# Local buyers' exclusion through price in the farmland market: a potential outcome approach

Olivier Beaumais, Sauveur Giannoni, Caroline Tafani

# Local buyers' exclusion through price in the farmland market: a potential outcome approach

Olivier Beaumais[a,b], Sauveur Giannoni[b] and Caroline Tafani[b]

[a]CREAM EA-4702, University of Rouen-Normandy; [b]UMR CNRS 6240 LISA; Université de Corse - Pascal Paoli; Campus Mariani, BP 52, 20250 Corte, France.

**ARTICLE HISTORY**
Compiled July 19, 2024

**ABSTRACT**
The influx of non-local buyers into the land market is commonly held responsible for the exclusion of local buyers through price. We study the seaside farmland market in Corsica. A massive price gap between non-local and local buyers is observed. To explore this gap, we rely on a treatment effect approach. We first estimate hedonic price models, while controlling for omitted variable bias using an innovative method. We go further and estimate a general potential outcome model which allows capturing observable and unobservable preference heterogeneity between buyers. We find that the price gap indeed reflects preference heterogeneity, suggesting market segmentation at work.

**KEYWORDS**
local buyers' exclusion, potential outcome model, price gap, heterogeneous preferences

**JEL CLASSIFICATION**
R30, C30

## 1. Introduction

The influx of non-local buyers into land and housing markets has become a major concern for many cities, regions and countries across the world. Numerous columnists stress the fact that the presence of non-local buyers raises prices and prevents local people from buying a house or plot of land, thus leading to the exclusion of local (domestic) buyers from the market. In general, these non-local buyers are willing to buy or build second homes and are characterized by preferences and opinions on the market value of the land in question that differ from local buyers' preferences and opinions. Non-local buyers also incur higher research costs, which results in a selection process that differs from one category of buyers to the other. Finally, exclusion would be reinforced by the fact that non-local buyers can easily afford a higher price than local buyers.

Pre publication version. Cite as: "Beaumais, O., Giannoni, S., & Tafani, C. (2021). Local buyers' exclusion through price on the farmland market: a potential outcome approach. Applied Economics, 53(19), 2206–2217. https://doi.org/10.1080/00036846.2020.1856767"

CONTACT Sauveur Giannoni. Email: giannoni@univ-corse.fr.

The case of Australia (Nicholls 2016) or New Zealand (Davidson 2016) facing a substantial inflow of Chinese buyers aptly illustrates this phenomenon. Britain (Williams 2015) and Canada (Sturgeon 2015) are also confronted with a similar issue.
Despite the worldwide nature of what we call in this paper the local buyers' exclusion problem, to the best of our knowledge it has received little attention from researchers. A notable exception is Chao and Yu (2015), who studied the optimal taxation scheme to suppress the inflationary influence of non-local buyers on the housing market, applied to the case of Hong Kong. This study illustrates that when scarcity of land is reinforced by specific constraints, as is the case for small or medium-sized islands, the exclusion phenomenon is likely to be accentuated.
As a prime example, the island of Corsica in France has, since the late 1990s, experienced a substantial influx of non-local buyers. Locals are said to suffer exclusion from the seaside farmland market, which provides most of the plots of land for housing construction and is consequently under heavy pressure. This causes political and social problems, with some political representatives asking for public intervention in order to protect local buyers.
To empirically assess the reality and extent of exclusion through price, we used a database of more than 5,600 observations covering all the seaside farmland sale agreements between 1998 and 2008. A simple descriptive analysis of the data shows a massive entry of both local and non-local buyers on the farmland market and overall stability in the share of non-local buyers. Rather than local buyers being excluded, both types of buyers are observed to coexist. Over the study period, prices increased faster for non-local than for local buyers, with the seaside farmland market exhibiting a high price gap: non-local buyers pay an average premium of roughly €8.75 per square metre. As mentioned earlier, preference heterogeneity between local and non-local buyers and differences in search costs, could explain all or part of the price gap and thus the coexistence of local and non-local buyers instead of locals' exclusion.
However, an in-depth exploration of the sources of the price gap requires an appropriate econometric strategy. A key feature of land and housing markets is that prices are set through a decentralized two-person search and bargaining process (King and Sinden 1994; Zhou, Gibler, and Zahirovic-Herbert 2015). Considering the buyer being non-local as a 'treatment', assessing the price gap as a treatment effect means controlling for omitted bias and/or endogeneity of the selection process (non-local versus local buyers) and for land attributes, while allowing for observable and unobservable preference heterogeneity. As a benchmark, we first estimate standard OLS hedonic price models, while controlling for omitted bias using an innovative approach recently proposed by Oster (2017). We find the treatment effect to be very stable across the specifications (around €8.10 for the most advanced hedonic model) and not subject to omitted variable bias. We argue further by suggesting that the effect of observable and unobservable heterogeneity between non-local and local buyers can be elegantly captured by using the general potential outcome model approach. In so doing, we find that the non-local premium (price gap) does indeed have its roots in observable and non-observable preference heterogeneity and is even higher, around €12 per square metre, than the estimation of standard OLS hedonic models suggests.
The rest of the paper is organized as follows. Section 2 presents Corsica's seaside farmland market and the database. Econometric method and results are explained in Section 3. Section 4 briefly discusses the results and draws a conclusion.

## 2. The seaside farmland market in Corsica

### *2.1. Context*

Corsica is a French island in the Mediterranean Sea with 326,000 inhabitants and a total surface area of about 8,680 $km^2$, two-thirds of which is very mountainous. Farming used to be the most important activity in Corsica but, due to a sharp decline, the farming sector nowadays accounts for only 1.5% of the regional value added. On the other hand, the island is one of the most popular tourist destinations in France. According to official data (INSEE, 2015), 35 million overnight stays are registered each year, with total tourism expenditure of €2.5 billion a year, one third of the regional GDP. Along with tourism development, the number of second homes is steadily growing. The share of second homes in total housing was 36.4% in 2012. Furthermore, the population grew by more than 25% between 1999 and 2015. Due to these demographic dynamics, land is becoming scarcer and scarcer, and the pressure on the farmland market has been increasing for several years. This pressure is even stronger in the seaside regions of Corsica due to their distinct appeal. In this context, the local population is concerned about the difficulty of buying land, especially for housing purposes. The local buyers' exclusion problem has accordingly become a prominent political issue. In particular, local nationalist political parties have put pressure on the French national government to obtain protective legislation for local buyers. In a 2011 official report[2]entitled *Local commitment to a land and housing policy*, the President of the local government stated:

> A measure of the reality of problems in the land market, and this is a major issue for Corsica, lies in the large number of disputed or invalidated urban plans, which is evidence of the pressure on the land market and highlights the difficulties of finding a long-term equilibrium point between the preservation of the general interest, on the one hand, and the pressure coming from private interests in high market value areas, on the other hand.

In France, the SAFER is a semi-public institution in charge of the promotion and development of agriculture and rural areas. When the owner of a parcel of farmland and a potential buyer have reached a preliminary sale agreement, they must make a declaration of intent to the SAFER before being allowed to complete the transaction. This prior declaration is called a DIA. The present study draws on a database which covers all the DIA in Corsican seaside municipalities between 1998 and 2008[3]. The data illustrate some interesting features of Corsica's seaside farmland market. Between 1998 and 2008, the SAFER in Corsica registered over 5,600 DIA for seaside municipalities. In total, between 1998 and 2008, 662 land plots were bought by farmers and only 700 land plots remained intended for farming use. These figures emphasize the fact that most of the farmland is sold for recreational and housing purposes. In the remainder of the paper, we focus on the sales for which the plot of land is not intended for farming.

The increase in the annual number of sale agreements also reflects an increasing demand related to demographic and tourism pressure: as an illustration, Figure 1 depicts the rise in the number of sale agreements per year. These agreements quadrupled in 11 years, from 155 agreements in 1998 to 616 in 2008.

A common belief in Corsica is that the growing demand for land from non-local buyers results in higher prices. Yet, Figure 1 and Table 1 clearly indicate that both local and non-local demand is increasing, and that their relative shares are almost constant

[2]In French, the report is entitled *Engagement territorial pour une politique du foncier et du logement, p. 43.*
[3]Note that more recent data were not available.

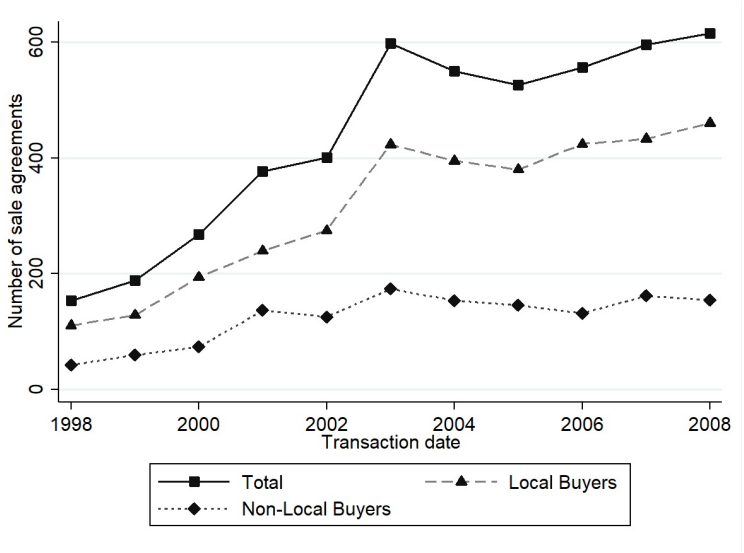


**Figure 1.** Annual number of sale agreements

| Year | Local buyers | Non-local buyers |
|---|---|---|
| 1998 | 0.72 | 0.28 |
| 1999 | 0.68 | 0.32 |
| 2000 | 0.72 | 0.28 |
| 2001 | 0.64 | 0.36 |
| 2002 | 0.69 | 0.31 |
| 2003 | 0.71 | 0.29 |
| 2004 | 0.72 | 0.28 |
| 2005 | 0.72 | 0.28 |
| 2006 | 0.76 | 0.24 |
| 2007 | 0.73 | 0.27 |
| 2008 | 0.75 | 0.25 |
| 1998-2008 | 0.72 | 0.28 |

**Table 1.** Share of sale agreements involving local and non-local buyers

over the study period. As shown in Figure 2, the idea of land scarcity induced by growing demand is supported by the fact that the price of a square metre of farmland (in constant 2008 euros) grew over fivefold (5.3) between 1998 and 2008. But this figure also highlights an even more interesting feature of the farmland market. While the average price paid by a local buyer increased around fourfold (3.97) between 1998 and 2008, the average price paid by a non-local buyer rose nearly ninefold (8.76).

Figure 2 and Table 2 support the existence of a price gap between local and non-local buyers. These figures are not consistent with the idea of local buyers' exclusion. The average farmland price is increasing, but local buyers continue to pay a significantly lower price. It should be noted that the existence of a price gap between non-local and local buyers, of a price premium paid by non-local buyers on the land and/or housing market, is well documented in the literature (Miller, Sklarz, and Ordway 1988; Lambson, McQueen, and Slade 2004). Recently, Levkovich, Rouwendal, and Brugman (2018) provided tantalizing evidence of a price gap between agricultural, commercial and industrial land prices due to land use restrictions in the Netherlands. However,

| Variable | Mean | Median | Std. Dev. | Num. of observations |
|---|---|---|---|---|
| $m^2$ price | 16.152 | 6.789 | 25.416 | 4,827 |
| $m^2$ price for local buyers | 13.681 | 5.622 | 21.174 | 3,464 |
| $m^2$ price for non-local buyers | 22.432 | 9.818 | 33.073 | 1,363 |

**Table 2.** Summary statistics on $m^2$ prices (€ 2008) within the full sample

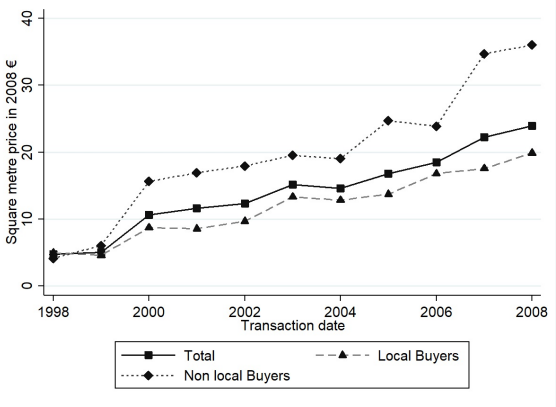


**Figure 2. Average annual $m^2$ price in 2008 constant €**

another potential explanation for the existence of this price gap could simply lie in differences in the valuation of land plot attributes by local and non-local buyers, i.e., in heterogeneous preferences. The aim of this paper is to test the latter hypothesis. Next section provides more details on the variables available in the database.

### *2.2. Data*

As mentioned earlier, the database is restricted to land plots that are not sold to farmers and that will not be used for farming. Table 3 reports some descriptive statistics for the overall sample and both subsamples of locals and non-locals buyers.

Once transactions involving farmers or for farming use dropped, our unique database includes 4827 transaction agreements. It provides information on the intrinsic characteristics of the land plot:

- the land plot square metre price in 2008 euros;
- the year of the sale;
- the cadastral category of the land plot;
- the municipality and the region in which the land plot is located.

We also have some administrative information, notably the place of residence of the buyer and a number of other variables were added to the database:

- the location of the land plot in a city;
- the undeveloped vs developed nature of the cadastral section in which the land plot is located (no existing agricultural building on the land plot vs existing agricultural building on the land plot; the land plot is fully serviced).

At this stage, Table 3 deserves a brief comment. Locals seem to be more prone to buy land in the Ajaccio or Bastia areas, that are the main employment areas on the island, while non-locals seem to favour places such as the Southern or Valinco areas, that are

renown tourism destinations. Furthermore, these figures also suggest that non-locals could be attracted by the countryside since their propensity to buy in an area close to a city is lower.

| Variable | Description | | | |
|---|---|---|---|---|
| price | Land plot price in €2008 | - | | |
| Variable | Description | overall % | non-locals % | locals % |
| YEAR | Year of sale | | | |
| | 1998 | 3.19% | 3.15% | 3.20% |
| | 1999 | 3.92% | 4.40% | 3.72% |
| | 2000 | 5.55% | 5.43% | 5.60% |
| | 2001 | 7.81% | 10.05% | 6.93% |
| | 2002 | 8.31% | 9.24% | 7.94% |
| | 2003 | 12.37% | 12.77% | 12.21% |
| | 2004 | 11.37% | 11.30% | 11.40% |
| | 2005 | 10.90% | 10.71% | 10.97% |
| | 2006 | 11.52% | 9.68% | 12.24% |
| | 2007 | 12.33% | 11.89% | 12.25% |
| | 2008 (Reference year) | 12.74% | 11.37% | 13.28% |
| CAD CAT | Cadastral category of the land plot | | | |
| | No predominant feature | 33.33% | 35.73% | 32.39% |
| | Meadow (Reference level) | 37.46% | 34.34% | 38.68% |
| | Vineyard and orchard | 3.73% | 3.60% | 3.78% |
| | Scrubland | 4.33% | 5.80% | 3.75% |
| | Wilderness | 7.23% | 7.34% | 7.19% |
| | Wood | 10.90% | 10.35% | 11.11% |
| | Garden | 3.02% | 2.86% | 3.09% |
| Region | Area in which the land plot is located | | | |
| | Ajaccio area | 15.19% | 13.28% | 15.94% |
| | Southern area | 22.81% | 28.47% | 20.58% |
| | Valinco area | 5.70% | 8.95% | 4.42% |
| | Sevi-Sorru area | 5.14% | 6.31% | 4.68% |
| | Bastia area (Reference region) | 26.56% | 18.56% | 29.71% |
| | Balagna area | 10.96% | 13.43%% | 9.99% |
| | Eastern area | 13.65% | 11% | 14.70% |
| DEVELOPED | The land plot is developed (1/0) | 7.23% | 11.37% | 5.60% |
| CITY | Land plot located in a city (1/0) | 29.17% | 23.25% | 46.46% |
| NOURB | Land plot preserved from urbanization (1/0) | 43.67% | 54.59% | 39.38% |
| NONRES | Non-resident buyer (1/0) | 28.24% | - | - |

**Table 3.** Description of the variables used in the model

Basically, estimating the price gap between non-local and local buyers amounts to estimating the treatment effect associated with being a non-local buyer on the sale price. In our empirical strategy, this treatment effect will be captured by a binary variable *nonres* which takes the value 1 if the buyer is non-local (non-resident), and 0 otherwise. Of course, a common concern regarding the estimation of treatment effects is the omitted variable bias, which, when not controlled for, can lead to wrong conclusion about the phenomenon under study. Consequently, the econometric strategy we implement in the following section aims both to control for the omitted variable bias and to take into account the observable and unobservable heterogeneity of preferences

between non-local and local buyers.

## 3. Estimation strategy and results

### *3.1. Simple hedonic pricing models and omitted variable bias*

As a benchmark, we estimate simple OLS hedonic models and check for omitted bias using the method proposed by Oster (2017).
By elaborating on the current practice of examining the stability of the treatment effect after the inclusion of control variables, while emphasizing the need to simultaneously take the R-squared movements into account, Oster (2017) approximates a bias-adjusted treatment effect. The idea is to first run a basic regression of the dependent variable (here the sale price) on the treatment variable (here the *nonres* variable); $\mathring{\beta}$ and $\mathring{R}$ are the treatment effect and R-squared resulting from this basic regression. Then, control variables are included; $\tilde{\beta}$ and $\tilde{R}$ are the treatment effect and R-squared from this full regression. The bias-adjusted treatment effect, $\beta^*$, is

$$\beta^* \approx \tilde{\beta} - \delta \left[ \mathring{\beta} - \tilde{\beta} \right] \frac{R_{\max} - \tilde{R}}{\tilde{R} - \mathring{R}}$$

Where $\delta$ is a measure of the relative degree of selection on unobservable versus observable variables and $R_{\max}$ denotes the R-squared from a hypothetical regression of the dependent on treatment and both observable and unobservable controls. In the extreme case where the full set of control variables would be available, $R_{\max} = 1$. Based on the study of papers relying on randomized data, Oster (2017) further suggests $R_{\max} = min\{1.3\tilde{R}, 1\}$.

Thus, following Oster (2017), we first propose a lower bound of the non-local buyer effect assuming equal selection on unobservable and observable variables ($\delta = 1$). The idea, here, is to check whether this lower bound is included in the 95% confidence interval around the estimated treatment effect. If so, omitted bias is unlikely to explain the treatment effect.

Then, we calculate the degree of proportionality (the value of $\delta$) for which the non-local buyer effect would vanish. A value of $\delta$ greater than one means that selection on unobservable variables should be more important than selection on observable variables to make the treatment effect vanish.

| Basic Effect<br>[95% confidence interval] | Controlled Effect<br>[95% confidence interval] | Bias-Adjusted $\beta^*$<br>[95% confidence interval] | $\delta$ for $\beta = 0$<br>given $R_{\max}$ |
|---|---|---|---|
| 8.75 ($\mathring{R} = 0.024$)<br>[6.86,10.64] | 8.77 ($\tilde{R}$=0.180)<br>[7.05,10.49] | 8.77<br>[7.01,10.54] | 34.29 |

Note: $R_{\max} = 1.3\tilde{R} = 0.234$

**Table 4.** Benchmark Hedonic Models and Omitted Bias Check

From the results reported in Table 4[4], two main conclusions can be drawn. First, the treatment effects are found very close to each other: the basic hedonic model, without control variables, estimates the treatment effect at €8.75 per square metre, while the hedonic model with control variables provides an estimate of €8.77 per square metre. More importantly, the bias-adjusted treatment effect is about €8.77, suggesting that no omitted variable bias is at play. Second, the absence of omitted variable bias is further confirmed by the value of $\delta$ which would produce a treatment effect of zero.

[4]We used the user written Stata command psacalc provided by Oster (2017).

Here, $\delta = 34.29$ which means that the unobservables would need to be about 34 times as important as the observables to produce a treatment effect of zero. Taken together, these results suggest that the treatment effect measured by the hedonic pricing model with control variables is robust to omitted variable bias.

However, our simple hedonic pricing model with controls appears quite poor: in particular, it does not allow us to explore the heterogeneity of preferences between non-local and local buyers. A first way to model observable preference heterogeneity is to introduce interaction variables between the treatment variable (*nonres*) and the controls into the hedonic pricing model (see Table 5)[5]. Indeed, when an interaction variable between non-residents and one of the observable characteristics (controls) of the land plot is significant, it implies that non-locals weight this characteristic differently than locals in their preferences. In doing so, we find a treatment effect of about €8.10 (95% confidence interval: [6.33,9.88]), very close from the treatment effects reported in Table 4. In doing so, we also find evidence of *observable* preference heterogeneity. Some of the interaction variables are found to be significant: for example, non-local buyers value developed plots of land more than locals (positive and significant interaction between *nonres* and *developed*), and favour certain locations, such as the Ajaccio area (positive and significant interaction between *nonres* and the binary coding a location in the Ajaccio area).

However, beyond observable heterogeneity, it seems necessary to us to control for *unobservable* heterogeneity, especially because price variability appears higher for non-locals than for locals (see Table 2). Controlling for unobservable heterogeneity leads us to change our econometric approach to a more general model than those we have estimated so far.

### *3.2. Allowing for observable and unobservable heterogeneity: the potential outcome model*

As mentioned earlier, a key feature of land and housing markets lies in the fact that prices are set through a decentralized two-person search and bargaining process. Empirically, this bargaining process can be viewed as an endogenous switching or as an endogenous selection mechanism, which can thus be presented within the potential outcome framework (outstanding contributions to the potential outcome model include Holland 1986, Heckman 2010 and Pearl 2012).

Define the price a seller gets from sale $i$ if he/she sells to a resident (local) and a non-resident (non-local) as, respectively, $p_{0i}$ and $p_{1i}$. As mentioned earlier, $nonres_i$ is a binary variable which takes the value 1 if the buyer is non-local (non-resident), and 0 otherwise. We never observe both $p_{0i}$ and $p_{1i}$, only one or the other. Thus, we observe the selling price

$$p_i = (1 - nonres_i)p_{0i} + nonres_i p_{1i} \tag{1}$$

The general potential-outcome model is

[5]Note that the model is now non-linear, and the Oster (2017) approach no longer applies, which was confirmed by personal correspondence with E. Oster.

$$
\begin{aligned}
p_{0i} &= x_i\beta_0 + \epsilon_{0i} \\
p_{1i} &= x_i\beta_1 + \epsilon_{1i} \\
nonres_i &= \begin{cases} 1, & \text{if } w_i\gamma + u_i > 0 \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
$$

where $x_i$ is a vector of covariates which are used to model the price (i.e., land-plot characteristics, location, etc.), $w_i$ is the vector of covariates used to model the selection of a non-local buyer, $\beta_0$, $\beta_1$, $\gamma$ are vectors of parameters to be estimated. $\beta_0$ and $\beta_1$ can be different from one another, which allows local and non-local buyers to weight the $x_i$'s differently. Put another way, the model specification captures heterogeneous responses, across local and non-local buyers, to observationally identical land plot characteristics, thus reflecting observable preference heterogeneity. Note that, in practice, this is strictly equivalent to introducing interaction variables between the *nonres* variable and land plot characteristics, as we did in the previous section for the hedonic price model with observable heterogeneity of preferences.

The generality of the model also stems from the fact that the vector of error terms $(\epsilon_{0i}, \epsilon_{1i}, u_i)$ comes from a trivariate normal distribution with mean 0 and covariance matrix

$$
\begin{bmatrix}
\sigma_0^2 & \sigma_{01} & \sigma_0\rho_0 \\
\sigma_{01} & \sigma_1^2 & \sigma_1\rho_1 \\
\sigma_0\rho_0 & \sigma_1\rho_1 & 1
\end{bmatrix}
$$

Thus, the model allows for separate variance and correlation parameters for the local and non-local groups, which means that the *unobservable* variables that determine the selling price can be different for local and non-local buyers. Note that the covariance between $\epsilon_{0i}$ and $\epsilon_{1i}$ , $\sigma_{01}$, cannot be estimated because the potential outcomes $p_{0i}$ and $p_{1i}$ are never observed simultaneously. However, identification of $\sigma_{01}$ is not necessary to estimate the other parameters (see Maddala 1983, p. 224). Additionally, the unobservable variables governing the bargaining/selection process $(u_i)$ may be correlated with the unobservable variables governing the price ($\epsilon_{0i}$ and/or $\epsilon_{1i}$); $\lambda_0 = \sigma_0\rho_0$ and $\lambda_1 = \sigma_1\rho_1$ are the covariances between $\epsilon_0, u$ and $\epsilon_1, u$ ($\rho_0$ and $\rho_1$ being the correlations).

To sump-up, this very general model allows to capture observable preference heterogeneity, by allowing $\beta_0$ and $\beta_1$, the coefficients associated with the characteristics of the land plots, to be different depending on whether the buyer is local ($\beta_0$) or non-local ($\beta_1$). This model also allows capturing unobservable preference heterogeneity, by allowing $\sigma_0$ and $\sigma_1$ to be different, and by allowing the unobservables governing the price and the selection process to be correlated. The model can be considered as an endogenous treatment regression model, $nonres_i$ being the endogenous 'treatment' variable. Therefore, the average treatment effect $(ATE)$ is

$$
ATE = E(p_{1i} - p_{0i}) = E\left\{x_i(\beta_1 - \beta_0)\right\} \tag{2}
$$

The $ATE$ measures the average difference in the potential selling price for non-local buyers and the potential selling price for local buyers, i.e., the price gap $(\mu)$ between non-local and local buyers. The model is estimated by maximum likelihood (see Mad-

dala 1983, p. 117-122 for the derivation of the likelihood function)[6].

### *3.3. Results of the potential outcome model*

Note that, under our normality assumptions, no exclusion restriction regressors are required to identify the mean treatment effects (see, e.g., Aakvik, Heckman, and Vytlacil, 2005, p. 34). However, we introduce one variable in the *nonres* equation, which can be viewed as a plausible exclusion restriction: *nourb* (the land plot is preserved from urbanization) is potentially an important determinant of the selection process, while being non-significantly correlated with *price*[7]. The estimation of the model was much more stable when using this exclusion restriction.

In order to ensure that the results of potential model can be compared with the hedonic pricing model with observable heterogeneity (estimated by OLS), we first estimate a constrained version ($\sigma_0 = \sigma_1$). Thus, the constrained version of the model allows for observable heterogeneity but relies, as the hedonic pricing model with interaction variables, on the assumption of no unobservable heterogeneity. In that case, too, there is only one parameter $\rho$ to estimate. Thus, testing the exogeneity of the *nonres* variable in that constrained version of the potential outcome model amounts to testing $\rho = 0$. A Wald test indicates that we cannot reject the null hypothesis ($\chi^2(1) = 0.12$, $p - value = 0.728$), thus suggesting that the unobservable variables in the price equations are not correlated with the unobservable variables governing the selection process. Thus, as we applied the Oster (2017) approach to the simple hedonic pricing models, we find no evidence of selection bias due to unobservables. As can be seen from Table 6, we also find evidence of observable heterogeneity, *i.e.* differences in the way locals and non-locals weight plot characteristics, which we will comment in more details for the unconstrained version of the potential outcome model. More importantly, from the constrained version of the potential outcome model, we find a treatment effect of about €8.49 per square meter (95% confidence interval: [5.57,11.42]), again very close to the treatment effect derived from the hedonic pricing model with observable heterogeneity. The estimation of the constrained model therefore confirms the robustness of the results obtained from the simple hedonic pricing models under the assumption of *no unobservable heterogeneity*. Precisely, what happens when we relax this assumption and estimate the unconstrained version of the potential outcome model?

First, according to the Bayesian Information Criteria (BIC), the unconstrained version of the potential outcome model clearly fits better the data than the constrained version ($BIC = 49555.05$ *versus* 49891.76). Note that, in the context of the unconstrained version of the potential outcome outcome model, this result basically means that the unobservables governing the selection process are neither correlated with the unobservables governing price for the locals, nor with those governing price for the non-locals.

Second, it should be noted that testing the exogeneity of the *nonres* variable in the unconstrained version of the potential outcome model now amounts to testing the following null hypothesis: $\rho_0 = 0$ and $\rho_1 = 0$. Here, a Wald test indicates that we cannot reject the null hypothesis ($\chi^2(2) = 4.28$, $p - value = 0.117$). Again, we find no evidence of selection bias due to unobservables.

[6]The command is implemented in Stata MP 15 as etregress.

[7]The correlation between *nourb* and *price* is -0.01 and not significant.

Additionally, the *nonres* equation provides some insights that merit consideration. It should be recalled that we have no information regarding the buyers' characteristics, such as gender, income, exact place of residence, etc., nor do we have any information regarding the sellers. Thus, $u_i$ in the *nonres* equation, actually captures the effects of these unobservable variables, while the $w_i$ reveals some observable variables that count in the selection process, notably because non-local and local buyers may have very different preferences underlying their decision. As stated in the introduction, non-local buyers are mainly in search of a second home, while local buyers are in search of a principal residence near employment areas.

Three highly significant and positive regional indicators were identified (Southern, Valinco, Balagna). However, non-local buyers seem to avoid crowded places (*city* is negative) and look for preserved sites (*nourb* is positive) probably in order to build a second home (*developed*, which captures the fact that the land plot is fully developed, is significant and positive).

Turning next to the price equations, the $\beta_1$-vector is reported in column 2 of Table 7 and the $\beta_0$-vector is reported in column 3 of Table 7. The estimate of the price error standard deviation parameter for the non-local group ($\sigma_1$) is clearly[8] larger than that of the domestic group parameter ($\sigma_0$), indicating a greater variability in the unobservable variables among the non-local group, providing evidence of unobservable preference heterogeneity[9]. It should first be noted that most of the year indicator coefficients are statistically significant. The reference year is 2008, and the negative coefficients associated with previous years reflect a positive trend in the land plot price (in constant 2008 euros). Both local and non-local buyers significantly and positively value garden-type land plots, reflecting the fact that gardens are easy to convert in order to build a house. Conversely, the scrubland and wilderness types of land plot are less highly valued. Additionally, local and non-local buyers respond heterogeneously to some of the land plot characteristics, which reflects observable preference heterogeneity (differences in $\beta's$). For example, the 'wood type' coefficient is significant and negative for local buyers, while non-significant for non-local buyers. Similarly, the southern area (positive and significant coefficient) is valued by non-local but not by local buyers, and the eastern area is found to be unattractive to local buyers, while leaving non-local buyers indifferent.

We next computed the average treatment effect, i.e., the *average causal effect* of being a non-local buyer on the selling price (the price gap). We found that

$ATE = E\{x_i(\beta_1 - \beta_0)\} =$€12.32 per square metre with a 95% confidence interval of [€6.68, €17.96].

Thus, allowing for unobservable heterogeneity results in two main effects. First, the magnitude of the treatment effect is significantly higher when we allow for unobservable heterogeneity than when we do not allow for unobservable heterogeneity. Part of the variability which was constrained by a common variance parameter is now more subtly captured by two variance parameters, which induces changes in the estimated coefficients and thus change in the estimated treatment effect. Second, in the case of our data the treatment effect is less precisely estimated when we allow for unobservable heterogeneity, due to the fact that variability is found greater for non

[8]The null hypothesis that $\sigma_0 = \sigma_1$ is unambiguously rejected ($\chi^2(1) = 51.46$, $p - value = 0.000$).

[9]Note that this reflects the greater variability in the sale price for non-local buyers compared to local buyers, see standard deviations in Table 2.

locals than for locals ($\sigma_1 > \sigma_0$). We believe that the results from the unconstrained version of the potential outcome model, given its generality, better describe the essence of the price gap between local and non-local buyers than results from models allowing only for observable heterogeneity.

Beyond that, the analysis is enriched by computing the marginal effects of the observable land plot characteristics on the average treatment effect. The marginal effects on the average treatment effect are reported in column 4 of Table 7. Very few marginal effects are found to be significant. If we reflect back to the definition of the average treatment effect, the marginal effect of a variable is unlikely to be statistically significant when the difference between $\beta_1$ and $\beta_0$ for this variable is actually not significant. Notably, the marginal effects of the year indicators are highly significant only for three years, mainly concentrated at the beginning of the study period (1998 and 1999), thus suggesting that price trends have now converged. Among the land plot characteristics, three locations (Ajaccio area, southern area, Valinco area) significantly and positively impact the $ATE$. The fact that the land plot is fully developed is associated with the largest marginal effect on the $ATE$ (about €14.06). As noted above, developed land plots are attractive to non-locals, as such land plots are available for immediate construction of a second home.

## 4. Conclusion

This paper has investigated the issue of local buyers' exclusion through price in the land market. The case of the seaside farmland market of Corsica between 1998 and 2008 was studied. This market is of special interest since the data show the existence of a price gap between local and non-local buyers. Comparing estimates from standard OLS hedonic models to those from a general potential outcome model, we show that the latter better capture observable and unobservable heterogeneity in preferences. An *average causal effect* of being a non-local buyer on the sale price of roughly €12 per square metre was found, which can be confidently related to observable and unobservable preference heterogeneity between local and non-local buyers. Our findings contribute to the debate on the consequences of the presence of non-local buyers in the land market: local and non-local buyers value the characteristics of land plots differently, suggesting a market segmentation effect rather than full exclusion of local buyers.
However, the question of the incidence of a large influx of non-local buyers into land and housing markets on market segmentation remains understudied. From that point of view, our paper paves the way for future research in at least two directions. First, there is a need for replication studies using data from other regions experiencing strong external demand pressure in order to confirm the general validity of our findings. Second, it is important to understand the mechanisms leading sellers to contract with local buyers at lower prices. The literature (e.g. Lambson, McQueen, and Slade, 2004; Turnbull and Sirmans, 1993; Levitt and Syverson, 2008) suggests that the explanation could lie in the existence of search costs and in time on the market, without our data allowing us to properly assess the relevance of this explanation. Building a database including good proxies for these variables could help us to deepen our understanding of sale agreements and market segmentation in these specific markets.

## References


Aakvik, A., J.J. Heckman, and E.J. Vytlacil. 2005. "Estimating treatment effects for discrete outcomes when responses to treatment vary: An application to Norwegian vocational rehabilitation programs." *Journal of Econometrics* 125: 15–51.

Chao, Chi-Chur, and Eden S. H. Yu. 2015. "Housing Markets with Foreign Buyers." *Journal of Real Estate Finance and Economics* 50: 207–218.

Davidson, Isaac. 2016. "John Key: Land tax plan could happen by 2017." *New Zealand Herald, April 26 2016* .

Heckman, J.J. 2010. "Building bridges between structural and program evaluation approaches to evaluating policy." *Journal of Economic Literature* 48 (2): 356–398.

Holland, P.W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–960.

INSEE. 2015. "Dossier Corse n °2 : Le poids du tourisme dans l'économie corse." INSEE. Octobre.

King, David A., and J. A. Sinden. 1994. "Price Formation in Farm Land Markets." *Land Economics* 70 (1): 38–52.

Lambson, Val E., Grant R. McQueen, and Barrett A. Slade. 2004. "Do Out-of-State Buyers Pay More for Real Estate? An Examination of Anchoring-Induced Bias and Search Costs." *Real Estate Economics* 32 (1): 85–126.

Levitt, Steven D., and Chad Syverson. 2008. "Market distorsions when agents are better informed: the value of information in real estate transactions." *The Review of Economics and Statistics* 90 (4): 599–611.

Levkovich, Or, Jan Rouwendal, and Lars Brugman. 2018. "Spatial Planning and Segmentation of the Land Market: The Case of the Netherlands." *Land Economics* 94 (1): 137–154.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press. Cambridge Books Online.

Miller, Norman G., Michael A. Sklarz, and Nicholas Ordway. 1988. "Japanese Purchases, Exchange Rates and Speculation in Residential Real Estate Markets." *Journal of Real Estate Research* 3 (3): 39–49.

Nicholls, Sean. 2016. "NSW Budget 2016: Foreign property buyers in NSW to be hit with stamp duty and land tax hikes." *New South Wales News, June 14 2016* .

Oster, Emily. 2017. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business and Economic Statistics* .

Pearl, J. 2012. *Handbook of Structural Equation Modeling*, Chap. The Causal Foundations of Structural Equation Modeling, 68–91. New-York.: Guilford Press.

Sturgeon, Jamie. 2015. "Seeds of crisis? Historic boom for farmland raises red flags." *Globalnews.ca, June 3 2015* .

Turnbull, GK, and CF Sirmans. 1993. "Information, search and house prices." *Regional Science and Urban Economics* 23 (4): 545–557.

Williams, Zoe. 2015. "My fix for the housing crisis: ban ownership by foreign non-residents." *The Guardian, June 22 2015* .

Zhou, Xiaorong, Karen Gibler, and Velma Zahirovic-Herbert. 2015. "Asymmetric buyer information influence on price in a homogeneous housing market." *Urban Studies* 52 (5): 891–905.

| | Controlled Treatment Effect | Controlled Treatment Effect with Interactions |
|---|---|---|
| nonres | 8.769*** | 6.410 |
| | (0.877) | (4.1266) |
| developed | 17.069*** | 10.611*** |
| | (2.128) | (2.271) |
| nonresXdeveloped | - | 15.445*** |
| | | (4.381) |
| 1998 | -19.586*** | -15.457*** |
| | (1.707) | (1.725) |
| 1999 | -19.184*** | -14.785*** |
| | (1.582) | (1.536) |
| 2000 | -13.194*** | -10.860*** |
| | (1.719) | (1.610) |
| 2001 | -14.439*** | -12.340*** |
| | (1.586) | (1.479) |
| 2002 | -12.480*** | -9.937*** |
| | (1.589) | ( 1.464) |
| 2003 | -9.752*** | -6.925*** |
| | (1.487) | (1.498) |
| 2004 | -10.739*** | -8.582*** |
| | (1.645) | (1.630) |
| 2005 | -6.860*** | -5.554** |
| | ( 1.586) | ( 1.612) |
| 2006 | -5.076** | -2.899 |
| | (1.607) | (1.628) |
| 2007 | -1.852 | -2.030 |
| | (1.677) | (1.596) |
| nonresX1998 | - | -13.599** |
| | | (4.551) |
| nonresX1999 | - | -15.981*** |
| | | (4.257) |
| nonresX2000 | - | -8.099 |
| | | (4.868) |
| nonresX2001 | - | -7.250 |
| | | (4.261) |
| nonresX2002 | - | -8.497 |
| | | (4.424) |
| nonresX2003 | - | -10.420* |
| | | (4.057) |
| nonresX2004 | - | -7.791 |
| | | (4.453) |
| nonresX2005 | - | -4.591 |
| | | (4.270) |
| nonresX2006 | - | -8.082 |
| | | (4.414) |
| nonresX2007 | - | 0.761 |
| | | (4.705) |
| Meadow | -1.216 | -1,803 |
| | (1.012) | (1.053) |
| Vineyard | -6.197*** | -5.026** |
| | (1.330) | (1.490) |
| Scrubland | -6.274*** | -5.217** |
| | (1.466) | (1.624) |
| Wilderness | -6.735*** | -5.824*** |
| | (1.250) | (1.248) |
| Wood | -4.045** | -5.288*** |
| | (1.387) | (1.312) |
| Garden | 21.510*** | 20.919*** |
| | (3.532) | (3.961) |
| nonresXMeadow | - | 2.110 |
| | | (2.606) |
| nonresXVineyard | - | -4.838 |
| | | (2.994) |

| | | |
|---|---|---|
| nonresXScrubland | - | -2.318 |
| | | (3.330) |
| nonresXWilderness | - | -2.103 |
| | | (3.322) |
| nonresXWood | - | 5.473 |
| | | (3.780) |
| nonresXGarden | - | 2.872 |
| | | (8.457) |
| Ajaccio | 9.812*** | 7.582*** |
| | (1.245) | (1.280) |
| South | 4.393*** | 1.628 |
| | (0.972) | (1.006) |
| Valinco | -2.331 | -5.510*** |
| | (1.607) | (1.523) |
| Sevi-Sorru | -2.677** | -2.790* |
| | (1.179) | (1.325) |
| Balagna | 1.202 | 0.721 |
| | (1.118) | (1.122) |
| Eastern Corsica | -3.179*** | -3.155*** |
| | (0.811) | (0.803) |
| nonresXAjaccio | - | 10.663** |
| | | (3.522) |
| nonresXSouth | - | 10.994*** |
| | | (2.574) |
| nonresXValinco | - | 11.403** |
| | | (3.588) |
| nonresXSevi-Sorru | - | 3.411 |
| | | (2.861) |
| nonresXBalagna | - | 5.058 |
| | | (2.892) |
| nonresXEastern Corsica | - | 1.090 |
| | | (2.460) |
| city | 4.979*** | 4.860*** |
| | (0.872) | (0.864) |
| nonresXcity | - | 0.612 |
| | | (2.553) |
| constant | 18.566*** | 18.561*** |
| | (1.492) | (1.457) |
| N | 4827 | 4827 |
| $R^2$ | 0.180 | 0.203 |
| F-statistic | 25.81*** | 15.48*** |

Robust standard errors in parentheses
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 5.: Results from standard OLS hedonic models without and with interaction variables

| | Selection equation | Price equation (non-locals) | Price equation (locals) |
|---|---|---|---|
| nourb | 0.330*** | - | - |
| | (0.040) | - | - |
| developed | 0.450*** | 25.983*** | 10.557*** |
| | (0.073) | (3.755) | (2.259) |
| 1998 | -0.028 | -29.053*** | -15.455*** |
| | (0.128) | (4.189) | (1.716) |
| 1999 | 0.073 | -30.778*** | -14.792*** |
| | (0.115) | (3.949) | (1.528) |
| 2000 | 0.031 | -18.969*** | -10.864*** |
| | (0.102) | (4.552) | (1.601) |
| 2001 | 0.250** | -19.526*** | -12.368*** |
| | (0.089) | (3.986) | (1.475) |
| 2002 | 0.189* | -18.466*** | -9.958*** |
| | (0.090) | (4.151) | (1.456) |
| 2003 | 0.073 | -17.355*** | -6.932*** |
| | (0.079) | (3.752) | (1.488) |
| 2004 | 0.023 | -16.376*** | -8.583*** |
| | (0.087) | (4.123) | (1.621) |
| 2005 | 0.066 | -10.154** | -5.561** |
| | (0.082) | (3.933) | (1.601) |
| 2006 | -0.055 | -10.968** | -2.893 |
| | (0.082) | (4.082) | (1.619) |
| 2007 | 0.073 | -1.275 | -2.037 |
| | (0.080) | (4.403) | (1.588) |
| No predominant feature | -0.140* | 0.327 | -1.787 |
| | (0.055) | (2.374) | (1.046) |
| Vineyard and orchard | -0.086 | -9.848*** | -5.015** |
| | (0.112) | (2.584) | (1.482) |
| Scrubland | 0.182 | -7.570** | -5.239** |
| | (0.101) | (2.885) | (1.621) |
| Wilderness | -0.007 | -7.935** | -5.824*** |
| | (0.084) | (3.062) | (1.242) |
| Wood | -0.051 | 0.193 | -5.278*** |
| | (0.075) | (3.528) | (1.306) |
| Garden | -0.049 | 23.792*** | 20.927*** |
| | (0.122) | (7.435) | (3.941) |
| Ajaccio area | 0.106 | 18.223*** | 7.573*** |
| | (0.065) | (3.257) | (1.269) |
| Southern area | 0.570*** | 12.525*** | 1.564 |
| | (0.059) | (2.366) | (1.030) |
| Valinco area | 0.719*** | 5.769 | -5.594 |
| | (0.088) | (3.334) | (1.524) |
| Sevi-Sorru area | 0.243* | 0.563 | -2.828* |
| | (0.096) | (2.527) | (1.322) |
| Balagna area | 0.316*** | 5.722* | 0.680 |
| | (0.072) | (2.657) | (1.128) |
| Eastern area | 0.067 | -2.059 | -3.154*** |
| | (0.069) | (2.311) | (0.798) |
| city | -0.461*** | 5.541** | 4.906*** |
| | (0.049) | (2.396) | (0.867) |
| constant | -0.907*** | 25.280*** | 18.470*** |
| | (0.080) | (3.951) | (1.472) |
| $\rho$ | -0.010 | | |
| | (0.029) | | |
| $\sigma$ | 22.695*** | | |
| | (0.595) | | |
| $\lambda$ | -3.824 | | |
| | (0.670) | | |
| $N$ | 4827 | | |

| | Selection equation | Price equation (non-locals) | Price equation (locals) |
|---|---|---|---|
| Log-likelihood | -24615.08 | | |
| BIC | 49891.76 | | |

Robust standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 6.: The potential-outcome model - constrained model

| | Selection equation | Price equation (non-locals) | Price equation (locals) | Marg. effects on $ATE$ |
|---|---|---|---|---|
| nourb | 0.332*** | - | - | 0.813 |
| | (0.040) | - | - | (0.422) |
| developed | 0.452*** | 24.860*** | 10.816*** | 14.065** |
| | (0.073) | (3.813) | (2.240) | (4.456) |
| 1998 | -0.028 | -29.020*** | -15.463*** | -13.646** |
| | (0.128) | (4.187) | (1.717) | (4.529) |
| 1999 | 0.073 | -30.967*** | -14.756*** | -15.981*** |
| | (0.115) | (3.965) | (1.529) | (4.338) |
| 2000 | 0.031 | -19.133*** | -10.840*** | -8.187 |
| | (0.102) | (4.571) | (1.601) | 4.831 |
| 2001 | 0.250** | -20.200*** | -12.228*** | -7.860 |
| | (0.089) | (4.034) | (1.455) | (4.284) |
| 2002 | 0.188* | -18.970*** | -9.858*** | -9.003* |
| | (0.090) | (4.160) | (1.455) | (4.402) |
| 2003 | 0.071 | -17.508*** | -6.894*** | -10.387** |
| | (0.079) | (3.766) | (1.486) | (4.033) |
| 2004 | 0.023 | -16.424*** | -8.575*** | -7.774 |
| | (0.087) | (4.131) | (1.621) | (4.428) |
| 2005 | 0.067 | -10.296** | -5.524** | -4.558 |
| | (0.082) | (3.944) | (1.602) | (4.242) |
| 2006 | -0.056 | -10.776* | -2.917 | -8.041 |
| | (0.082) | (4.093) | (1.619) | (4.388) |
| 2007 | 0.074 | -1.383 | -2.001 | 0.854 |
| | (0.080) | (4.411) | (1.58) | (4.682) |
| No predominant feature | -0.140* | 0.644 | -1.862 | 2.496 |
| | (0.055) | (2.384) | (1.042) | (2.602) |
| Vineyard and orchard | -0.087 | -9.616*** | -5.065** | -4.580 |
| | (0.111) | (2.686) | (1.481) | (2.995) |
| Scrubland | 0.182 | -8.123** | -5.130** | -2.999 |
| | (0.101) | (2.884) | (1.628) | (3.319) |
| Wilderness | -0.008 | -8.062** | -5.818*** | -2.245 |
| | (0.084) | (3.073) | (1.241) | (3.313) |
| Wood | -0.052 | 0.306 | -5.321*** | 5.624 |
| | (0.075) | (3.537) | (1.307) | (3.772) |
| Garden | -0.050 | 23.830*** | 20.885*** | 2.933 |
| | (0.122) | (7.459) | (3.942) | (8.436) |
| Ajaccio area | 0.104 | 17.898*** | 7.614*** | 10.281 ** |
| | (0.065) | (3.245) | (1.267) | (3.484) |
| Southern area | 0.569*** | 11.037*** | 1.868 | 9.193*** |
| | (0.059) | (2.451) | (1.027) | (2.663) |
| Valinco area | 0.717*** | 3.863 | -5.190 | 9.083* |
| | (0.088) | (3.321) | (1.505) | (3.650) |
| Sevi-Sorru area | 0.241* | -0.334 | -2.639* | 2.318 |
| | (0.095) | (2.580) | (1.320) | (2.900) |
| Balagna area | 0.316*** | 4.865 | 0.873 | 4.018 |
| | (0.072) | (2.698) | (1.133) | (2.928) |
| Eastern area | 0.068 | -1.974 | -3.160*** | 1.183 |
| | (0.069) | (2.318) | (0.798) | (2.453) |
| city | -0.460*** | 6.614** | 4.682*** | 1.910 |
| | (0.049) | (2.455) | (0.853) | (2.600) |
| constant | -0.908*** | 30.044*** | 18.905*** | - |
| | (0.080) | (4.826) | (1.479) | - |
| $\rho_0$ | 0.045 | | | |
| | (0.039) | | | |
| $\rho_1$ | -0.129 | | | |
| | (0.070) | | | |
| $\sigma_0$ | 19.526*** | | | |
| | (0.674) | | | |
| $\sigma_1$ | 29.526*** | | | |

| | Selection equation | Price equation (non-locals) | Price equation (locals) | Marg. effects on $ATE$ |
|---|---|---|---|---|
| | (1.203) | | | |
| $\lambda_0$ | 0.883 | | | |
| | (0.767) | | | |
| $\lambda_1$ | -3.824 | | | |
| | (2.138) | | | |
| $N$ | 4827 | | | |
| Log-likelihood | -24438.25 | | | |
| BIC | 49555.05 | | | |

Robust standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 7.: The potential-outcome model - unconstrained model