University of St.Gallen

Web mining – Term project

Spring 2023

Quentin Leoni

19-406-560

"Effect of cross-border labor mobility on national real estate price"

Credit: https://www.fondation-idea.lu/2019/04/30/combien-y-a-t-il-vraiment-de-travailleurs-frontaliers-au-luxembourg/

Table of contents

1. Abstract

Cross border labor mobility has drastically increased in Europe after creation of the EU and adoption of different treaties with its neighbors. Academic literature studied this phenomenon and its consequences in many regards including unemployment, competition, and price. The latter constitutes a core domain of research in this context and contains a whole subcategory in real estate, including Micheli, Rouwendal, Dekkers (2018), Dumeignil (2021) and Gonzalez, Ortega (2012) amongst other. Thanks to the heterogeneity dimension of political decision, most of those studies use changes in the legal framework as natural experiment for their analysis. In this case however, real time market data are gathered from the internet to isolate real estate price and its determinants in four different cities, two in France and two in Germany. half are located inland and the other half at the border of Switzerland to constitute the treatment group due to the number of cross-border commuters driven by minimum income discrepancy.

A naïve mean comparison between treated and non-treated is first used to test the effect. It returns poor results quality and does not suffice for causal claim. A linear regression is then performed to identify the weight of each determinant in price composition. Based on the results and backed by Grether, Mieszkowski (1973), footage appears to be the most important. Therefore, interaction effect of real estate footage and treatment are analyzed to test the impact of cross border labor mobility on real estate price. The assumption being that, by controlling for density, land tax, job attractiveness, and unemployment, any variation of footage's coefficient would reflect the treatment effect. No statistically significant difference is found for neither for France between region near Geneva and Lyon nor Germany comparing region near Basel and Stuttgart. Despite literature support, this research design doesn't allow to conclude in the presence of a treatment effect.

2. Data

Natural experiment is usually preferred for such analysis, but it is well known that adequate circumstances to perform it are rare. Therefore, web mining technics are used in this study to generate the dataset. Indeed, thanks to many real estate's announcement websites, it is easy to obtain market data such as price, size, number of rooms and type of estate. One French and one German property advertisement website were picked, namely, seloger.com and Wohnungsboerse.net.
While the power of web scrapping would allow to deal with enormous amount of data, nowadays many hosting servers are equipped with bot detection algorithm restricting the scope of possibilities. Indeed, seloger.com as the first French website per volume, automatically block the access after too many requests. It nevertheless remains the best option for remote location. In Germany, Immobilienscout24.de goes even further and has anti-bot protection preventing any attempt. Second best option called wohnungsboerse.net is therefore held.

As mentionned, first step of research design is focused on identifying two municipalities per country as similar as possible. The choice fell on Miribel and Gex in France and Leinfelden-Echterdingen and Lörrach in Germany. That selection process raises important risk of selection bias and reflect the weakness of this study. However, comparisons of many criteria were made to limit systematic error as much as possible. First, by choosing towns in the same state, most difference in land tax can be controlled as for France and Germany, it is managed on a national or state level. In the case of France, only two municipal taxes are applied: Property tax on property (TFPB) and property tax on non-buildings (TFPNB) and will be considered in the results. No such tax was found within German state of Baden-Württemberg (BW). Regarding unemployment rate, communal French latest figures available are for 2019 only, a period when economic conditions were very different. State numbers are therefore retained for both French cities with 5.4% [(1)].

[1] Data of the 4th trimester 2022 from INSEE, https://www.insee.fr/fr/statistiques/2012804

Baden-Württemberg state on the other hand provides more recent figures for Lörrach and Leinfelden-Echterdingen with 4.1% and 3.4% [2] respectively. It's worth comparing with neighboring regions in Switzerland to identify any structural difference. At the latest, Geneva's state announces 3.7% [3] against 3% [4] for Basel-Stadt. Those differences must be acknowledged.

Intimately connected, job attractivity of the location has its importance. Therefore, all cities are within a circle of 10-15 km of major agglomerations, namely Geneva or Lyon for France and Basel or Stuttgart for Germany. For that reason, the locations are assumed to be likely attractive in this regard. It should be noted that no discrimination in hiring foreign workers were assumed.

Finally, density of cities must be similar as land shortage remains one of the most important metrics when it comes to property price. The total number of inhabitant and inhabitants per squared kilometer have hence been controlled. According to Wikipedia as of 2020, Miribel has a population of 10,202 and a density of 420/km^2 while Gex announces 13,177 and 410/km^2. On the other hand, Leinfelden-Echterdingen counts 40,100 for 1,300/km^2 while Lörrach declares 49,318 inhabitants with 1,300/km^2 as well.

The table below summarizes those characteristics:

Town	State	Unemployment	Distance to city	Nearest city	Population	Density
Miribel	Aim	5.4%	13.1 km	Lyon	10,202	420/km^2
Gex	Aim	5.4%	16.2 km	Geneva	13,177	410/km^2
Leinfelden	BW	3.4%	11.3 km	Stuttgart	40,100	1,300/km^2
Lörrach	BW	4.1%	13.7 km	Basel	49,318	1,300/km^2

Figure 1 – cities characteristics

From those 4 locations, every announcement available on the respective website are collected, including: Price (1), the type of estate (2), number of rooms (3) and footage (4). Properties other than apartment and villa are excluded to focus the research on so-called "traditional" estate. The same process is applied on Wohnungsboerse.net. At the moment of the study, a total of 297 property advertisements were collected for Germany including 127 in Leinfelden-Echertdingen against 170 in Lörrach and 292 for France including 95 for Miribel against 197 for Gex.

Figure 2 – announcement example

[2] Data of March 2022 from Statistisches Landsamt, https://www.statistik-bw.de/Arbeit/Arbeitslose/03033022.tab?R=KR336
[3] Data of April 2023 from Republique et Canton de Genève, https://statistique.ge.ch/domaines/apercu.asp?dom=03_03
[4] Data of April 2023 from Statistisches Amt, https://www.statistik.bs.ch/zahlen/tabellen/3-arbeit-erwerb/arbeitslose.html

The proportion of apartment for sale on seloger.com is of is of 52% in the non-treated group and 69% for the treated. For Germany, those figures are 60% and 56% respectively. Type is slightly unbalanced in the data but will be controlled in the research design.

Here's a summary of the three main characteristics collected for both countries:

```
> summary(data_f$Price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 109900  259000  346000  447561  523750 2980000
> summary(data_f$Room)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   2.000   2.000   2.664   4.000   7.000
> summary(data_f$Footage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   28.0    62.0    78.0   108.9   135.0   517.0
> print(sumup_f)
      Miribel Gex
Flat        50 137
House       45  60
```

```
> summary(data_d$Price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 168800  299000  466900  578742  768200 3100000
> summary(data_d$Room)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    3.00    4.00    4.69    6.00   30.00
> summary(data_d$Footage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   25.0    71.0   109.0   130.3   155.0   880.0
> print(sumup_d)
      Leinfelden Lörrach
Flat           76      96
House          51      74
```

Figure 3 – data statistics

3. Empirical Strategy/Methods

In 2014 Switzerland voted to adopt what would have been the highest minimum wage in the world at CHF 22. While the initiative was rejected at 76.3% [5], many states implemented similar laws within their jurisdiction. Notably, Geneva and Basel-Stadt with an amount of CHF 24 [6] and CHF 23[7] respectively. On the other side of the borders, France, and Germany have also put in place similar measures but at a significant lower level with 11.52€ [8] and EUR 12€ [9]. This discrepancy represents a serious arbitrage opportunity that many have figured out. Cross-border mobility labor is therefore very important in those areas implying major population movement.

To test the causal effect of this phenomena on real estate price, a regression discontinuity design could be used with border as the intervention. Indeed, by selecting two cities of similar characteristics in the neighbor countries with only one at the border, property price difference would reflect the treatment effect. While the risk for bias selection is important, controls can be made to limit the impact (see Data).

Two methods are used to determine the impact of cross-border labor mobility. First, simply comparing the average price of the treated and the non-treated for both geographical areas. An additional approach based on the property type is also tested to control for the effect of neighborhood as a potential confounder.

a) Naïve approach:

$$\widehat{ATE} = \frac{1}{N}\sum(price_1(i) - price_0(i))$$

b) Property type controlled:

$$\widehat{ATE} = \frac{1}{N}\sum(price_{1,t}(i) - price_{0,t}(i)) \ for\ t = 1{,}2\ (apartment\ or\ house)$$

[5] Volksinitiative vom 23.01.2012 «Für den Schutz fairer Löhne» https://www.bk.admin.ch/ch/d/pore/va/20140518/det583.html
[6] Appliquer le salaire minimum genevois, https://www.ge.ch/appliquer-salaire-minimum-genevois
[7] Gesetz über den kantonalen Mindestlohn (MiLoG), https://www.gesetzessammlung.bs.ch/app/de/texts_of_law/812.200
[8] Smic (salaire minimum de croissance), https://www.service-public.fr/particuliers/vosdroits/F2300
[9] Eine Frage des Respekts, https://www.bundesregierung.de/breg-de/aktuelles/mindestlohn-faq-1688186

The second method aims to be more subtle in its design and analyzes the effect of treatment indirectly. With the use of a linear regression, price determinants effect is decomposed while including the interaction effect with treatment to test its impact on coefficients.

c) Linear regression with interaction effect

$$Price(i) = \beta_0 + \beta_1 Type(i) + \beta_2 Room(i) + \beta_3 Footage(i) + \beta_4 Treatment(i)$$
$$+\beta_5 Type(i) * Treatment(i) + \beta_6 Room(i) * Treatment(i)$$
$$+\beta_7 Footage(i) * Treatment(i)$$

4. Results

d) Mean difference methods

France returns a price difference of 136,299 € without controlling for the property type. With apartment only, it decreases to 66,925€ while for houses it jumps to 418,955€. Those differences must be treated with cautious as the proportion of type in the dataset is unequal.
Germany, on the other hand, returns a price difference of 12,102€ without controlling for the property type. With apartment only, it turns negative with -102,424€ while for house it goes upside down at 134,189€.

Such differences betray a poor approach to testing the treatment effect. Indeed, with those little observations, the law of large numbers doesn't hold. As counterfactual can't be observed, the assumption that characteristics of the treatment group are, on average, identical the non-treated is hence weakly supported. In addition, using data from only one location per group is certainly prone to many biases.

e) Linear regression method:

France data returns the following table:

MODEL INFO:
Observations: 292
Dependent variable: data_f$Price
Type: OLS linear regression

MODEL FIT:
$F(7,284) = 106.10$, $p = 0.00$
$R^2 = 0.72$
Adj. $R^2 = 0.72$

Standard errors: OLS

	Est.	S.E.	t val.	p
(Intercept)	79968.51	44015.84	1.82	0.07
data_f$Type2	29934.30	55718.60	0.54	0.59
data_f$Room	29313.54	20793.85	1.41	0.16
data_f$Footage	1757.84	626.72	2.80	0.01
data_f$Treatment1	22566.02	53068.72	0.43	0.67
data_f$Type2:data_f$Treatment1	149227.61	68600.09	2.18	0.03
data_f$Room:data_f$Treatment1	-15609.58	26413.68	-0.59	0.56
data_f$Footage:data_f$Treatment1	913.89	674.33	1.36	0.18

Figure 4 – France OLS results

Based on standard convention, only two coefficients seem to be statistically significant: β_2 (Footage) and β_7 (interaction term of treatment and type). While failing 5% confidence test, the constant term remains

important. In this context, it could be interpreted as the average price of land before construction and return a figure of 79,968€. As expected, footage plays a relevant part with an average effect of 1,757€ per additional square meter of built property (not to be confused with land). In addition, the interaction term of the treatment with property type shows an average effect of 149,277€. This reflects the unbalanced proportion of villa-type estate in Gex compared to Miribel. Finally, the interaction term between treatment and footage (β_7) embodies the intended effect (ATE) of the study and seems not to be significant.

In other words, labor mobility between Gex and Geneva has no positive impact on real estate prices in Gex. This does not support the presence of a treatment effect from French working in Switzerland on the national real estate market. Before confirming this claim, the German results must be analyzed.

MODEL INFO:
Observations: 297
Dependent variable: data_d$Price
Type: OLS linear regression

MODEL FIT:
$F(7,289) = 109.19$, $p = 0.00$
$R^2 = 0.73$
Adj. $R^2 = 0.72$

Standard errors: OLS

	Est.	S.E.	t val.	p
(Intercept)	197816.06	36538.05	5.41	0.00
data_d$Type2	65996.05	50920.75	1.30	0.20
data_d$Room	-13784.39	15008.48	-0.92	0.36
data_d$Footage	3646.66	619.72	5.88	0.00
data_d$Treatment1	-46729.33	45596.37	-1.02	0.31
data_d$Type2:data_d$Treatment1	185748.06	65266.48	2.85	0.00
data_d$Room:data_d$Treatment1	-18934.70	17757.06	-1.07	0.29
data_d$Footage:data_d$Treatment1	-271.05	677.26	-0.40	0.69

Figure 5 – Germany OLS results

Here, three coefficients are statistically significant: β_0 (the intercept), β_2 (Footage) and β_6 (interaction term of treatment and type). The average price of the land appears to be highly significant and more than double at 197,816€. Same observation for the impact of any additional square meter at 3,646€. Finally, the interaction term of the treatment with property type rises to 185,748€. There's a lot more to draw attention here. To begin with, the negative "Room" coefficient which raises concern. Indeed, it doesn't make any sense to observe a decrease in price for any additional chamber. Relevance of those numbers should then be address cautiously. Secondly, the treatment effect appears to be negative here, even though below any statistically relevant threshold. When looking at its effect on price determinants, footage remains below zero contrasting with France analysis.

As the relevance of footage remains constant across both datasets, a plot comparison allows to better visualize the interaction effect of treatment and footage representing the ATE in this method.

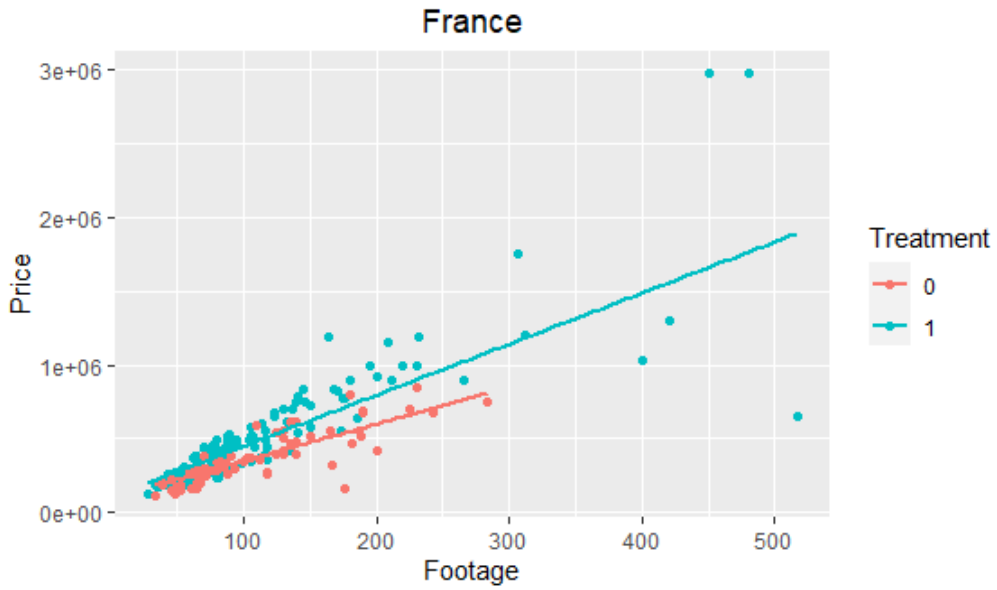

Figure 6 – Treatment effect, France

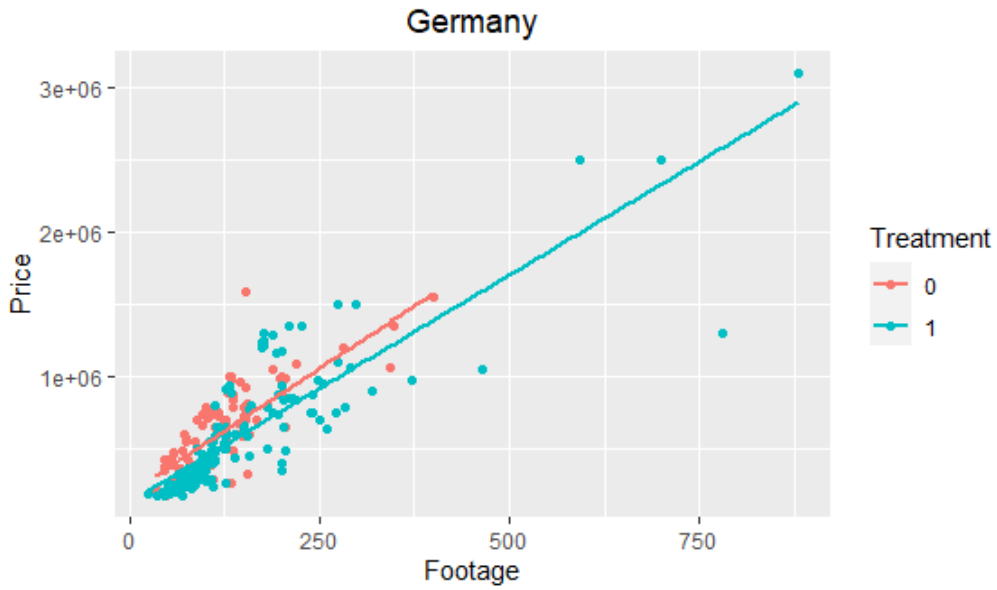

Figure 7 – Treatment effect, Germany

Both regressions show an adjusted R^2 of 0.72 but return important coefficient differences. Represented in the graph above as the difference of regression lines' slopes, the interaction term can be distinguished as the closest form of treatment effect. Indeed, France dataset clearly shows a slightly steeper relationship between footage and property price for the treated than the non-treated reflecting regression tab's results. German's coefficients on the other hand are very similar which translates as almost parallel regression lines.

Despite similar conclusions, it seems that this decomposition of the price is incomplete. Indeed, according to hedonic pricing theory, footage, type, and room are internal factors and only explain half of the market data. External factors include neighborhood appealing features such as access and quality of school or public transport, presence of a downtown area, cultural offering, air quality, etc. Those are implicit to the city selection process and hard to control for. Therefore, conclusions drawn from this research design must be treated with utmost caution as new announcements or different town choice could challenge it all.

5. Concluding remarks

Real estate market is highly location-dependent and one slight difference could have drastic consequences in attractivity, volume and price. While some controls were made in the city selection, countless others were forgotten. In addition, even amongst the retained criteria, finding a perfect match is impossible. A data collection across many more location and throughout time could be an interesting amelioration of this research design. Furthermore, the study didn't make distinctions between foreign workers coming to Switzerland and Swiss citizens moving across the border as both benefit from arbitrage opportunity. However, the composition of cross-border movement and its proportion would be a very interesting information to consider as well.

Finally, one should also keep in mind that any causality test relies on assumptions which hence exposes its conclusions to debate.

6. References

Micheli M., Rouwendal J., Dekkers J, 2018 “Border Effects in House Prices” Real Estate Economics 47 (3): 757-783. https://doi.org/10.1111/1540-6229.12255

Dumeignil C. 2021 “The Impact of Cross-border Labor Mobility on Real Estate Price Trends: A Natural Experiment” Journal of Regional Science 45 (1): 108-132. https://doi.org/10.1177/0160017621989394

Gonzalez L., Ortega F. 2012. “Immigration and Housing Booms: Evidence from Spain.” Journal of Regional Science 53 (1): 37–59. https://doi.org/10.1111/jors.12010

Buch T., Schmidt T., Niebuhr A. 2009 “Cross-border commuting in the Danish-German border region” Journal of Borderland Studies 24 (2): 38-54. https://doi.org/10.1080/08865655.2009.9695726

d’Albis H., Boubtane E., Coulibaly D. 2019. “International Migration and Regional Housing Markets: Evidence from France.” International Regional Science Review 42 (2): 147–80. https://doi.org/10.1177/0160017617749283

Grether D. M., Mieszkowski P. 1974. “Determinants of Real Estate Values.” Journal of Urban Economics 1 (2): 127–45. https://doi.org/10.1016/0094-1190(74)90013-8

Liu, M., Ma, QP. Determinants of house prices in China: a panel-corrected regression approach. Ann Reg Sci 67, 47–72 (2021). https://doi.org/10.1007/s00168-020-01040-z

Eryk Lewinson, April 2023. “A Comprehensive Guide to Interaction Terms in Linear Regression” Nvidia Developer, https://developer.nvidia.com/blog/a-comprehensive-guide-to-interaction-terms-in-linear-regression/

“The French-Swiss border real estate market”, April 2021. Stone and Wood, https://www.stoneandwood.fr/en/french-swiss-bordel_reat-estate-market/