

ECO 102: Topics in Economics

Ecole Polytechnique, Spring 2022

Professors: Geoffroy Barrows, Benoit Schmutz

Teaching Assistants: Arnault Chatelain, Maddalena Conte

TD 5 (part 2): Migrant segregation

In this TD we will analyse migrant segregation in France. As usual, download the folder TD_2_5.zip, write your answers directly in the Tex script and type your Stata commands in a do file.

Exercise

In this TD, we try to understand why non-European immigrants tend to live more in public housing, based on the French Household Survey.

1. Prepare the dataset: load the dataset and (following the help dofile) create:
 - *non_european*: a variable indicating the origin
 - *deprived*: a variable indicating if the individual lives in a "sensitive urban area"
 - various control variables
2. Regress *deprived* on *non_european*. What is the unconditional effect of being a non-European migrant on living in a deprived area? Add household characteristics (*many_kids*, *employed*, *female*, *married_couple*, *low_diploma*, *high_occupation*) as additional independent variables. What is the conditional effect? Export the results.

$$deprived = \alpha + \beta * non_european + \varepsilon$$

$$deprived = \alpha + \beta * non_european + c * dummy_variables$$

3. Which individual characteristic makes immigrants more likely to live in deprived areas? Regress the variable *deprived* on all individual characteristics (*many_kids*, *employed*, *female*, *married_couple*, *low_diploma*, *high_occupation*), interacting each of these variables with *non_european*, one at the time, (in 6 separate regressions). What do you see?

$$deprived = \alpha + \beta_1 * non_european + \varepsilon + \beta_2 * many_kids + \beta_3 * non_euro \times manykids$$

We find that the variable of low diploma and many kids have a huge effect on the possibilities of living in the deprived area. For the low diploma the coefficients of regression increased by 0.08 and 0.16 for Europeans and Non-Europeans. For many_kids variable we have increased by 0.11 and 0.17 for Europeans and Non-Europeans. Other variables do not have significant effect on the possibilities of living in the deprived area.

| | (1) | (2) |
|-----------------|-------------------|--------------------|
| | deprived | deprived |
| non_european | 0.28*** (0.01) | 0.27*** (0.01) |
| many_kids | | 0.12*** (0.01) |
| employed | | -0.01** (0.01) |
| female | | 0.04*** (0.01) |
| married_couple | | -0.06*** (0.00) |
| low_diploma | | 0.09*** (0.01) |
| high_occupation | | -0.05*** (0.01) |
| Constant | 0.17*** (0.00) | 0.14*** (0.01) |
| # Observations | 27137 | 27137 |
| R squared | 0.020 | 0.058 |
| Mean Dep. Var | | |

Table 1: The regressions table of the unconditional and additional variables

4. What is the role of push and pull factors at the local level? What role does public housing play?

- (a) Create a location identifier at the city-size-region level (use the variables *tu2010* and *rg* and the Stata function *group*). Run the same regression as in part (2) using household characteristics, but now add the location fixed effects (using the Stata function *reghdfe* and the option *absorb*). How does the conditional effect change? What does this suggest?

With location fixed effects, the effect of non european is smaller, This means that if we considered european residents and non european residents who live in a same city scale, the gap of the living condition between these two groups are smaller.

- (b) Generate a variable *hlm* to identify individuals living in social housing (*lsy1* == "1"). Calculate the fraction of non-Europeans living in social housing (note: we subtract each individual from this fraction). Calculate the fraction of all individuals living in social housing. Calculate the fraction of non-Europeans not living in social housing (see code in Stata help do file).
- (c) Calculate pull factor to understand whether non-European migrants are more attracted to places with a higher concentration of non-European migrants. Calculate push factor to understand if more limited access to public housing reduces the presence of non-European migrants in distressed areas (see code in Stata help do file).

| | (1) | (2) | (3) |
|-----------------|--------------------|--------------------|--------------------|
| | deprived | deprived | deprived |
| non_european | 0.20*** (0.01) | 0.20*** (0.01) | 0.20*** (0.01) |
| many_kids | 0.09*** (0.01) | 0.09*** (0.01) | 0.09*** (0.01) |
| employed | 0.00 (0.00) | 0.00 (0.00) | 0.00 (0.00) |
| female | 0.02*** (0.00) | 0.02*** (0.00) | 0.02*** (0.00) |
| married_couple | -0.03*** (0.00) | -0.03*** (0.00) | -0.03*** (0.00) |
| low_diploma | 0.10*** (0.00) | 0.10*** (0.00) | 0.10*** (0.00) |
| high_occupation | -0.06*** (0.01) | -0.06*** (0.01) | -0.06*** (0.01) |
| Constant | 0.13*** (0.01) | 0.13*** (0.01) | 0.13*** (0.01) |
| # Observations | 27137 | 27137 | 27137 |
| R squared | 0.200 | 0.200 | 0.200 |
| Mean Dep. Var | | | |

Table 2: Caption

- (d) Run the same regression as in part (4a), without fixed effects and adding the following independent variables: *inter_1_non_european*, *inter_1_other*, *frac_hlm*, *frac_non_european*. What do you see? Run the same regression adding the following independent variables: *inter_2_non_european*, *inter_2_other*. What do you see? Export results of the 3 regressions in this section.

inter_1_non_european: Pull factor: Non=european living attracted to more non-european migrants,
inter_1_other_european: European more attracted to living with non-european migrants
inter_2_non_europeanL: Push factor: if non-europeans don't live in public housing, the non-european immigrants are less likely to live in deprived areas

For pull factor, we can see that chances are much higher that for non-european who tends to live with more non-european migrants to live in deprived area. For push factor, if non-europeans don't have access to public housing, non-european not living in deprived area will have a positive effect on immigrants living in deprived area. If non-europeans don't have access to public housing, European not living in deprived area will cause less non-european live in deprived area.

5. What is the role of past location? Does living in social housing in the past impact whether one lives in a deprived area in the current period?

- Identify whether individuals lived in social housing in the past. Create a variable called *past_hlm* with the following condition: *vsy* == "1".
- Identify whether the previous municipality in which households lived had a deprived area using the variable *vcomzus* (if it is equal to 0 and non-missing, the individual did not live in deprived area in the past). Create a new variable called *past_non_deprived*.
- Run the same baseline regression as before, but now interact *non_european* with these *past_hlm*. Do the same as before, but now use a triple-interaction term which also includes *past_non_deprived*. Do the same as before but now add location fixed effects. How do you interpret these results? Export regression results.
- The regressions in the previous section identified movers. We now look at non-movers. Create a dummy *non_mover* equal to 1 if *vlr* == "1" or *vlr* == "2". Run the baseline regression on non-movers only. Run this same regression, but now interact *non_european* with *hml* (whether the individual lives in social housing). Run this same regression but add fixed effects. How do you interpret these results? Export regression results.

. **reg deprived non_european#many_kids employed female married_couple low_diploma high_occupation**

| Source | SS | df | MS | Number of obs | = | 27,137 |
|----------|------------|--------|------------|---------------|---|--------|
| Model | 231.186128 | 8 | 28.8982661 | F(8, 27128) | = | 209.84 |
| Residual | 3735.92151 | 27,128 | .137714594 | Prob > F | = | 0.0000 |
| | | | | R-squared | = | 0.0583 |
| | | | | Adj R-squared | = | 0.0580 |
| Total | 3967.10764 | 27,136 | .14619353 | Root MSE | = | .3711 |

| | deprived | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|------------------------|----------|-----------|-----------|--------|-------|----------------------|
| non_european#many_kids | | | | | | |
| | 0 1 | .1106364 | .0085897 | 12.88 | 0.000 | .0938001 .1274727 |
| | 1 0 | .253977 | .0135902 | 18.69 | 0.000 | .2273395 .2806145 |
| | 1 1 | .422801 | .0241609 | 17.50 | 0.000 | .3754444 .4701577 |
| employed | | -.012016 | .0052197 | -2.30 | 0.021 | -.0222469 -.0017851 |
| female | | .0353071 | .0050139 | 7.04 | 0.000 | .0254797 .0451346 |
| married_couple | | -.0554296 | .0049753 | -11.14 | 0.000 | -.0651814 -.0456777 |
| low_diploma | | .0851636 | .005146 | 16.55 | 0.000 | .0750773 .09525 |
| high_occupation | | -.0530419 | .0062603 | -8.47 | 0.000 | -.0653124 -.0407714 |
| _cons | | .1411093 | .0062326 | 22.64 | 0.000 | .1288932 .1533255 |

Figure 1: The regression result of the variable *many_kids*

.

| | | | | | | |
|----------|------------|--------|------------|---------------|---|--------|
| Source | SS | df | MS | Number of obs | = | 27,137 |
| Model | 232.328645 | 8 | 29.0410807 | F(8, 27128) | = | 210.94 |
| Residual | 3734.77899 | 27,128 | .137672478 | Prob > F | = | 0.0000 |
| | | | | R-squared | = | 0.0586 |
| | | | | Adj R-squared | = | 0.0583 |
| Total | 3967.10764 | 27,136 | .14619353 | Root MSE | = | .37104 |

| deprived | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] | |
|--------------------------|-----------|-----------|--------|-------|----------------------|-----------|
| non_european#low_diploma | | | | | | |
| 0 1 | .0818908 | .0052347 | 15.64 | 0.000 | .0716306 | .0921511 |
| 1 0 | .2239268 | .0171136 | 13.08 | 0.000 | .1903833 | .2574703 |
| 1 1 | .3897262 | .01695 | 22.99 | 0.000 | .3565033 | .4229492 |
| many_kids | .1149027 | .0082267 | 13.97 | 0.000 | .098778 | .1310275 |
| employed | -.0130143 | .0052235 | -2.49 | 0.013 | -.0232527 | -.0027759 |
| female | .0349223 | .0050141 | 6.96 | 0.000 | .0250944 | .0447501 |
| married_couple | -.0554064 | .0049745 | -11.14 | 0.000 | -.0651567 | -.0456562 |
| high_occupation | -.0530993 | .0062592 | -8.48 | 0.000 | -.0653676 | -.040831 |
| _cons | .1433447 | .0062727 | 22.85 | 0.000 | .1310499 | .1556395 |

Figure 2: The regression result of the variable *low_diploma*