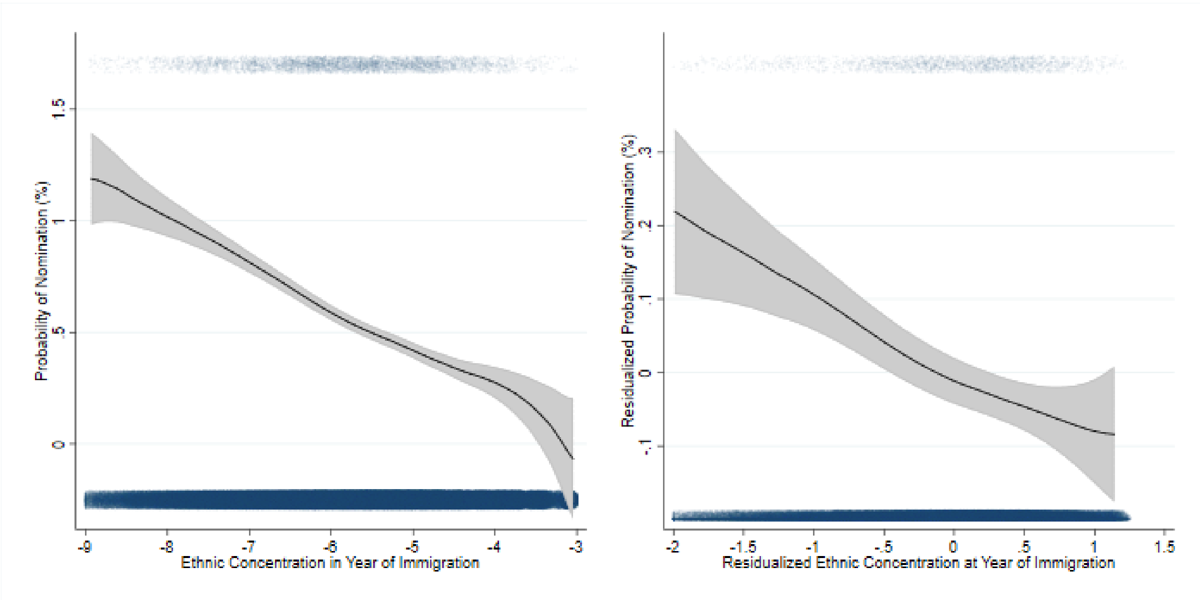

Figure A5: Local polynomial graph for the bivariate reduced form relationship

Although there is a slight kink in the residualized relationship for high levels of ethnic concentration, it nevertheless seems reasonable to use a linear approximation for the reduced form relationship. Figures A6 and A7 display the bivariate and residualized reduced form relationships for each election year separately. Admittedly, in some years the residualized reduced form relationship appears somewhat non-linear, but there is no clear pattern to this non-linearity. Moreover, it should be remembered that the number of nominated individuals in a given year is very small so running our full model for each year separately is very taxing on the data. Overall, we therefore believe that these analyses support the reasonableness of the linearity assumptions underlying our preferred specifications.

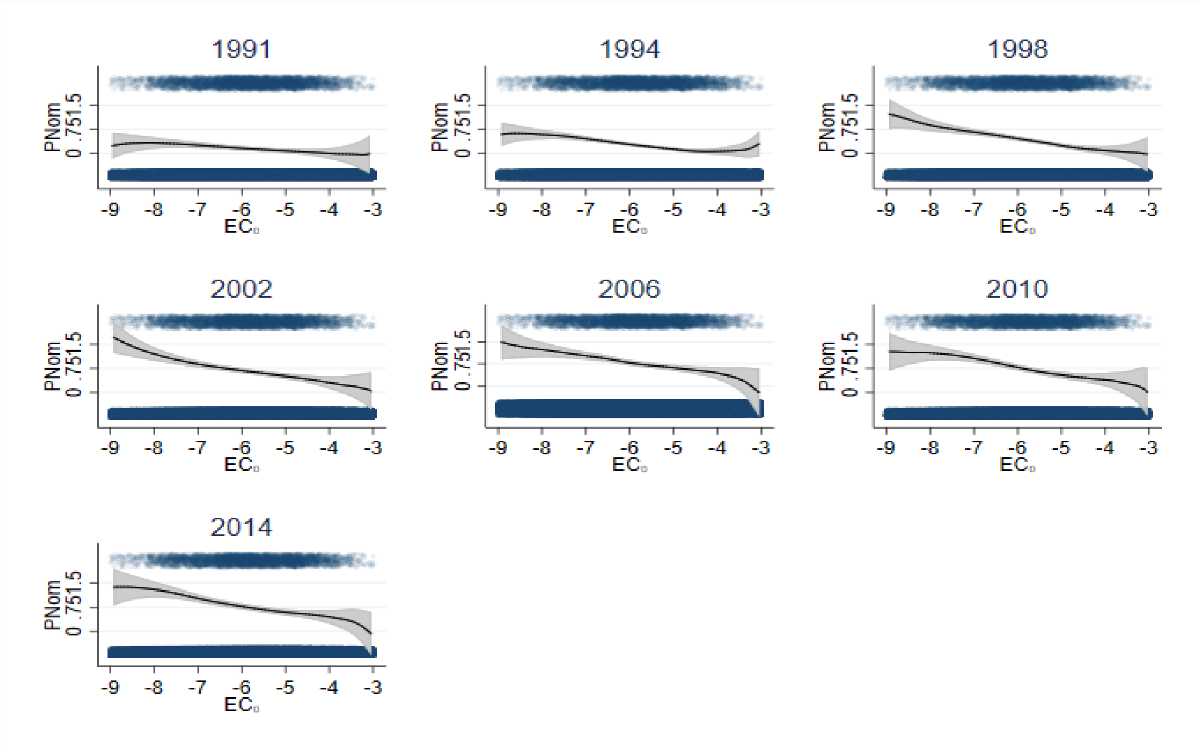

Figure A6: Local polynomial graph for the reduced form relationship by year

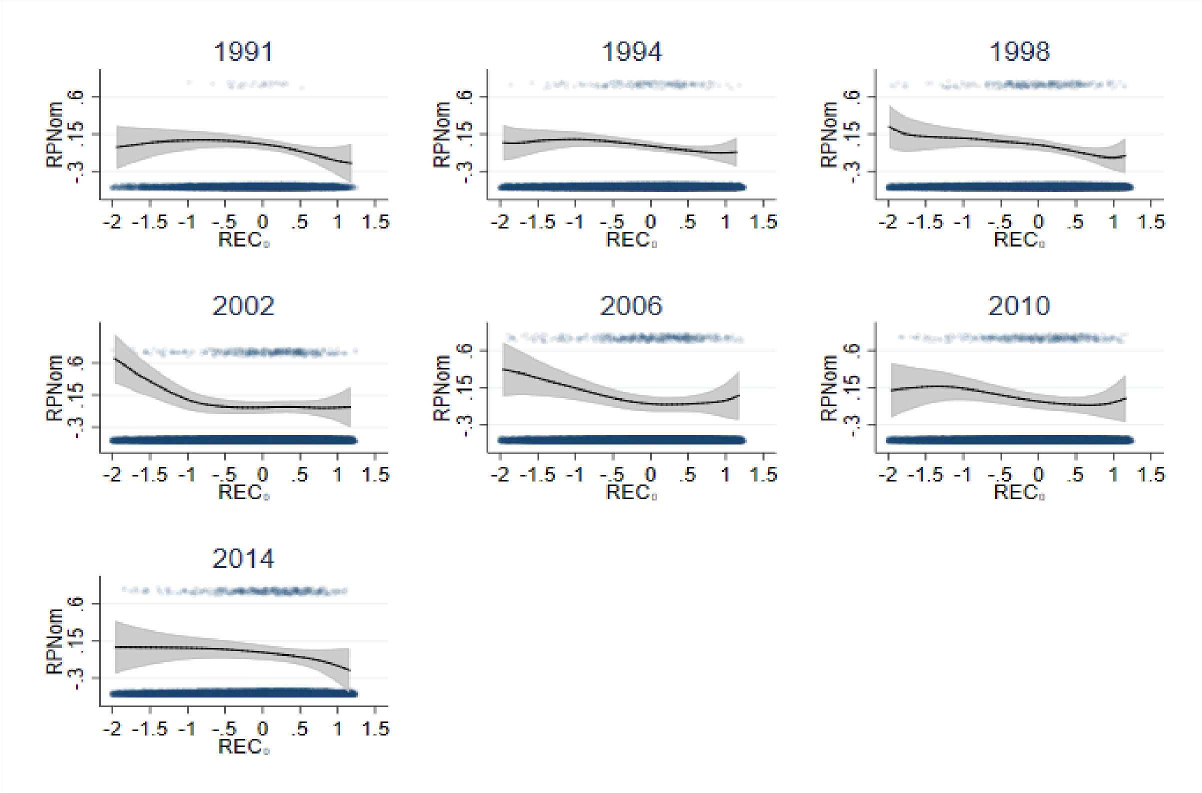

Figure A7: Local polynomial graph for the residualized reduced form relationship by year

D.2. Profile of the compliers

In causal analyses, using the potential outcome framework it is standard to partition the overall sample into three different groups named compliers, never-takers, and always-takers. It can be shown that the instrumental variable approach will identify the treatment effect among the subgroup of compliers, i.e., those who abide by the treatment assignment so that they are in the treatment group when assigned to the treatment and in the control group when not assigned to the treatment (e.g., Angrist and Pischke 2009). When considering to what extent the instrumental variable results can be generalized to the population at large, it can therefore be useful to examine whether the group of compliers are systematically different from the overall sample in any important respects.

To this end, we utilize a simple profiling approach developed by Marbach and Hangartner (2019). This method uses information on instrument and treatment values to identify the different subgroups and describe these groups by means of simple averages and confidence intervals for various observed characteristics. Unfortunately, the method presupposes that both the treatment and instrument are binary. We have therefore dichotomized the treatment and instrument indicators by splitting each variable at the sample median.

Figure A8 shows the means and confidence intervals for seven different demographic and social characteristics, all measured at the time of immigration. The estimated share of compliers is about 22%, whereas the never-takers and always takers each make up about 39% of the sample. Compared to the sample as a whole, we see that the subgroups of compliers have immigrated at slightly higher age, are more likely to be female and have children, and are somewhat less educated. With respect to the three remaining variables–year of immigration, municipality size, and marital status–the covariate means for compliers come very close to the overall sample means.

The question then becomes how to interpret these results. Can the LATE estimated for compliers be generalized to the study population at large? This question can only be answered by additional research using instruments affecting other groups. However, in absolute terms, the differences found in Figure A8 are relatively minor so we do not believe they will have any huge importance, although some care should be taken when attempting to generalize the effect to the group of highly educated men, for instance.

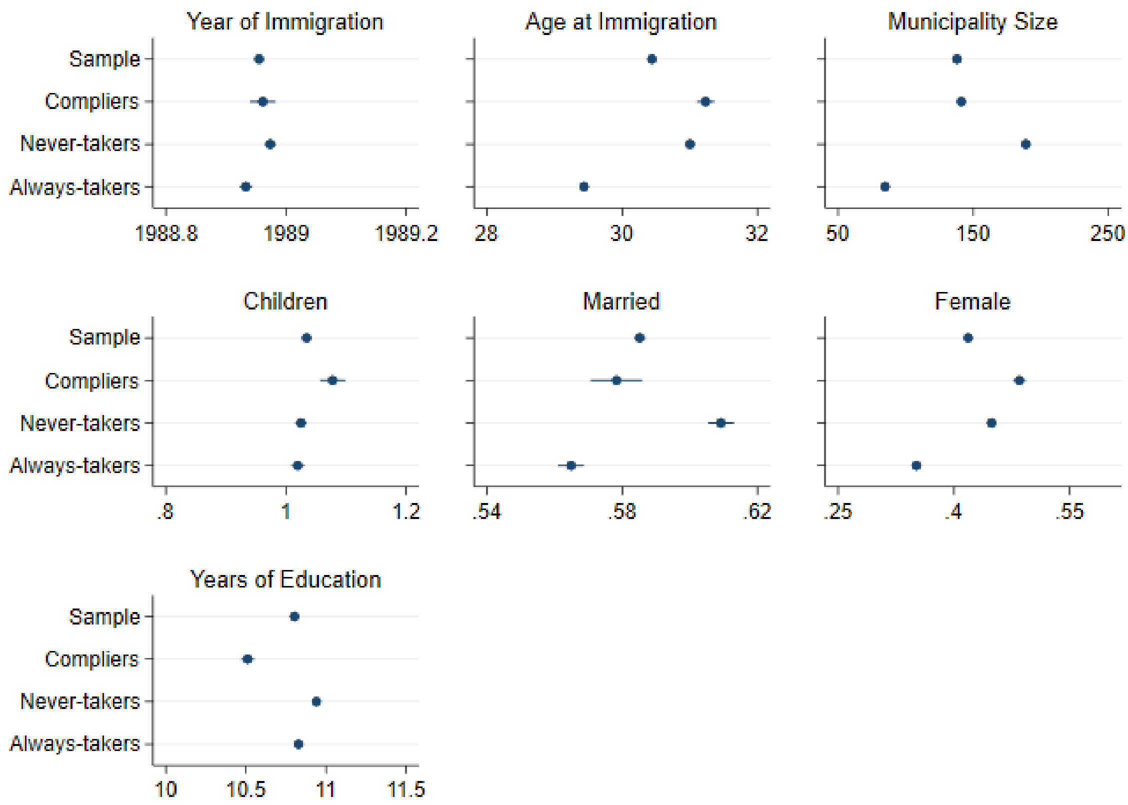

Figure A8: Descriptive statistics (mean and 95% bootstrap intervals) for the complier and non-complier subpopulations.

D.3. Alternative measures and methods

To judge from the results presented in the main text, ethnic residential segregation seems to have a negative effect on immigrants' likelihood of nomination. In this section we examine the robustness of these findings. Table A4 presents results from models in which we alternate the baseline specification in a number of different ways.

Table A4. Alternative specifications.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	2SLS	2SLS	IV-Probit	2SLS	IV-Probit
Ethnic density	-0.425**	-0.752**	-0.235***	-0.054	-0.125**
	(0.186)	(0.311)	(0.086)	(0.047)	(0.056)
Female	-0.192***	-0.163**	-0.129***	-0.167***	-0.128***
	(0.049)	(0.065)	(0.027)	(0.040)	(0.026)
Married	0.185**	0.237***	0.098***	0.162***	0.098***
	(0.069)	(0.081)	(0.037)	(0.056)	(0.037)
Education	0.110***	0.085***	0.062***	0.099***	0.063***
	(0.012)	(0.015)	(0.006)	(0.012)	(0.005)
Children	-0.005	-0.038	0.001	-0.010	-0.001
	(0.024)	(0.035)	(0.013)	(0.020)	(0.013)
First stage	0.252***	0.268***	0.268***	1.001***	0.994***
	(0.020)	(0.026)	(0.023)	(0.087)	(0.092)
Observations	281,286	172,683	327,466	349,383	327,466

Note: All models include fixed effects for election year, year of birth, country of origin, year of immigration, and assigned municipality. Standard errors are shown in parentheses and allow for clustering within assigned municipalities. ***/**/* indicate significance at the 1/5/10% level. In models 1-4 Ethnic density has been log transformed, whereas the unlogged version of the variable is used in models 5-6. The probit models are estimated by the ivprobit command in Stata, and the coefficients for these models are probit coefficients.

In Column (1) of Table 4, we examine to what extent the results are unduly driven by immigration to the large urban areas by excluding the individuals that were placed in the three big cities Stockholm, Gothenburg, and Malmö. Although this means that we lose almost a quarter of our original sample, the substantive results remain very similar.

A potential drawback with our data is that immigrants from small source countries have been grouped together with immigrants from neighboring countries for reasons of confidentiality. To examine whether this poses a problem, we have re-run the model including only immigrants from countries that we can identify uniquely. The results are presented in Column (2) and, if anything, the negative effect of ethnic density on political candidacy becomes even more pronounced when restricting the analysis to immigrants from individual source countries. According to these results, increasing ethnic density by one log unit can be expected to decrease the probability of nomination by as much as 0.7 percentage points.

Throughout this paper, we have employed a linear probability model, adjusting for heteroscedasticity in the error terms. We did so because instrumental variable estimation is much more involved, and requires additional strong model assumptions, when applied to non-linear models such as logit or probit. Nonetheless, in Column (3) of Table A4 we present the results from the instrumental variable probit model discussed by Newey (1987). An advantage with this model is that we can take the binary nature of our dependent variable into account. However, a disadvantage is that we now have to assume that initial ethnic density is not only *a* valid instrument for ethnic density in later years, but that it is the *only* relevant instrument (Lewbel et al. 2012).

This said, it is reassuring to note that the substantive results of the probit model are very similar to those of the linear probability model. The probit coefficients are not directly comparable to those of the linear probability model but as can be seen they all have the same sign and statistical significance. Moreover, if we compute the average marginal effect for ethnic density in model 3 it is .32, which is very close to the linear probability results.

Finally, in the main analyses we log transform the ethnic density variable. The reasons for taking the log of this ratio are theoretical as well as methodological. Theoretically, as argued by Bertrand et al. (2000), even if a small ethnic group was fully concentrated in a single municipality, the group would never constitute a large fraction of the population in that municipality. Yet, because individuals tend to self-segregate into social networks rather than match randomly, we could still imagine that members of a small ethnic community spend considerable time with their co-ethnics even if that group makes up only a very small proportion of the municipality's population. From a theoretical standpoint, using logarithmic shares is preferable since it prevents us from underweighting small ethnic groups in the analysis.

Methodologically, as shown by Gerdes (2011), it is usually preferable to work with logarithmic, rather than actual, shares when estimating fixed effects models. This is because, when using actual shares, the observations are implicitly weighted by the denominators used to calculate the shares, which implies scope for spurious correlation between the shares and the dependent variable. Moreover, the logarithmic transformation considerably reduces the skewness of the ethnic density variable.[1]

Yet, it might nevertheless be interesting to examine what happens to the results if we use actual shares instead of log shares in the model. This we do in columns (4) and (5). Starting with the linear probability specification in Column (4) we see that the coefficient of the ethnic density variable remains negative, but it is no longer statistically significant at conventional levels.

However, when using the instrumental variable probit model, the ethnic density effect increases in magnitude and regains statistical significance, as indicated by the results reported in Column (5). One likely reason for the greater difference between the probit and linear probability results in this case is that the probit model implicitly accounts for the decreasing marginal effect of ethnic share on the probability of nomination. These results could thus be taken as support for our choice to use the logarithmic transformation of ethnic density in the main analysis. The fact that the reduced form relationship appears approximately linear when log transforming ethnic density further corroborates the reasonableness of this approach (see Figure A5).

We have also performed some additional analyses to make sure that our results are not driven by functional form bias. These results are presented in Table A5. The first two columns of Table A5 shows the results from saturated regression models where the specifications include a separate parameter for all possible values taken on by the explanatory variables (Angrist and Pischke 2009:48). That is, in the first model we construct a variable that contains all unique combinations of election year, year of immigration, year of birth, and country of origin and add a full set of fixed effects for this new variable to the model.[2] In the second column we take this approach one step further by generating all unique combinations of the variables just mentioned *and* sex, marital status, education, and the number of children.

[1] It should also be noted that because our models include assigned municipality fixed effects, our measure is observationally equivalent to using the logged size of the ethnic group in a municipality as done by Edin et al. (2003).

[2] Assigned municipality is not included in this set because then there would not be any remaining "within-cell" variation in ethnic density. Instead we add assigned municipality fixed effects to the models in the same way as we have done in the previous models.

Table A5. Saturated controls and binary treatment

	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic density	-0.433***	-0.602**	-1.206**	-1.113**	-1.300**	-0.932
	(0.166)	(0.265)	(0.545)	(0.542)	(0.557)	(0.667)
Treatment	Continuous	Continuous	Binary	Binary	Binary	Binary
Ext. controls	No	Yes	No	Yes	No	Yes
Saturated controls	Yes	Yes	No	No	Yes	Yes
Observations	347,800	197,965	349,383	349,383	347,800	197,965

Note: All models include controls for election year, year of birth, country of origin, year of immigration, and assigned municipality, the models with extend controls also include sex, marital status, education, and the number of children in the set of control variables. In the saturated models the models include controls for all unique combinations of the covariates, except assigned municipality for which we add separate fixed effects. In the non-saturated models we control for the covariates in the same way as we do in the main analyses. ***/**/* indicate significance at the 1/5/10% level.

The advantage of the saturated regression model is that we do not have to make any assumptions about the joint distribution of our right-hand side variables. A disadvantage, however, is that we lose data since many observations will have a unique combination of the explanatory variables, and then they cannot be used for estimation. This problem becomes more aggravated as the number of explanatory variables increases, which is also evident from the drop in observations from columns 1 and 2.

As can be seen from the two first columns of the table, the negative effect of ethnic density on political candidacy remains also when we control for the explanatory variables in a more flexible way. If anything, the negative effect appears somewhat stronger in these more flexible specifications. However, the strengthening of the coefficient in Column 2 is mainly driven by the change in sample, not by the flexibility of the controls. If we use our main specification to estimate the effect of ethnic density in this subsample we obtain a coefficient of -0.54.

As an alternative means to examine whether the results could be affected by functional form bias, we have run models where both the instrument and treatment have been dichotomized by splitting each variable at the sample median. That is, we now only differentiate between high (above median) and low (below median) ethnic concentration.

Columns 3 and 4 of Table A5 shows the 2SLS results when both the instrument and the treatment variable have been coded as dummy variables. As can be seen the results are well in line with those presented in the main text. To judge from these estimates an individual is about 1 percentage point less likely to pursue candidacy if he or she lives in a municipality with high ethnic concentration.

For reasons of completeness columns 5 and 6 of the table report the results from models where the saturated specifications of columns 1 and 2 are combined with the binary instrument/treatment. Although, the coefficient of the binary treatment is no longer statistically significant in the last column, due to the decreased precision when losing a large share of the observations, the point estimates from the saturated models are very similar to that of our main specification.

So far, the analysis has abstracted away from the surrounding political context. The main justification for this is that our identification strategy rests on the comparison of individuals initially placed within the same political context (i.e., municipality). That is, we believe that the municipality fixed effects included in the models will soak up most of the impact of the political opportunity structure variables commonly believed to foster immigrant political representation (e.g., Bird 2011; Dancyiger et al. 2015). Moreover, by not conditioning our analyses on any election-year municipality characteristics we have also been able to circumvent the risk of controlling for variables that could be an effect rather than a cause of ethnic concentration.

This being said, it may still be interesting to examine to what extent controlling for the political context affects our results. In Table A6 we do this by using the political opportunity structure variables discussed and analyzed by Dancygier et al. (2015).

Table A6. Results when controlling for political variables.

	(1)	(2)	(3)	(4)
Ethnic density	-0.373**	-0.365**	-0.306*	-0.318**
	(0.160)	(0.165)	(0.171)	(0.133)
Seats to voters	329.897***	330.480***		
	(46.364)	(60.684)		
Effective nr. of parties		-0.051		
		(0.045)		
Disproportionality		0.017		
		(0.024)		
Left share		1.621**		
		(0.632)		
Immigrant share		-0.927		
		(0.979)		
Ethnic fractionalization		0.789		
		(0.838)		
Native education		-0.022		
		(0.081)		
Mun. FE year-0 (FE0)	Yes	Yes	Yes	Yes
Mun. FE year-t (FET)	No	No	Yes	No
FE0 x FET	No	No	No	Yes
Observations	349383	349383	349383	349383

In the first column of the table we add the seats-to-voters ratio to our main specification. By controlling for this variable, we adjust for the strong and mechanically negative effect between the size of the electorate and the probability of running for office, which is due to the fact that the size of local assemblies does not increase proportionally to the size of the electorate (Dancygier et al. 2015). However, as can be seen the ethnic density variable is hardly affected at all by the inclusion of this variable.

In Column (2) we add a larger set of political opportunity structure variables to the model, including the effective number of parties[3], vote-seat disproportionality[4], left-party strength,

[3] This measure is defined as $(\sum s_i^2)^{-1}$, where s_i is the seat share of party i.
[4] To measure this we use the Gallagher index, i.e., $G = \sqrt{.5(\sum(v_i - s_i)^2)}$, where v_i and s_i indicates votes and seat shares of party i, respectively.

immigrant share, overall ethnic fractionalization[5], and average educational attainment among the natives in the municipality. The most important thing to note is that controlling for the political opportunity structure variables does not affect the impact of ethnic density. Second, the only opportunity structure variable, besides the seats-to-voters variable, that is clearly related to the likelihood of being nominated to political office is left-party strength. That is, immigrants are more likely to become nominated as the support for leftist parties increases.

An alternative means to capture the importance of the political context is to add election-year municipality fixed effects to the model. This is done in Column (3) and as can be seen the coefficient of ethnic density decreases by about one fifth and is no longer statistically significant when adding election-year municipality fixed effects to the model.

However, in Column (4) we instead create fixed effects for all unique combinations of initial (assigned) municipality and election-year municipality. That is, we now only compare individuals who were assigned to the same municipality at year 0 *and* reside within the same municipality in year *t*, either because they have stayed in their assigned municipality or because they have moved to the same municipality by time *t*. Consequently, this is a more demanding specification than that used in Column (3). As can be seen, we now again find a statistically significant negative effect of ethnic density on the probability of nomination. The overall message of the results presented in Table A6 is thus that we obtain very similar results also when controlling for the, potentially endogenous, political opportunity structure facing the immigrants.

D.4. Heterogeneity analyses

In Figure 3 of the main text, we present the results from a set of heterogeneity analyses graphically. Table A7 contains the point estimates and standard errors used to construct this graph.

[5] Overall ethnic fractionalization is measured as the inverse of the Herfindahl index, i.e., $1 - \sum d_i^2$, where d_i is the share of the immigrant group coming from region *i*.

Table A7. Effect of Ethnic Concentration by Political Context.

	Left party strength		Right pop. strength		Disproportionality	
	Low	**High**	**Low**	**High**	**Low**	**High**
Ethnic density	-0.426*	-0.367*	-0.444*	-0.213	-0.478**	-0.120
	(0.222)	(0.212)	(0.218)	(0.216)	(0.213)	(0.172)
Observations	176,209	173,174	143,782	185,268	174,792	174,589

Note: All models include fixed effects for election year, year of birth, country of origin, year of immigration, and assigned municipality, as well as controls for education, marital status, sex, and the number of children. Standard errors are shown in parentheses and allow for clustering within assigned municipalities. ***/**/* indicate significance at the 1/5/10% level.

To further examine potential heterogeneities in the data, Table A8 presents results separated by sex, education (at time of immigration), and citizenship status. For education, individuals with at least 11 years of education are coded as highly educated.

Table A8. Effect of Ethnic Concentration by Individual Characteristics.

	Sex		Education		Swedish citizen	
	Male	**Female**	**Low**	**High**	**No**	**Yes**
Ethnic density	-0.310	-0.394**	-0.346*	-0.458*	-0.224*	-0.495**
	(0.211)	(0.190)	(0.187)	(0.236)	(0.127)	(0.248)
Observations	203217	146166	143544	205838	119509	229874

Note: All models include fixed effects for election year, year of birth, country of origin, year of immigration, and assigned municipality, as well as controls for education, marital status, sex, and the number of children. Standard errors are shown in parentheses and allow for clustering within assigned municipalities. ***/**/* indicate significance at the 1/5/10% level.

As can be seen from the table, we find slightly more negative coefficients of ethnic density for females, highly educated, and Swedish citizens. However, the results for the different subgroups are rather imprecisely estimated and none of the differences reach conventional levels of statistical significance.

D.5. Ethnic concentration and candidate performance

To judge from the main results, the likelihood for immigrants to be nominated to political office decreases as ethnic concentration increases. However, we may also be interested in how ethnic concentration affects the qualifications or performance of the immigrants that are nominated to political office.

In Table A9 we therefore restrict attention to the subset of nominees and study four different "performance" indicators. In the first column we examine the relationship between ethnic concentration and educational attainment among the candidates, which can be considered a rough proxy for a candidate's formal qualifications. The remaining three indicators are of more direct political nature and indicate whether a candidate was elected (column 2), the list position of the candidate (column 3), and the number of preference votes received by the candidate.[6]

Table A9. Effect of Ethnic Concentration Candidate Performance.

	Education	**Elected**	**List position**	**Pref. votes**
Ethnic density	0.099	9.960	-1.081	12.590
	(0.619)	(6.848)	(2.897)	(11.766)
Observations	2008	2008	1885	1822

Note: All models include fixed effects for election year, year of birth, country of origin, year of immigration, and assigned municipality, as well as controls for education, marital status, sex, and the number of children. Standard errors are shown in parentheses and allow for clustering within assigned municipalities. ***/**/* indicate significance at the 1/5/10% level.

Admittedly, the coefficients presented in Table A9 are very imprecisely estimated due to the small number of nominated individuals. The results should therefore be interpreted with great care. With that said, we find some indication of a positive association between ethnic concentration and candidate qualifications/performance. On average, the immigrant candidates coming from municipalities with high ethnic concentration are somewhat more educated, more likely to be elected, are placed further up on the party lists, and receive slightly more preference votes. However, none of the coefficients reach conventional levels

[6] Information on the latter two variables are only available for a subset of the elections.

of statistical significance and more research is clearly needed before we can draw any firm conclusion about the impact of ethnic concentration on candidate qualifications/performance.

REFERENCES

Angrist D., and Pischke J.-S. 2009. *Mostly harmless econometrics*. Princeton university Press: Princeton.

Bertrand, M., E. Luttmer and S. Mullainathan. 2010. "Network effects and welfare cultures." *Quarterly Journal of Economics* 115(3):1019–1055.

Bird,K.,T. Saalfeld,and A. Wüst, eds. 2011. *The Political Representation of Immigrants and Minorities: Voters, Parties and Parliaments in Liberal Democracies*. New York: Routledge.

Edin, PA., P. Fredriksson and O. Åslund. 2003. "Ethnic enclaves and the economic success of immigrants." *Quarterly Journal of Economics* 118(1):329–357.

Gerdes, C. 2011. "Using "shares" vs. "log of shares" in fixed-effects estimations." *Journal of Economics and Econometrics* 54(1):1–6.

Lewbel, A., Y. Dong and T. T.Yang. 2012, "Comparing features of convenient estimators for binary choice models with endogenous regressors." *Canadian Journal of Economics* 45(3): 809–829.

Marbach, M., and D. Hangartner. 2019. "Profiling compliers and non-compliers for instrumental variable analysis", unpublished manuscript, https://dx.doi.org/10.2139/ssrn.3380247.

Newey, W. 1987. "Efficient estimation of limited dependent variable models with endogenous explanatory variables," *Journal of Econometrics* 36(3):231–250.