# More than Winners and Losers: The Consequences of Price Deviation on Capital Gains Formation

Rémi Lei*
remi.lei@psemail.eu

Laure Casanova Enault†
laure.casanova@univ-avignon.fr

Antoine Peris†
antoine.peris@univ-avignon.fr

August 2025

**Abstract**

Whereas location is considered the main driver for capital gains formation, we assess how individuals shape their wealth accumulation. Making the best of the universe of transactions in France between 2012 and 2021, we estimate group and individual price deviations being orthogonal to housing characteristics through a repeat sales approach with *Heckman* procedure to account for potential selection bias. Then, we regress realized capital gains on the individual deviation and homeowners' characteristics to assess how much of the capital gains is explained by individuals' characteristics, controlling for housing location. Whereas experience and financial constraints play a major role in explaining both purchase price and capital gains, we estimate that an individual purchase price deviation of €1 is associated with a decrease of capital gains by €0.41. Drivers such as individual efficiency, inattention, and financial literacy are likely to explain these differences.

**JEL classification:** R2 ; R3 ; D1

**Keywords:** Purchaser behaviour ; wealth accumulation ; repeat sales ; spatial confounding

*EU Tax Observatory, Paris School of Economics, F-75014 Paris, France

# 1 Introduction

Housing represents the main asset of households' portfolio. The evolution of housing price conditions affects wealth accumulation, access to housing, and wealth inequality. Understanding the underlying mechanisms of price formation, local housing market dynamics, and homeowners' portfolio decisions is fundamental to better understanding how housing shapes wealth inequality. One of the major sources of housing appreciation is location (Kiel and Zabel, 2008; Wind and Hedman, 2018). Factors such as land availability, job attractiveness, natural amenity, or transport infrastructure are known to affect housing prices and market dynamics. Hence, homeowners are mostly considered as winners and losers on the housing market, depending on the location of the housing unit they acquire (Eggum and Røed Larsen, 2024) and how their local environment evolves. It implicitly assumes that homeowners cannot condition their wealth accumulation, apart from anticipating potential housing market reversal (Martínez-Toledano, 2022). Nonetheless, besides location and housing market dynamics, both the purchase and selling processes can influence the price depending on the bargaining power, which might condition capital gains and wealth accumulation as well.

Among numerous aspects, the experience and the financial constraints faced by both purchasers and sellers are likely to shape the exchange value. Less experienced individuals cannot adjust their bidding price based on past purchases. We expect the most experienced individuals to better value housing units and bid accordingly. In addition, the financial constraints and individual timing will affect bargaining intensity and potentially exchange price accordingly. Despite some research on the tenure premium (Biagi et al., 2021), distant purchasers (Ling, Naranjo and Scheick, 2021), or the role of specific purchasers in housing market such as Small and Medium Real Estate Investors (SMREI) (Garriga, Gete and Tsouderou, 2023), the role of experience, financial constraints and individual efficiency in the housing market remains understudied to understand the formation of capital gains. Moreover, the exchange value for housing units may condition price beliefs for future purchasers and sellers, and then affect the local housing market dynamics (Szumilo, 2021).

In this paper, we fill two gaps in the current housing literature. First, we extend previous research on the impact of purchasers' characteristics on the formation of housing prices. We adopt a new classification that accounts for experience and financial constraints. It includes common categories such as small investors, while it includes new levels such as first-time owners or consumer-investors (for more detail, see Section 2). Second, we link the individual purchase price deviation to the relative realized capital gains. We isolate an individual component of the purchase price that is orthogonal to housing characteristics, experience, and financial constraints using housing fixed effects regression. This

component is then introduced to explain the relative realized capital gains, in addition to housing characteristics and the purchaser's category. Whereas we cannot interpret these results causally, we highlight that individuals shape their purchase price and capital gains accordingly.

Our paper exploits a dataset about the universe of housing transactions in France that occurred between 2012 and 2022. It comes from administrative sources with detailed characteristics about housing units, such as floor area or location, with high spatial granularity. An invariant housing ID allows us to track multiple sales over time and compute realized capital gains. In addition, we make the best of Lei et al. (2024) improvement of French administrative data about homeowners to link housing exchange to purchasers and sellers. We characterize individuals according to their portfolio size (i.e., number of properties), municipality of residence, or age. We leverage the longitudinal dimension of these administrative data about homeowners to observe whether purchasers are simultaneously selling one of these properties to finance the purchase. These portfolio characteristics enable us to construct the new purchaser category that proxies for experience and financial constraints.

The main threat to the estimation of purchase price deviation, both at the individual and group level, is sorting into unobservable housing characteristics (Halket, Nesheim and Oswald, 2020). Despite the recent developments in the hedonic method of modelling non-linear relationships (see for instance Wood, 2017; Khoshnoud, Sirmans and Zietz, 2023), this approach still assumes that unobservable characteristics are randomly distributed across purchaser groups. In practice, it implicitly states that thermal insulation, amenities views, or quality of construction are similar for first-time owners, real estate investors, or secondary home purchases, which is very unlikely to hold. First-time owners smooth their consumption to access homeownership, considering the housing affordability crisis (Waxman et al., 2020), which might downgrade their housing quality as well. Hence, recovering group-specific price deviations from a hedonic method is not adequate.

We rather adopt a repeat sales approach that enables us to introduce housing fixed effects (for application, see e.g. Dorsey et al., 2010). It relaxes the assumption that unobservable characteristics are randomly distributed between purchasers' categories and provides group and individual deviations that are orthogonal to invariant housing characteristics. The drawback of the introduction of housing fixed effects is the sample restriction to units being sold at least twice between 2012 and 2022. We account for the potential selection bias through the estimation of the probability of each housing transaction being sold at least twice. We introduce the inferred probability as a control variable in the main regression *à la Heckman* (Heckman, 1979), although we check Inverse Probability Weighting (IPW) as an alternative method. The selection model includes, in addition to common observable housing characteristics, a spatial smoothing function using transaction coordin-

ates to account for spatial heterogeneity, including non-observable one (Dupont, Marques and Kneib, 2023; Gilbert et al., 2024). Considering the error term as the individual component, we regress realized capital gains on the individual deviation, the purchasers' category, and housing characteristics, including tenure. It links the relative capital gains to the individual efficiency on the housing market, experience, and financial constraints.

We estimate that a €1 individual purchase price deviation being orthogonal to housing characteristics translates on average to a lower capital gain by €0.41. Hence, local housing market dynamics cannot be considered as the only driver for capital gains and wealth accumulation. We expect drivers such as inattention, financial literacy (Behrman et al., 2012), timing decision, and bargaining intensity to drive these results, although we cannot disentangle each channel. In addition, experience and financial constraints proxied by our purchaser category significantly affect purchase price. The less experienced purchasers (i.e., first-time owners) pay the highest price (+4.5% compared to homeowners achieving a residential mobility). On the contrary, the most experienced and less financially constrained individuals pay the lowest price. Finally, we estimate a tenure discount of 3.2% for a buy-to-rent purchase compared to an owner-occupied purchase. We expect investors to bid more aggressively in the bargaining process as they seek profitability. However, despite the purchase discount, buy-to-rent housing units generate lower relative capital gains by 1.4pp. Our results highlight that the housing location is not the only driver for capital gains and wealth accumulation.

Our contribution to the existing literature is threefold. First, we contribute to the literature focusing on housing wealth inequality through the formation of capital gains. Whereas measurement errors in self-assessed value are widely studied with impact on inequality measures (Tur-Sinai, Fleishman and Romanov, 2020), we demonstrate that both group and individual price deviations affect the realization of capital gains. The wealthiest homeowners better anticipate housing in turn (Martínez-Toledano, 2022) and benefit from both the largest price increase after their purchase (Aiello, Kotter and Schubert, 2022) and capital returns (Jordà et al., 2019; Bach, Calvet and Sodini, 2020; Fagereng et al., 2020). We uncover that all purchasers affect their capital gains through their behaviour. While we do not estimate spillover effects resulting from the signalling effect, our results are consistent with Szumilo (2021) ones. Considering the impact of capital gains on self-employment (Harding and Rosenthal, 2017), wealth transmission (Daysal, Lovenheim and Wasser, 2023), consumption (Ben-Shahar, Gabriel and Golan, 2019), or even income inequality (Roine and Waldenström, 2012), these individual deviations are likely to condition life-cycle outcomes. Second, we contribute to the literature focusing on the determinants of purchase price being orthogonal to housing characteristics. Whereas previous papers have focused on the role of tenure status (Turnbull and Van Der Vlist, 2022), information asymmetries (Li and Chau, 2023) based on distance (Biagi et al.,

2021; Cvijanović and Spaenjers, 2021; Favilukis and Van Nieuwerburgh, 2021), or the small real estate investors (Wong, Deng and Chau, 2022; Garriga, Gete and Tsouderou, 2023), we provide new estimation that accounts for experience and bargaining power for a broader population of homeowners. Our classification is the first to include the less experienced purchasers (namely the first-time owners), known to be a relevant category to understand the housing wealth inequality (Turner and Luea, 2009; Wainer and Zabel, 2020). We highlight that the less experienced and financially constrained purchasers are more likely to pay a higher price, which is consistent with the bargaining intensity and inattention mechanisms. Finally, accounting for unobservable characteristics of housing units and purchaser characteristics, we find that buy-to-rent purchases are more likely to bargain as they pay at a lower price compared to owner-occupied ones ($-3.2\%$ for rental purchase compared to owner-occupied ones). Our estimations are opposed to those obtained by Biagi et al. (2021). We expect the source of heterogeneity within owner-occupier purchasers to drive these differences. Third, we contribute to the literature focusing on homeownership development. Considering the consequences of the affordability crisis (see e.g. Favilukis, Mabille and Van Nieuwerburgh, 2021), renters have increasing difficulty transitioning into homeownership, despite trading off consumption for savings (Waxman et al., 2020). Our results provide two new insights. Potential first-time owners are likely to bargain less intensively, which affects the financial burden in the long term due to the leveraging effect of interests. In addition, they smooth housing quality to achieve their transition to homeownership. Whereas demand-side policies supporting homeownership are mostly inefficient (Bäckman and Lutz, 2020; Braakmann and McDonald, 2020; Kunovac and Zilic, 2022), further investigations are needed to determine mechanisms to mitigate these differences and favour homeownership development.

The paper proceeds as follows. We first detail the transaction dataset we exploit in Section 2. We then detail our identification strategy to recover group and individual purchase price deviations and link them to capital gains realization in Section 3. Then, we present our results in Section 4, including robustness checks, and discuss potential underlying mechanisms and related consequences in Section 5. We conclude in Section 6.

# 2 Data

We take advantage of two distinct raw datasets to recover price differences and their impact on capital gains formation. These data sources improve the qualification of purchasers, extending the traditional categories that distinguish owner-occupiers, buy-to-rent, and buy-to-let investments.

## 2.1 Transaction Dataset

Our main dataset to study the formation of capital gains is derived from the DV3F database, which has registered housing transactions for France since 2010 in a comprehensive manner. For each housing transaction, we observe structural characteristics such as floor area, housing type, location with high granularity, or the presence of certain facilities (basement, parking space, etc.). We can track housing units over time using the invariant ID and identify multiple sales over time (see in Appendix A.1). In addition, characteristics of the transaction are provided, such as the agreement date and the exchange price. This joint presence of invariant ID and purchase price enables us to compute the universe of realized capital gains between 2012 and 2023.

However, the dataset lacks information about purchaser and seller characteristics. The current transaction dataset allows disentangling the purchaser based on whether it is a natural or a legal person. This information is not sufficient to control for experience and financial constraints, two variables that are likely to affect purchase price through bargaining power.

## 2.2 Property Tax Files

To address the lack of information about purchasers in the transaction dataset, we exploit the French property tax files. These data link housing units to their current owners annually since 2011. Using the recent development that turns these data into panel data (Lei et al., 2024), we identify the purchasers (homeowners the year following the purchase) and the sellers (homeowners the year preceding the purchase) at the transaction level. We merge this dataset with the transaction dataset using the common housing ID.

The information about purchasers is available at the individual level, while multiple individuals can purchase similar housing units at the same time (for instance, a household that purchases a home). As wealth becomes more individualized over time (Frémeaux and Leturcq, 2020), the number of held properties before the purchase is not always homogeneous within the purchasers' group. Rather than aggregating purchasers' characteristics at the transaction level, we choose to keep the full raw data and weight transactions according to the number of purchasers and sellers. This is particularly convenient for the purchaser definition, as it is not always consistent among purchasers at the transaction level.

## 2.3 Portfolios-Based Categories

Our purchaser definition aims to proxy for experience and financial constraints faced at the time of the purchase. We leverage the set of information about the size of the housing portfolio and the joint sale of the current housing portfolio to finance the new purchase as key information to classify individuals. The number of housing units held before the purchase is a good proxy for experience, as it indicates the knowledge of the purchasing process for an individual. The underlying assumption is that individuals with many housing units are more experienced, whereas individuals that never gone through this process are the less experienced ones. In addition, the simultaneous sale of a housing unit held in the portfolio is a good proxy for financial constraint. We consider that if individuals do not sell any of their housing assets simultaneously or in the previous year, they are more likely to finance their purchase through debt. On the contrary, if individuals sell a housing unit or multiple housing units, they are more likely to reuse the gain from the sale to finance their new purchase. Consequently, we consider this variable as a proxy for financial constraints.

Based on these two variables, we construct an eight-level classification: first-time owners, single purchaser-sellers, which are likened to homeowners achieving residential mobility, multiple purchaser-sellers, small investors, big investors, i.e., purchasers with respectively less than five and more than five housing units and increased their portfolio size, and consumer-investor, i.e., purchaser that sells more properties than they purchase. Our classification does not overlap strictly with previous classifications in the literature, which focus specifically on purchasers' location and tenure status. Yet, we introduce the post-purchase tenure status in the main regression to control for differences in behaviour attributable to the purchase objective.

## 2.4 Descriptive Statistics

Our final dataset is composed of 2,251,773 observations from 2012 to 2021. We remove from our sample observations-related legal persons as they are out of scope. Most housing transactions are achieved by first-time owners (nearly 45% of the sample, for more details, see Figure 1), followed by small investors and simple buyer–seller (i.e., mostly residential mobility of homeowners). However, their choices differ significantly according to purchase price and unit size.

First-time owners and investors purchase at a lower price than those who benefit jointly from housing sales. These differences in purchase price are likely to result from larger financial constraints that affect their housing choice, including unit size and location.

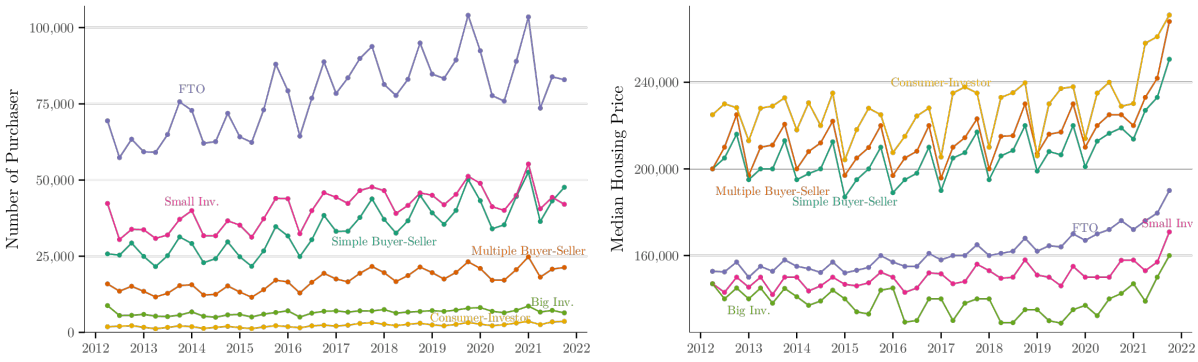


Figure 1: Quarterly number of individuals and median price for each purchaser category

*Notes:* This figure shows the number of housing units (left panel) and the median housing price (right panel) for our purchaser category. We aggregate data to the quarter. For both metric, we account for the number of purchasers within the sale. Housing price is expressed in euros.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

Hence, they tend to decrease their floor area.[1] The purchaser structure according to the sales number for repeat sales is balanced (see in Appendix A.3). These results are consistent with our proxy definition. Purchasers who do not benefit from a joint sale to finance their purchase, i.e., first-time owners, small and big investors, purchase housing units at the lowest price. Conversely, purchasers who benefit from multiple simultaneous sales to finance their purchase (i.e., consumer-investors) pay the highest price, regardless of the period of interest.

The main distinction between first-time owners and investors lies in the price per square metre. Investors purchase more expensive housing according to their size, which might be caused by two factors. First, location choices are likely to be different; investors prefer to purchase units located at a closer distance from the centre or in rural areas (Figure 2).[2] On the contrary, first-time owners are overrepresented in second-ring areas, while the purchaser-seller prefers first-ring areas. Beyond potential heterogeneity in preference, the expected tenure status after the purchase might drive this choice. The rental market is indeed spatially concentrated close to the urban centre, whereas homeownership is related to suburbs (for details, see Appendix A.4). Hence, small investors are dominant in these areas. Second, housing choices might differ according to unobservable characteristics. Investors might sort into higher quality housing units compared to first-time owners, as they might smooth consumption of unobservable characteristics in line with the affordability crisis (Halket, Nesheim and Oswald, 2020). Finally, the distribution of capital gains follows a normal distribution, both in value (Appendix A.5) and duration (Appendix A.6).

[1] For full statistics, see Appendix A.2
[2] For other categories, see Appendix A.7.

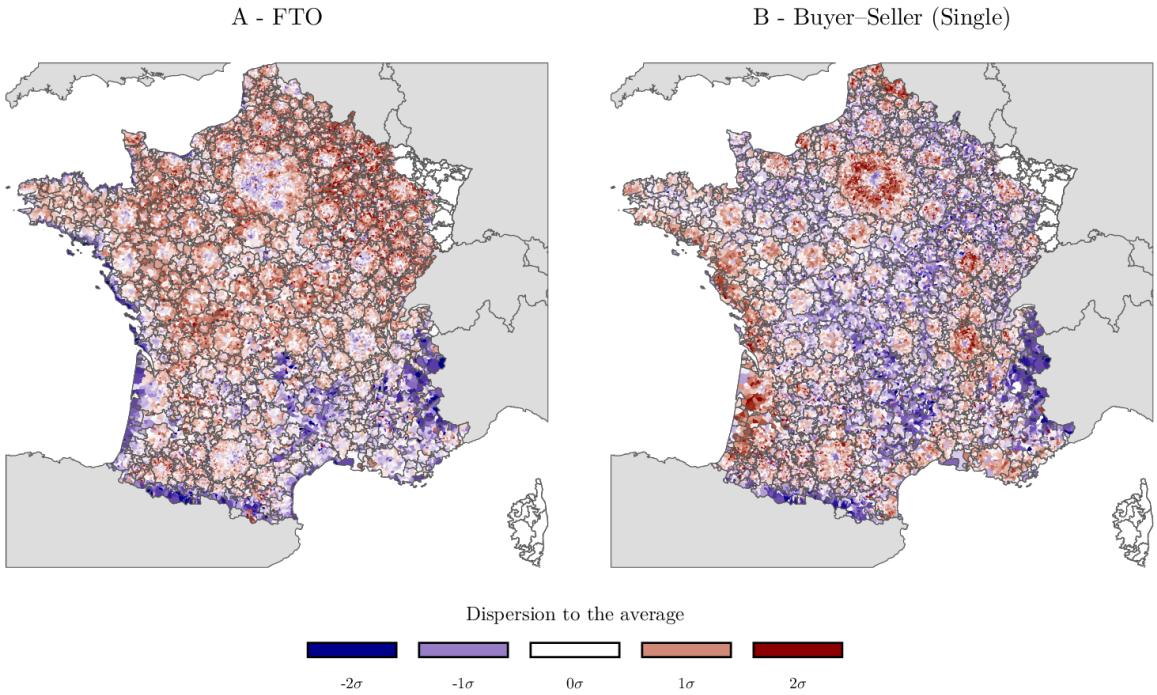


Figure 2: Preferred Location According to Purchaser Category

*Notes:* We represent the local deviation from the country average of categories within the purchasers' structure. Positive values (respectively negative) in red mean that the category is overrepresented (underrepresented) within the purchaser structure at the municipality level. The deviation is normalized using the standard deviation relative to the category. We select all transactions between 2011 and 2021. We also represent in black the perimeter of the nine most important urban areas.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

# 3 Identification Strategy

The main threat to identifying both an individual and group purchase price deviations is sorting into unobservable characteristics. Although the sources of housing heterogeneity are numerous, the hedonic method is convenient to control for observable characteristics, especially with the development of non-linear techniques. Yet, the sources of heterogeneity are often unobservable to the econometrician, with e.g., insulation, amenity view, or quality of construction. If the probability of purchasing these amenities is correlated with the purchasers' category, it would bias the group coefficient. In addition, the error term will be composed of an idiosyncratic component (faced by the individual) and the set of unobservable characteristics. Consequently, we prefer to adopt a repeat sales method to get rid of the unobservable variable component rather than the common hedonic method. From this method, we obtain a reliable group deviation while increasing the likelihood for the error term to include only the individual component.

## 3.1 Repeat Sales Approach

To control for both observed and unobserved housing characteristics, the repeat sales method represents an appealing alternative (Dorsey et al., 2010), widely used in the housing economics literature (see e.g. Biagi et al., 2021). By restricting the transaction samples to repeat sales, we explicitly introduce housing fixed effects to control for invariant characteristics such as

$$y_{ijt} = \eta_j^{RS} + \beta_1^{RS} C_i + \beta_2^{RS} \widehat{p}_{jt} + \mathbf{X}\gamma^{RS} + \mathbf{T}\delta^{RS} + \zeta_{ijt}^{RS} \tag{1}$$

with $y_{ijt}^{RS}$ outcome of interest for housing $j$ purchased by individual $j$ at time $t$, $C_i^{RS}$ purchaser category for individual $i$, $\widehat{p}_{jt}$ estimated probability to be sold at least twice, $\eta_j^{RS}$ housing fixed effects, $\mathbf{T}$ time dummies, and $\zeta_{ijt}$ the error term. The housing fixed effect controls for invariant characteristics of housing, including both observable and unobservable ones. We expect characteristics such as thermal insulation, amenity view, and construction quality to be strong determinants of housing prices. In addition, unobservable characteristics are likely to be different across the group of purchasers, providing biased estimation. The only remaining threat that we cannot control is variation in unobservable characteristics (e.g., gentrification or renovation works), which are included in the error term. To lower this threat, we introduce housing characteristics that vary over time (noted $\mathbf{X}$), including building age, for instance. However, based on robustness checks, we consider that variation in non-observable characteristics is likely to be evenly distributed within our purchaser category. Finally, we control for the post-purchase tenure status.

Then, we assume that the error term $\zeta_{ijt}^{RS}$ includes two terms. First, an individual deviation that is unexplained, representing an idiosyncratic shock. It might include preferences for specific housing assets, inattention, financial literacy, or aversion to bargaining. Second, variations in unobservable housing characteristics, such as quality, are included in the error term. Renovation works or greater deterioration of housing units are likely to affect housing value, not captured by the housing fixed effect. Empirically, we expect renovation works to be marginal and assume that the error term mainly aggregates individual shocks. Robustness checks, removing observations that are the most likely to be affected by renovation works, do not affect our conclusions.

## 3.2 Selection Model

The repeat sales approach restricts our sample to units being sold at least twice. Whereas providing results that are robust to unobservable characteristics raises concerns about selection bias, as these observations might differ significantly from the full sample. To correct for potential selection bias, we estimate the probability for each observation to be

included in the final sample, i.e., being sold at least twice, in the first step. Based on this probability, we will correct in the purchase price regression, either through the *Heckman* procedure or an Inverse Probability Weighting (IPW) scheme.

Empirically, we estimate the selection equation through a logistic regression as follows

$$log\left(\frac{p_{jt}}{1-p_{jt}}\right) = \alpha + \mathbf{X}\beta + \varepsilon_{jt} \tag{2}$$

with $p$ probability to be sold at least twice, $\mathbf{X}$ observable characteristics of the housing, and $\varepsilon$ the error term. Based on this first step, we estimate the probability for each housing unit $\widehat{p_j}$ to compute weights for the repeat sales regression.

To improve the credibility of the selection model, we exploit the recent developments of Generalized Additive Models (Wood, 2017) to determine the functional form of each explanatory variable that fits best. Using the Restricted Maximum Likelihood method (hereafter REML), the functional form for each variable is endogenously determined using penalized regressions that trade off between under-fitting and over-fitting. For instance, we expect the impact of the held duration to be non-linear, although we cannot specify properly the functional form. In addition, we exploit this framework to introduce spatial smoothing splines to account for the impact of location on the likelihood of being included in our sample. The spatial coordinates of housing units are used as the main variables for this function. In comparison to spatial fixed effects, the spatial contribution varies smoothly over space, without a strong discontinuity imposed by the boundaries. Similarly to univariate variables, the functional form is endogenously determined. The set of univariate variables includes the surface, the price per square meter, and the building age. We also add the housing type (single unit or multiple ones) and the tenure status one year before the purchase in a parametric way.

Based on the estimation of Equation (2), we estimate for each housing unit $j$ in the transaction sample the probability $\widehat{p_j}$ to be sold at least twice between 2012 and 2022 and consequently to be included in the final sample.

## 3.3 Consequences on capital gains formation

Finally, we assess the consequences of these group-specific differences on the formation of capital gains in the mid-term perspective. While relative capital gains should be independent of purchase price deviations, two mechanisms might be in place. First, positive price deviation might spill to the entire local market through a signalling effect, fuelling the housing price dynamics (Szumilo, 2021). In this setting, positive price deviations should generate, in the long term, higher price appreciation leading to greater capital

gains. Second, positive price deviations might result from individual inexperience, inattention, or aversion to bargaining. This mechanism is also likely to affect the selling process, thus reducing the selling price and capital gains, consequently. The equation is

$$\frac{y_{ijt+1}}{y_{ijt}} = \beta_1^{CG} G_i + \beta_2^{CG} \cdot \widehat{\zeta}_{ijt}^{RS} + \mathbf{X}\gamma^{CG} + \mathbf{T}\delta^{CG} + \xi_{it}^{CG} \tag{3}$$

with $y_{ijt+1}$ the selling price, $y_{ijt}$ the previous purchase price, $G_i$ group of interest according to the purchaser category, $\mathbf{X}$ housing characteristics as control, $\mathbf{T}$ time dummies for both purchase and selling time, whereas $\xi_{it}$ represents the error term. We also introduce $\widehat{\zeta}_{ijt}^{RS}$ which corresponds to the estimated individual price deviation faced by individual $i$ for purchasing housing units $j$ at time $t$ resulting from Equation (1). This equation is estimated through WLS, with standard errors being clustered at the municipality level. Finally, we introduce $\zeta^{RS}$, the individual unexplained deviation from Equation (1) is likely to affect the formation of capital gains while controlling for housing characteristics, including location.

Unlike the purchase price deviation analysis (detailed in Section 3.1), unobservable characteristics are less likely to bias our group-specific deviation. Indeed, considering that our outcome of interest is the relative capital gains estimated at the housing unit level, it captures the housing characteristics being invariant over time. As for the repeat sales approach, only the correlation between group and variation of unobservable housing characteristics is likely to bias our estimation. Yet, we assume renovation works to be marginal and evenly distributed across categories. Robustness checks are performed to support this assumption.

# 4 Results

Our results section proceeds as follows. First, we present the results of the selection model to assess the probability of being included in the final sample. Second, we highlight how the purchase price differentiates between purchaser categories. Yet, we also introduce results from a more common hedonic approach as a comparison. Third, we assess deviation for realized capital gains both at the individual and the group level.

## 4.1 Selection model

We model the probability for a housing unit to be sold at least twice between January 2012 and December 2022. We use a semi-parametric logistic regression. The introduction of smoothing splines for continuous variables accounts for potential non-linear relationships

and bi-variate smoothing splines with longitude and latitude as variables. It controls for the role of the location and local variables that may affect the probability of a housing unit being sold multiple times. In addition, structural characteristics such as surface, housing type, and date of construction are introduced as explanatory variables. We also include the price per square meter. For continuous variables, we introduce smoothing splines to endogenously determine their functional form. We report the contribution for the duration from the previous sale and location in Figure 3.

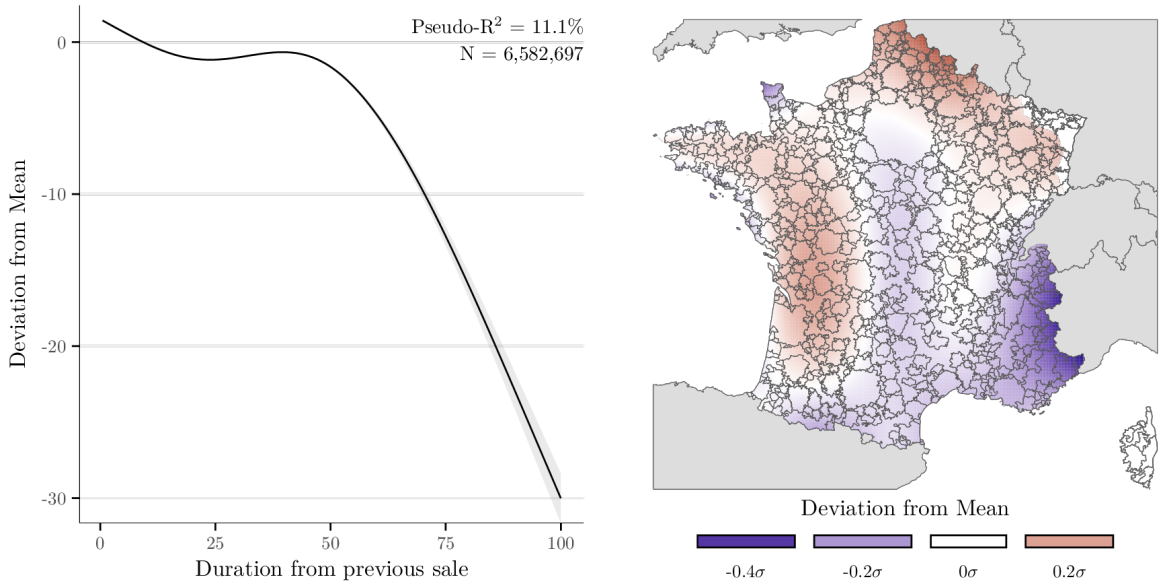


Figure 3: Logistic coefficients for held duration and housing location

*Notes:* The left panel represents the marginal contribution of held duration in the probability for a housing unit to be sold twice between 2012 and 2021. The right panel represents the spatial smoothing splines included in Equation (2). Positive values indicate that housing located in these areas is more likely to be included in the sample. These results derive from eq. (2) estimated using logistic regression. The 95% confidence interval is estimated in grey in the left panel. Additional functional forms for univariate smoothing splines are reported in Appendix B.1.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

As expected, the contribution of the held duration is non-linear (left panel, Figure 3). Observations with a small holding duration until 25 years old are more likely to be included in our sample. From 25 years, the marginal contribution drops significantly with the held duration. These housing units are likely to be off the market. In addition, location affects the probability of being sold at least twice between 2012 and 2021. Housing units located in the north of France and the western area are those with the highest probability to be sold at least twice, whereas observations in the south of France, and especially near the Italian border, are less likely to be included in our final sample. The share of explained variance is nonetheless relatively low despite the main housing characteristics being included in the selection step. It might result from some randomness in the selling decision that we cannot control for such as a job change, inheritance, or divorce. Probability distribution for both single- and repeat sales is reported on Appendix B.2.

## 4.2 Purchase price differences

We estimate price differences using the repeat sales approach with a subsample of transactions being sold at least twice to introduce housing fixed effects in the regression. We introduce the estimated probability in Section 4.1 as control variables *à la Heckman* to correct for potential selection bias. The estimation procedure is performed using Weighted Least Squares, with two-way clustered standard errors (municipality and quarter). As a baseline, we also estimate hedonic models with observable variables such as housing size, dependencies, or housing types. Results are reported in Table 1.

Table 1: WLS Estimation with Housing Fixed Effects about Price Effect According to Purchaser Category

| | *Dependent Variable:* Price (log) | | | | | |
|---|---|---|---|---|---|---|
| | Hedonic | | Housing FE | | | |
| Covariate<br>Category (ref. Simple Purchaser-Seller) | (1) | (2) | (3) | (4) | (5) | (6) |
| Multi Purchaser-Seller | 0.056* | 0.061*** | 0.010*** | 0.010*** | 0.009*** | 0.008*** |
| | (0.030) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| FTO | -0.170*** | -0.004*** | 0.046*** | 0.046*** | 0.049*** | 0.045*** |
| | (0.013) | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) |
| Small Inv. | -0.255*** | 0.014*** | 0.035*** | 0.035*** | 0.036*** | 0.033*** |
| | (0.027) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Big Inv | -0.309*** | 0.005*** | 0.014*** | 0.014*** | 0.015*** | 0.014*** |
| | (0.044) | (0.002) | (0.001) | (0.002) | (0.002) | (0.002) |
| Consumer-Investor | 0.136*** | 0.075*** | 0.010*** | 0.010*** | 0.009*** | 0.009*** |
| | (0.025) | (0.002) | (0.001) | (0.002) | (0.002) | (0.002) |
| Tenure status: Own-occupied (ref.) | | | | | | |
| Rent | – | -0.136*** | -0.012*** | -0.012*** | -0.014*** | -0.032*** |
| | | (0.001) | (0.001) | (0.001) | (0.001) | (0.002) |
| Vacant | – | -0.090*** | -0.007*** | -0.007*** | -0.009*** | -0.032*** |
| | | (0.001) | (0.002) | (0.002) | (0.002) | (0.003) |
| Other | – | -0.067*** | 0.004** | 0.004 | 0.002 | -0.013*** |
| | | (0.002) | (0.002) | (0.003) | (0.004) | (0.003) |
| N | 2,251,773 | 2,251,773 | 2,251,773 | 2,251,773 | 2,251,773 | 2,251,773 |
| $R^2$ | 0.028 | – | 0.958 | 0.958 | 0.959 | 0.962 |
| Adj. $R^2$ | 0.028 | 0.719 | 0.940 | 0.940 | 0.942 | 0.946 |
| Within $R^2$ | – | – | 0.296 | 0.296 | 0.274 | 0.359 |
| Housing Control | No | Spl. Space | FE | FE | FE | FE |
| Quarter Control | No | Yes | Yes | Yes | Yes | Yes |
| Selection Bias Control | No | No | No | No | IPW | Heckman |
| SE Clustered | No | No | No | Two-Way | Two-Way | Two-Way |

*Notes:* We report results for WLS estimation of Equation (1) and alternative results with hedonic method. The first three columns do not include housing fixed effects, whereas the last three do. In addition, the last three specifications vary according to whether the model is correct for probability to be included in the sample, and the cluster used for standard errors. We report the only difference between the purchaser category in comparison to the category of reference, being simple purchaser sellers. We report in parentheses standard errors. The estimation is performed on housing transactions about housing being sold at least twice between 2012 and 2021 at the country level. Additional control variables, including time and building age, are reported on respectively Appendix B.3 and Appendix B.4.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.
*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$

The introduction of housing fixed effects has a sizeable impact on purchase group-specific price deviations. Indeed, whereas the most flexible specification of the hedonic model with spatial smoothing splines explains 71.5% of explained variance (column 2, Table 1), first-time owners underpay by 2.6% in comparison to single purchaser-seller (our reference

category). Introducing housing fixed effects improves the explained variance by more than 20pp and significantly affects the group-specific price difference. The first-time owners represent the category that faces the highest premium controlling for observable and non-observable housing characteristics (+4.5%). In addition, small and big investors also overpay by respectively 3.3% and 1.4%. Purchasers with a lower level of experience and who are more financially constrained experience the highest price deviations. In addition, the observed differences in coefficients between hedonic and housing fixed effects specification (columns 2 and 5, Table 1) indicate that unobservable characteristics are likely to be unevenly distributed across purchaser characteristics based on their quality.

Besides the change for group-specific coefficients, the introduction of housing fixed effects affects tenure premiums. The introduction of housing fixed effects lowers these differences (from −12.4% to −1.2%, columns 2 and 3, Table 1), although the hierarchy remains unchanged, as buy-to-rent housing units are purchased at a lower price from those being purchased for owner-occupied purposes. We expect unobserved quality according to tenure status to drive these differences, as for group-specific price deviations.

The selection correction based on probability has little impact on our results (columns 4, 5, and 6, Table 1). The IPW approach (column 5) exacerbates slightly the group-specific differences compared to the Heckman correction, while it introduces sizeable differences for tenure deviations. In the same way, the introduction of two-way clustered standard errors has little impact on the significance of our results.

We add control variables as robustness checks, especially to control for potential information asymmetries based on geography (Cvijanović and Spaenjers, 2021; Li and Chau, 2023) that might correlate with a specific category. We compute the distance between the purchaser's main residence (available in the fiscal files) and the housing location. The introduction of these control variables does not affect our results, whereas we recover the expected sign for distance (purchase price deviation increases with distance, see in Appendix B.5).

Finally, we assess whether our results are robust to variation in unobservable housing characteristics, e.g., renovation works. First, we remove from our sample housing units that experience an important variation in housing per square meter during the two sales. We assume that renovation works significantly affect the price per square meter between two sales, correcting for local housing market dynamics. We remove observations belonging respectively to the top 1%, 5%, 10%, and 20% of the annual price per square meter variation. Second, we remove observations for which the duration between the two sales is superior to two, three, or four years. The underlying assumption is that the lower the duration between two sales, the more likely the unobservable characteristics are to be similar. In both cases, the hierarchy between our purchaser categories remains unaffected. Although it affects the magnitude of price difference (see in Appendix B.6 and

Appendix B.7), it does not affect our main results.

## 4.3 Consequences on realized capital gains

We close the results section by focusing on realized capital gains. We estimate Equation (3) with WLS with weights accounting for the number of purchasers within the sale. Our data is then restricted to the first wave of purchase to compute realized capital gains at the housing unit level. Standard errors are clustered at the municipality level. We add, in the full specification model, the municipality of the housing, the timing of the purchase, and the timing of the sale as control variables, using quarter dummies. Results are reported in Table 2, whereas we provide similar results with the level of capital gains in Appendix B.8.

Table 2: WLS Results for Selling Price Heterogeneity According to Each Group

| | Relative Difference | | | | | | |
|---|---|---|---|---|---|---|---|
| Covariate | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Multi Purchaser-Seller | 0.012*** | 0.002 | -0.007*** | – | -0.029*** | – | -0.028*** |
| | (0.002) | (0.002) | (0.002) | | (0.001) | | (0.001) |
| FTO | 0.011*** | 0.000 | 0.000 | – | -0.034*** | – | -0.038*** |
| | (0.003) | (0.002) | (0.001) | | (0.001) | | (0.001) |
| Small Inv. | 0.019*** | -0.002 | -0.009*** | – | -0.054*** | – | -0.056*** |
| | (0.002) | (0.002) | (0.001) | | (0.001) | | (0.001) |
| Big Inv | 0.052*** | 0.028*** | 0.014*** | – | -0.035*** | – | -0.039*** |
| | (0.003) | (0.004) | (0.003) | | (0.002) | | (0.001) |
| Consumer-Investor | 0.007** | -0.000 | -0.010*** | – | -0.032*** | – | -0.029*** |
| | (0.004) | (0.003) | (0.004) | | (0.002) | | (0.002) |
| Individual Deviation | – | – | – | -0.024*** | -0.024*** | -0.023*** | -0.024*** |
| | | | | (0.000) | (0.000) | (0.000) | (0.000) |
| Held Duration (years) | – | 0.000 | 0.000 | 0.008*** | 0.009*** | 0.008*** | 0.009*** |
| | | (0.001) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure status: Own-occupied (ref.) | | | | | | | |
| Rent | – | -0.000 | -0.014*** | -0.037*** | -0.034*** | -0.049*** | -0.047*** |
| | | (0.003) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Vacant | – | 0.010*** | 0.002 | -0.021*** | -0.020*** | -0.025*** | -0.024*** |
| | | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Other | – | -0.000 | -0.015*** | -0.031*** | -0.027*** | -0.034*** | -0.031*** |
| | | (0.004) | (0.003) | (0.002) | (0.002) | (0.002) | (0.002) |
| N | 1,354,529 | 1,354,529 | 1,354,529 | 1,354,529 | 1,354,529 | 1,354,529 | 1,354,529 |
| $R^2$ | 0.001 | 0.068 | 0.199 | 0.726 | 0.730 | 0.760 | 0.764 |
| Adj. $R^2$ | 0.001 | 0.068 | 0.184 | 0.726 | 0.730 | 0.756 | 0.760 |
| Within $R^2$ | – | 0.033 | 0.045 | 0.716 | 0.720 | 0.714 | 0.719 |
| Housing Control | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Time Control | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Muni FE | No | No | Muni | No | No | Muni | Muni |
| Heckman Correction | No | Yes | Yes | Yes | Yes | Yes | Yes |

*Notes:* We report results for WLS estimation of Equation (3). The first three columns only include housing characteristics and purchaser categories, whereas the last two include the individual deviation (namely the residuals of Equation (1)). We report only the difference between the purchaser category in comparison to the category of reference, being simple purchaser sellers, and the individual deviation. We report in parentheses standard errors using a 1,000-iteration procedure. The estimation is performed on paired housing transactions about housing being sold at least twice between 2012 and 2021 at the country level to compute capital gains.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.
*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$

Considering relative capital gains, our results indicate that individual purchase price deviation is correlated with the relative capital gains difference. We estimate that a 1pp

increase in purchase price deviation decreases the relative capital gains difference by 2.4pp (column 6, Table 2). Transformed in constant euros, a €1 increase in purchase price deviation decreases capital gains by €0.41. The results cannot be interpreted causally, as individuals who experience a positive deviation for purchase price might sell at a lower price, conditionally on housing price. Yet, it supports the fact that individual behaviour matters when considering the formation of capital gains, besides housing location. Moreover, we find that relative capital gains differ between groups. Our category of reference (single buyer-seller, i.e., mostly residential mobility) benefits from the highest capital gains (columns 6 and 7, Table 2). On the contrary, the small investors have the lowest relative capital gains, conditionally on housing characteristics. Bargain intensity in the selling process is likely to differ between those groups as well, as the classification is not aligned with those retrieved for the purchase price deviations (Figure 4).

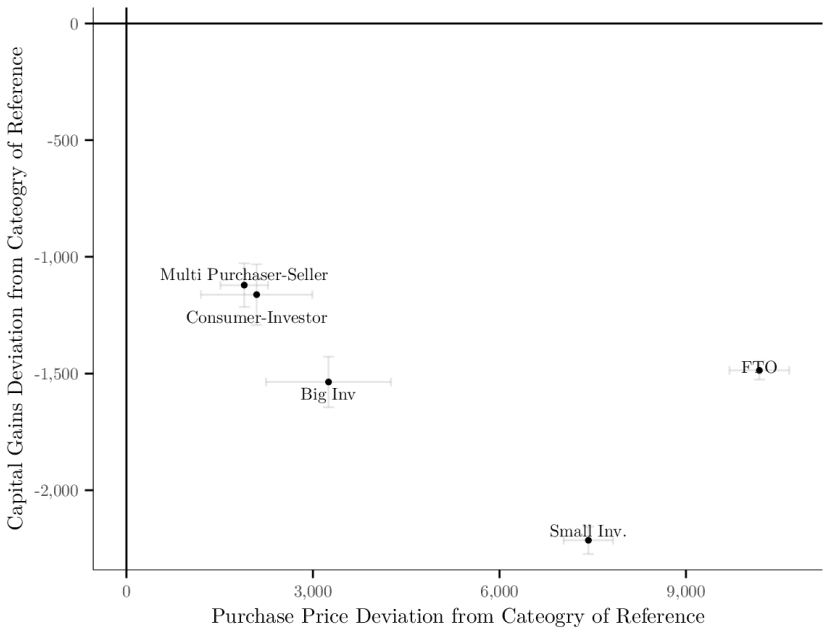


Figure 4: Bivariate Results for Purchase Price and Capital Gains

*Notes:* We report the difference in euros for each purchaser category for the purchase price difference (Equation (1)) and the capital gains (Equation (3)). The category of reference is homeowners achieving residential mobility. We report standard errors in light grey for both equations.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

From the purchase price process, we distinguish two distinct groups. On the one hand, the small investors and the first-time buyers are considered to be financially constrained and less experienced purchasers. On the other hand, big investors, multiple purchaser-sellers, and consumer-investors are considered more experienced and less financially constrained. The first group pays a higher premium compared to the latter one (the x-axis, Figure 4). When considering differences in capital gains, they all lose relatively to our category of reference (the y-axis, Figure 4). Yet, first-time owners and small investors differ, which might result from different bargaining intensity in the selling process. We find

similar results with big investors that face a higher discount in capital gains compared to consumer-investors and multiple purchaser-sellers.

Finally, the held duration increases the relative capital gains, which is consistent with the positive price appreciation over time. Turning to the tenure status premium, despite a purchase price discount for buy-to-rent investment, these types of investments generate lower relative capital gains compared to owner-occupied housing units. The fact that seller benefit from the use value compared to buy-to-rent housing units might shape price expectations when selling their housing, thus conditioning capital gains. Removing observations which are likely to correspond to renovation does not affect our conclusions (see Appendix B.9, Appendix B.10 and Appendix B.11).

# 5 Discussion

We close the article by discussing our results in regard with the underlying mechanisms that drive our main results and the sorting of non-observable housing characteristics.

## 5.1 Underlying Mechanisms

We demonstrate that housing characteristics, including housing location, are not the only drivers to understand purchase price and capital gains formation. We highlight that individual matters, whether it is their experience, their financial constraints, or inattention. Whereas we might expect capital gains to be independent of purchase price, we estimate a negative correlation between purchase price deviation and relative capital gains.

To drive these results, we expect three mechanisms to be in place: spillover effect through contamination of the price reference of local homeowners, individual performance on the housing market, and the catching-up effect. The first mechanism is expected to induce a positive relationship between housing prices and capital gains. Indeed, the positive price deviation might fuel the local housing price dynamics, transitioning into higher capital gains in the long term. The two other mechanisms tend to favour a negative relationship between purchase price and capital gains. Individuals facing a positive purchase price deviation might be less effective in the housing market, especially during the bargaining process. Consequently, they are more likely to be less effective when selling their housing unit, lowering capital gains. Finally, if the purchase price deviation is disproportionately positive, regardless of the two other mechanisms, the capital gains will be low if the selling process is close to the common value in the local housing market. Based on our results, we expect the latter two mechanisms to be dominant over the first one.

Besides the correlation between purchase price deviations on realized capital gains, the purchase price deviations are likely to exacerbate differences when accounting for financial burdens. As we assume that most housing purchases are financed through a mortgage, these price differences will affect reimbursement. Taking a 2% annual interest rate for 20 20-year loan maturity, the first-time owners will correspond to a total increase of €15,000 compared to homeowners achieving their residential mobility, i.e., nearly €63 per month. Hence, purchase price differences are exacerbated when accounting for interest rates. Finally, we consider that the borrowing rate is homogeneous between purchaser categories, which is unlikely to hold. Hence, the monthly differences accounting for interest rates are a lower bound.

## 5.2 Sorting Into Non-Observable Variables

We provide insightful results about sorting into non-observable characteristics between purchaser categories. We infer sorting into non-observable based on differences between coefficients derived from the model with housing fixed effects (column 6, Table 1) and those from the hedonic specification (column 3, Table 1). If the difference is positive, we consider that this category consumes lower quality housing on unobservable characteristics compared with single-purchaser sellers. Considering the price difference, we assume that first-time owners purchase housing that provides lower quality for non-observable characteristics. This result is consistent with the fact that first-time owners are more likely to smooth their consumption (including housing) to access homeownership (Waxman et al., 2020). More interestingly, a major difference occurs between small and big investors. Small investors tend to select housing with better non-observable characteristics than single purchaser-sellers, whereas big investors purchase lower-quality ones. These results are consistent with our expectations that small investors are more likely to purchase secondary homes (and thus owners are more willing to pay attention to housing characteristics), whereas big investors are more likely to buy-to-rent. The difference in the expected tenure status might explain differences in sorting in non-observable characteristics.

Lowering living standards for first-time owners may affect social and economic outcomes over the long term. Assuming that thermal insulation is one of the main characteristics that purchasers sort into, consequences on health are significant even in the short term (Angel and Bittschi, 2019; Lima, Ferreira and Leal, 2020). Whereas investors would not be concerned directly as they buy-to-rent, first-time owners would be significantly and permanently affected by lower living standards. This fact is also supported by the lower likelihood of renovation, regarding the long-term impact on monthly payments. Hence, the most constrained purchasers sort into the lowest quality building, limiting the potential for thermal renovation that would benefit public health (Levy, Nishioka and

Spengler, 2003). In the context of transitioning into a low-carbon emission society, the sorting between financially constrained households and low-quality housing is likely to be an important friction.

# 6 Conclusion

Homeowners are mostly considered as winners or losers in the housing market based on their assets' locations. Yet, we highlight that individual behaviour shapes purchase price and capital gains. We exploit a new category definition that proxies for experience and financial constraints, using both the size of the portfolio and the joint sale of a housing unit when purchasing a new one. Considering that housing choices are likely to differ between purchasers both in observable and unobservable characteristics, we introduce housing fixed effects to isolate group and individual price deviation. This approach significantly improves the share of explained variance and reshuffles the hierarchy for price differences between groups. However, it restricts our sample to housing units being sold at least twice between 2012 and 2021. We correct for potential selection bias through the estimation of the probability for housing units to be sold at least twice based on observable characteristics. The probability is then introduced in the price regression using the Heckman procedure. We then isolate idiosyncratic shocks at the individual level that we use as explanatory variables to explain realized capital gains.

We estimate a negative relationship between individual price deviation and realized capital gains. A €1 increase in purchase price deviations decreases capital gains by €0.41 realized capital gains. We expect inattention, individual efficiency in both the purchase and selling processes to drive these results. In addition, the less experienced purchasers (i.e., first-time owners) experience the greatest premiums for purchase price (+4.5% compared to our category of reference), followed by small investors. Individual characteristics are likely to shape housing prices and capital gains, in addition to location.

Further research might focus on two specific questions. First, while previous research uncovered a spillover mechanism for overpayment that fuels the housing market dynamics, the conditions of the emergence of this spillover mechanism are of interest. Whereas the main research has focused on the overpayments from the wealthiest, it raises concerns about whether it works similarly for other categories of purchasers. Second, although we uncover that individuals shape their wealth accumulation, we do not disentangle whether it results from heterogeneity in financial literacy, inattention or alternative channels. Hence, future research should focus on each mechanism to better understand how individuals shape their wealth accumulation.

# References


Aiello, Darren, Jason D. Kotter and Gregor Schubert (2022), 'Housing Wealth and Overpayment: When Money Moves In'. *SSRN Electronic Journal*.

Angel, Stefan and Benjamin Bittschi (2019), 'Housing and Health'. *Review of Income and Wealth* 65.3, pp. 495–513.

Bach, Laurent, Laurent E. Calvet and Paolo Sodini (2020), 'Rich Pickings? Risk, Return, and Skill in Household Wealth'. *American Economic Review* 110.9, pp. 2703–2747.

Bäckman, Claes and Chandler Lutz (2020), 'The Impact of Interest-Only Loans on Affordability'. *Regional Science and Urban Economics* 80.103376.

Behrman, Jere R., Olivia S. Mitchell, Cindy K. Soo and David Bravo (2012), 'How Financial Literacy Affects Household Wealth Accumulation'. *The American economic review* 102.3, pp. 300–304.

Ben-Shahar, Danny, Stuart Gabriel and Roni Golan (2019), 'Housing Affordability and Inequality:A Consumption-Adjusted Approach'. *Journal of Housing Economics* 45, p. 101567.

Biagi, Bianca, Steven B. Caudill, Laura Ciucci, Claudio Detotto and Franklin Mixon (2021), 'Relative Bargaining Power of Residential Home Traders and Real Estate Investors'. *Applied Economics* 53.34, pp. 3962–3971.

Braakmann, Nils and Stephen McDonald (2020), 'Housing Subsidies and Property Prices: Evidence from England'. *Regional Science and Urban Economics* 80, p. 103374.

Cvijanović, Dragana and Christophe Spaenjers (2021), '"We'll Always Have Paris": Out-of-Country Buyers in the Housing Market'. *Management Science* 67.7, pp. 4120–4138.

Daysal, N. Meltem, Michael F. Lovenheim and David N. Wasser (2023), *The Intergenerational Transmission of Housing Wealth*. Working Paper.

Dorsey, Robert E., Haixin Hu, Walter J. Mayer and Hui-chen Wang (2010), 'Hedonic versus Repeat-Sales Housing Price Indexes for Measuring the Recent Boom-Bust Cycle'. *Journal of Housing Economics* 19.2, pp. 75–93.

Dupont, Emiko, Isa Marques and Thomas Kneib (2023), 'Demystifying Spatial Confounding'.

Eggum, Terje and Erling Røed Larsen (2024), 'Is the Housing Market an Inequality Generator?' *Review of Income and Wealth* 70.3, pp. 697–722.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino and Luigi Pistaferri (2020), 'Heterogeneity and Persistence in Returns to Wealth'. *Econometrica* 88.1, pp. 115–170.

Favilukis, Jack and Stijn Van Nieuwerburgh (2021), 'Out-of-Town Home Buyers and City Welfare'. *The Journal of Finance* 76.5, pp. 2577–2638.

Favilukis, Jack Y., Pierre Mabille and Stijn Van Nieuwerburgh (2021), *Affordable Housing and City Welfare*. SSRN Scholarly Paper. Rochester, NY.

Frémeaux, Nicolas and Marion Leturcq (2020), 'Inequalities and the Individualization of Wealth'. *Journal of Public Economics* 184, pp. 104–145.

Garriga, Carlos, Pedro Gete and Athena Tsouderou (2023), 'The Economic Effects of Real Estate Investors'. *Real Estate Economics* 51.3, pp. 655–685.

Gilbert, Brian, Abhirup Datta, Joan A. Casey and Elizabeth L. Ogburn (2024), *A Causal Inference Framework for Spatial Confounding*.

Halket, Jonathan, Lars Nesheim and Florian Oswald (2020), 'The Housing Stock, Housing Prices, and User Costs: The Roles of Location, Structure, and Unobserved Quality'. *International Economic Review* 61.4, pp. 1777–1814.

Harding, John P. and Stuart S. Rosenthal (2017), 'Homeownership, Housing Capital Gains and Self-Employment'. *Journal of Urban Economics* 99, pp. 120–135.

Heckman, James J. (1979), 'Sample Selection Bias as a Specification Error'. *Econometrica* 47.1, p. 153.

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick and Alan M Taylor (2019), 'The Rate of Return on Everything, 1870–2015'. *The Quarterly Journal of Economics* 134.3, pp. 1225–1298.

Khoshnoud, Mahsa, G. Stacy Sirmans and Emily N. Zietz (2023), 'The Evolution of Hedonic Pricing Models'. *Journal of Real Estate Literature* 31.1, pp. 1–47.

Kiel, Katherine A. and Jeffrey E. Zabel (2008), 'Location, Location, Location: The 3L Approach to House Price Determination'. *Journal of Housing Economics* 17.2, pp. 175–190.

Kunovac, Davor and Ivan Zilic (2022), 'The Effect of Housing Loan Subsidies on Affordability: Evidence from Croatia'. *Journal of Housing Economics* 55.101808.

Lei, Rémi, Laure Casanova Enault, Antoine Peris and Martin Bocquet (2024), 'Évolution de La Structure de La Propriété Immobilière En France: Apports Des Fichiers Fonciers Enrichis Pour Analyser Les Patrimoines Des Individus'. *Revue d'Économie Régionale & Urbaine* Juin.3, pp. 375–400.

Levy, Jonathan I., Yurika Nishioka and John D. Spengler (2003), 'The Public Health Benefits of Insulation Retrofits in Existing Housing in the United States'. *Environmental Health* 2.1, p. 4.

Li, L. and K. W. Chau (2023), 'Information Asymmetry with Heterogeneous Buyers and Sellers in the Housing Market'. *The Journal of Real Estate Finance and Economics*.

Lima, Fátima, Paula Ferreira and Vítor Leal (2020), 'A Review of the Relation between Household Indoor Temperature and Health Outcomes'. *Energies* 13.11, p. 2881.

Ling, David C., Andy Naranjo and Benjamin Scheick (2021), 'There Is No Place like Home: Information Asymmetries, Local Asset Concentration, and Portfolio Returns'. *Real Estate Economics* 49.1, pp. 36–74.

Martínez-Toledano, Clara (2022), 'House Price Cycles, Wealth Inequality and Portfolio Reshuffling'. *Working Paper*, p. 80.

Roine, Jesper and Daniel Waldenström (2012), 'On the Role of Capital Gains in Swedish Income Inequality'. *Review of Income and Wealth* 58.3, pp. 569–587.

Szumilo, Nikodem (2021), 'Prices of Peers: Identifying Endogenous Price Effects in the Housing Market'. *The Economic Journal* 131.639, pp. 3041–3070.

Tur-Sinai, Aviad, Larisa Fleishman and Dmitri Romanov (2020), 'The Accuracy of Self-Reported Dwelling Valuation'. *Journal of Housing Economics* 48.101660.

Turnbull, Geoffrey K. and Arno J. Van Der Vlist (2022), 'Bargaining Power and Segmented Markets: Evidence from Rental and Owner-occupied Housing'. *Real Estate Economics* 50.5, pp. 1307–1333.

Turner, Tracy M. and Heather Luea (2009), 'Homeownership, Wealth Accumulation and Income Status'. *Journal of Housing Economics* 18.2, pp. 104–114.

Wainer, Allison and Jeffrey Zabel (2020), 'Homeownership and Wealth Accumulation for Low-Income Households'. *Journal of Housing Economics* 47, p. 101624.

Waxman, Andrew, Yuanning Liang, Shanjun Li, Panle Jia Barwick and Meng Zhao (2020), 'Tightening Belts to Buy a Home: Consumption Responses to Rising Housing Prices in Urban China'. *Journal of Urban Economics* 115.103190.

Wind, Barend and Lina Hedman (2018), 'The Uneven Distribution of Capital Gains in Times of Socio-Spatial Inequality: Evidence from Swedish Housing Pathways between 1995 and 2010'. *Urban Studies* 55.12, pp. 2721–2742.

Wong, Siu Kei, Kuang Kuang Deng and Kwong Wing Chau (2022), 'Do Short-Term Real Estate Investors Outperform the Market?' *Journal of Real Estate Research* 44.2, pp. 287–309.

Wood, Simon N. (2017), *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC.

# A Descriptive Statistics

## A.1 Number of Housing Units According to the Number of Sales

Table A.1: Number of Housing Units According to the Number of Sales

| Number of Sales | Number of Housing | Share of Transactions |
|---|---|---|
| 1 | 5,490,068 | 87.73% |
| 2 | 723,268 | 11.56% |
| 3 | 44,732 | 0.71% |

*Notes:* We report the number (and the share) of housing units being sold according to the number sales between 2012 and 2021. The observation unit is the housing.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## A.2 Descriptive Statistics about Repeat Sales Sample

Table A.2: Descriptive Statistics about Purchase by Category

| Category | N | Mean | Std. Dev. | Median | Q1 | Q3 |
|---|---|---|---|---|---|---|
| | | | A - Purchase Price | | | |
| Simple Purchaser-Seller | 289,096 | 224,150 | 150,610 | 190,000 | 130,000 | 274,310 |
| Multi Purchaser-Seller | 119,768 | 250,480 | 198,160 | 200,000 | 125,640 | 311,600 |
| FTO | 625,550 | 183,520 | 117,440 | 158,700 | 113,000 | 221,000 |
| Small Inv. | 248,135 | 185,420 | 151,200 | 148,000 | 90,000 | 231,000 |
| Big Inv | 33,409 | 186,940 | 178,570 | 135,500 | 79,000 | 231,800 |
| Consumer-Investor | 16,713 | 264,370 | 202,890 | 212,680 | 140,000 | 325,000 |
| | | | B - Housing Size | | | |
| Simple Purchaser-Seller | 289,096 | 90.1 | 37.8 | 85.0 | 66.0 | 108.0 |
| Multi Purchaser-Seller | 119,768 | 85.3 | 45.0 | 80.0 | 54.0 | 107.0 |
| FTO | 625,550 | 76.3 | 32.9 | 72.0 | 55.0 | 92.0 |
| Small Inv. | 248,135 | 69.0 | 39.8 | 63.0 | 39.0 | 89.0 |
| Big Inv | 33,409 | 64.5 | 42.4 | 55.0 | 33.0 | 83.0 |
| Consumer-Investor | 16,713 | 92.1 | 44.4 | 85.0 | 63.0 | 114.0 |
| | | | C - Price per square metres | | | |
| Simple Purchaser-Seller | 289,096 | 2,669 | 1,679 | 2,250 | 1,578 | 3,235 |
| Multi Purchaser-Seller | 119,768 | 3,160 | 2,011 | 2,630 | 1,783 | 3,900 |
| FTO | 625,550 | 2,686 | 1,775 | 2,200 | 1,533 | 3,223 |
| Small Inv. | 248,135 | 2,969 | 1,955 | 2,450 | 1,633 | 3,667 |
| Big Inv | 33,409 | 3,162 | 2,162 | 2,583 | 1,625 | 4,000 |
| Consumer-Investor | 16,713 | 3,078 | 1,992 | 2,530 | 1,741 | 3,770 |

*Notes:* We report descriptive statistics about purchase price (top panel), housing size (middle panel) and price per square meter (bottom panel) for each purchaser category. The transaction sample concerns all housing transactions being sold at least twice between 2012 and 2018 at the country level.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## A.3 Purchaser Structure for Each Mutation Number for Repeat Sales

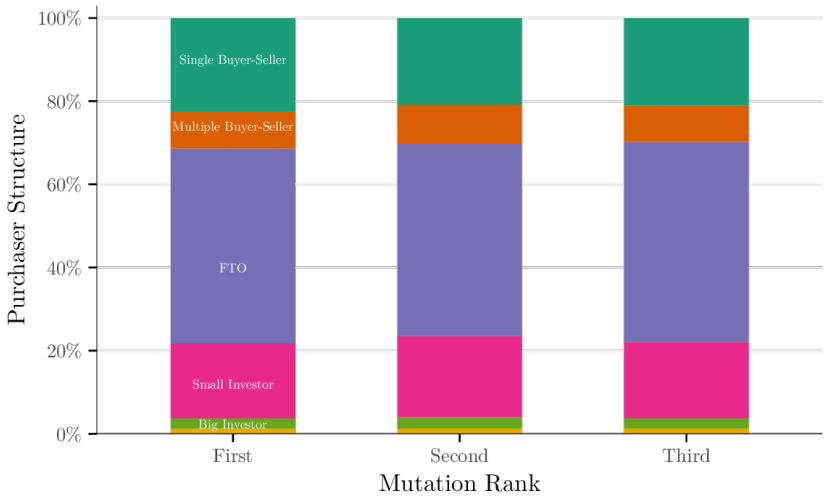


Figure A.3: Purchaser Structure for Each Mutation Number for Repeat Sales

*Notes:* We report the share of purchaser category within the repeat sales sample. We distinguish the purchasers structure according to whether it is the first, the second or the third sale of the housing units over our studied period.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## A.4 Spatial Distribution of Tenure Status

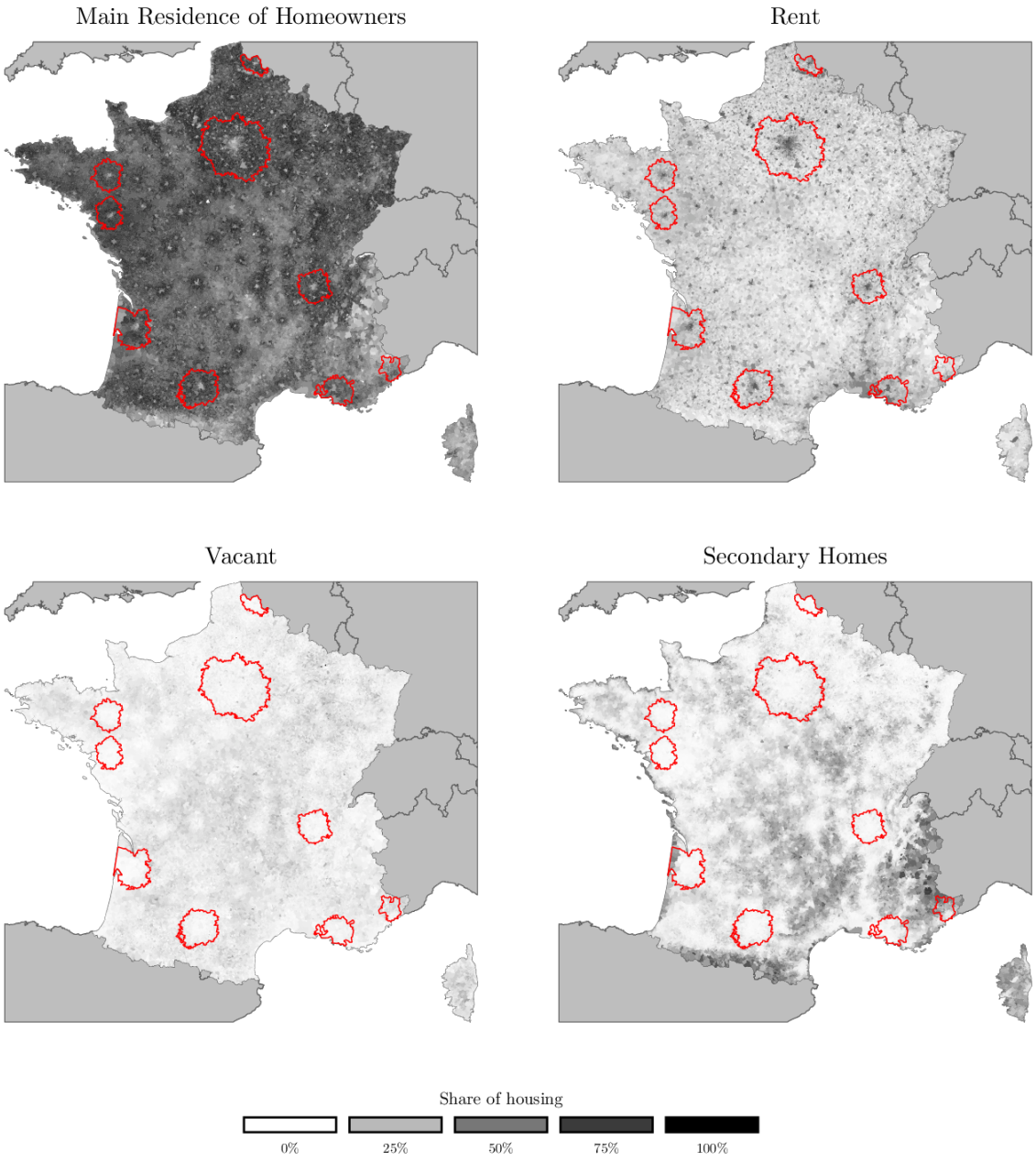


Figure A.4: Share of Housing per Tenure Status (2018)

*Notes:* We report the share of tenure status at the municipality level in 2018. We distinguish main residence of homeowners, rental housing, vacant housing and secondary homes. The share is expressed in percent. We add in red the nine most important urban areas in France.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## A.5 Distribution of Relative Capital Gains

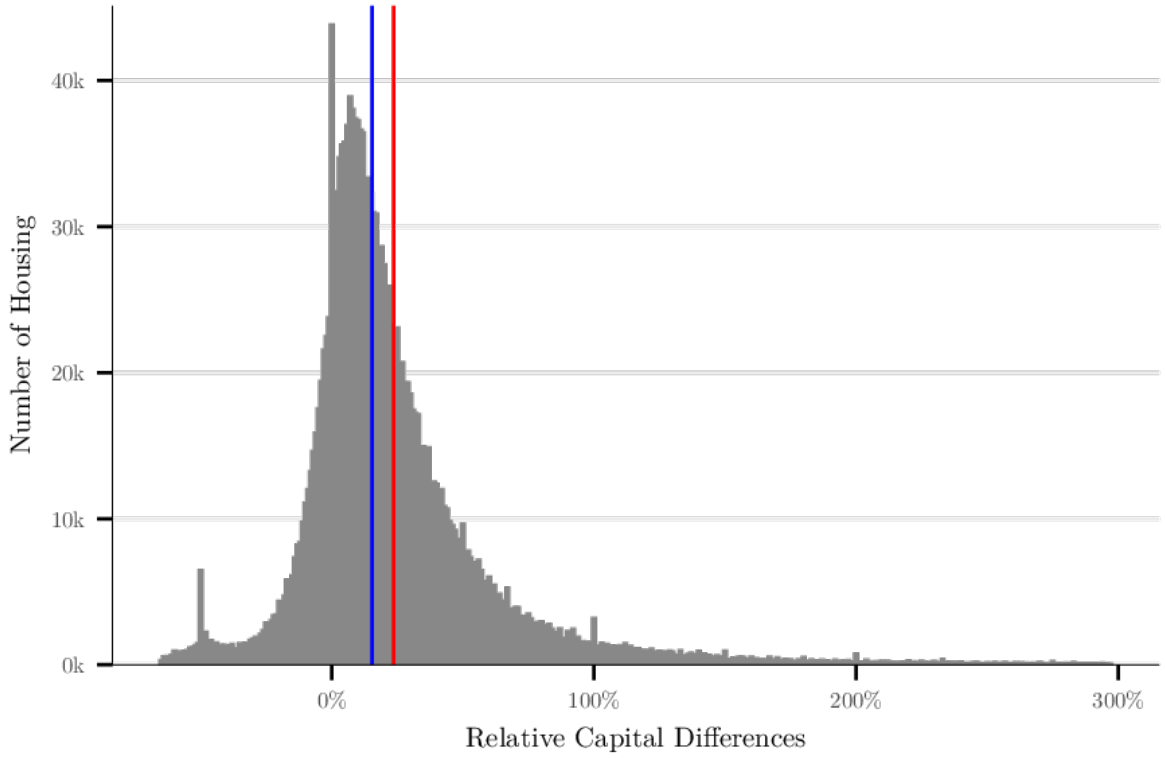


Figure A.5: Distribution of Realized Capital Gains

*Notes:* We report the distribution of relative capital gains in percent. For each housing unit, we appraise the capital gains by subtracting the purchase price to the selling price. The observation unit is the housing. The blue line represents the median capital gains, whereas the red line is the average one.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## A.6 Distribution of Duration Between Two Sales

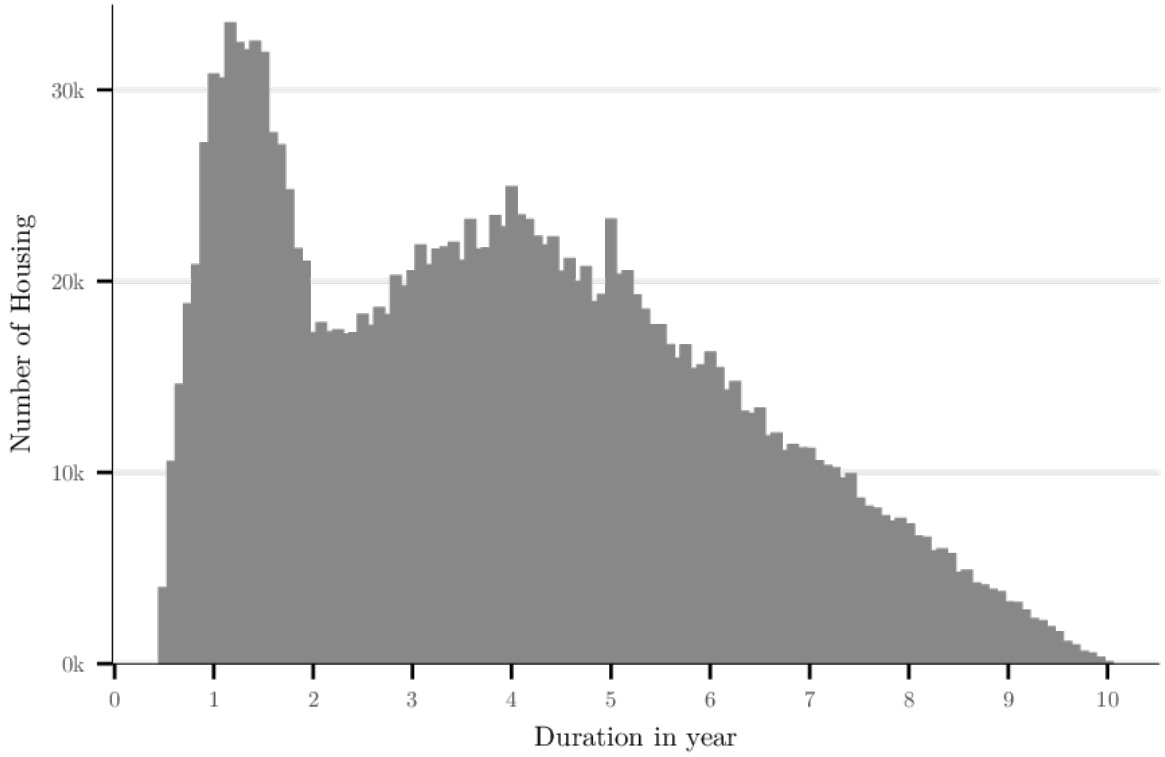


Figure A.6: Distribution of Duration between Two Sales

*Notes:* We report the distribution of housing units according to their duration between the two sales. For each housing unit, we appraise the duration between the two sales. The observation unit is the housing.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## A.7 Spatial Dispersion for Alternative Categories

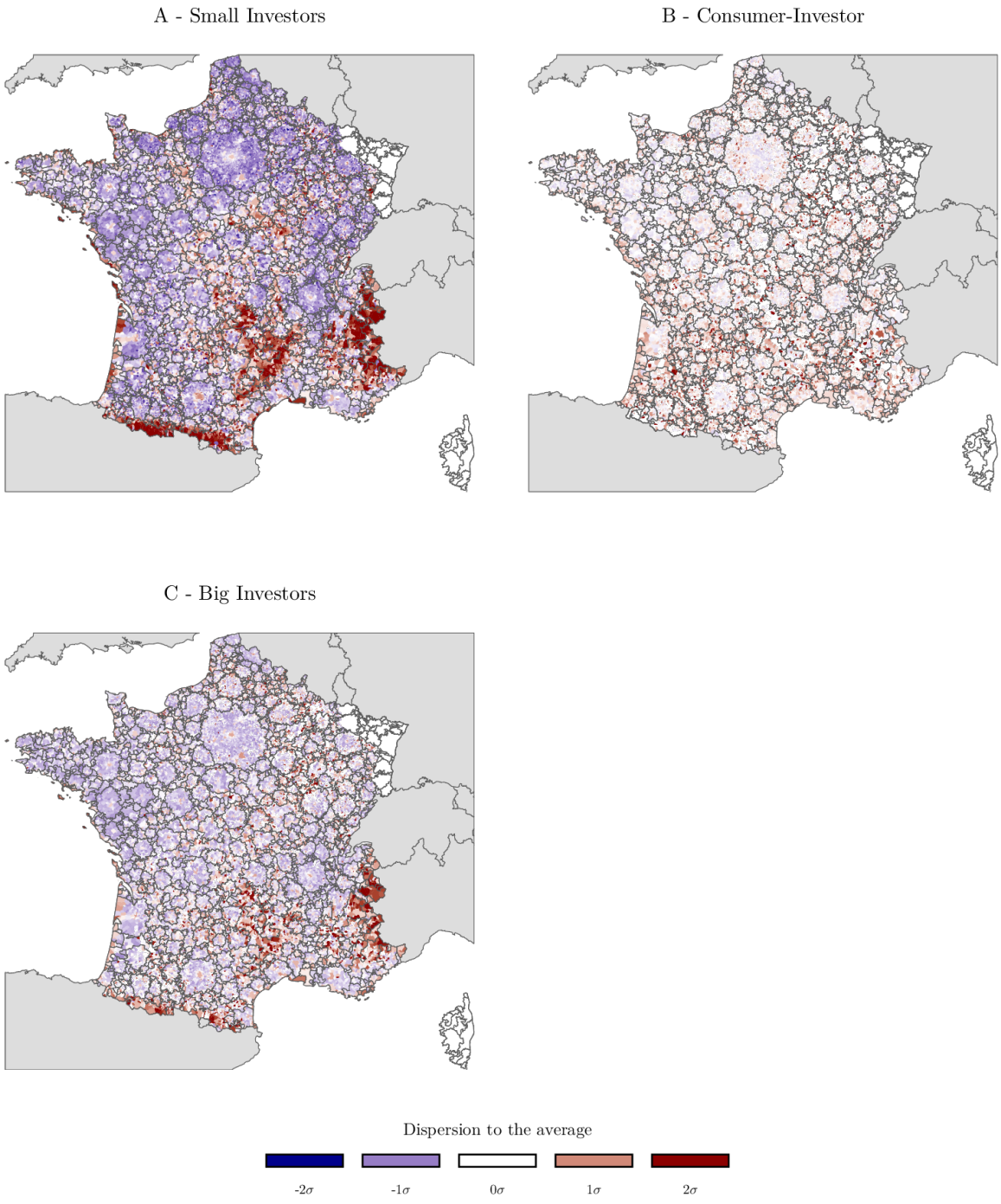


Figure A.7: Spatial Dispersion for Alternative Categories

*Notes:* We represent the local deviation to the country average of categories within the purchasers' structure. Positive values (respectively negative) in red means that the category is overrepresented (underrepresented) within the purchaser structure at the municipality level. The deviation is normalized using standard deviation relative to the category. We select all transactions between 2011 and 2021. We also represent in black the perimeter of the nine most important urban areas
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

# B Additional Results

## B.1 Selection Model: Marginal contribution of univariate variables

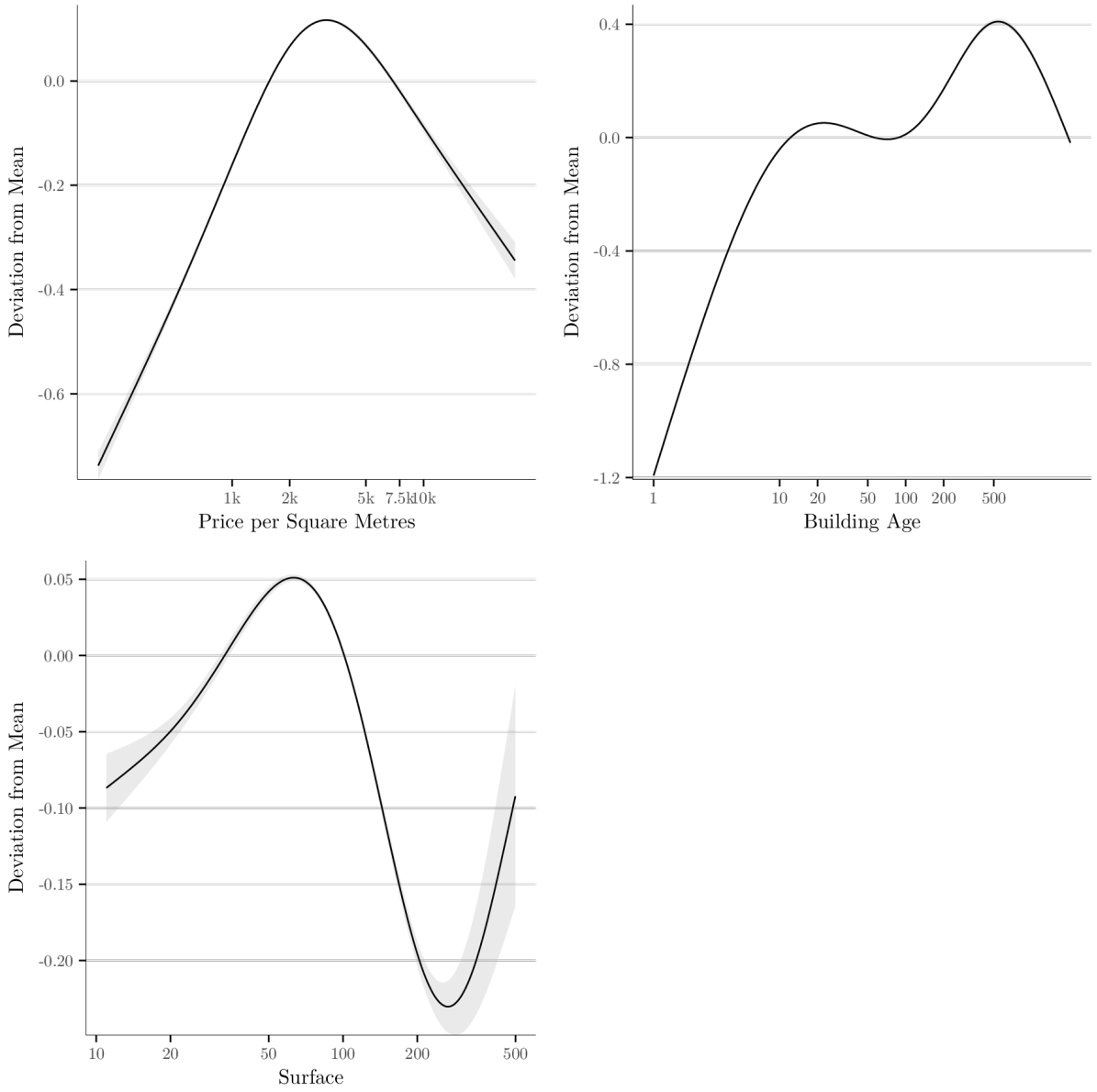


Figure B.1: Marginal contribution of univariate variables for selection

*Notes:* We report the marginal contribution of the univariate variable for the selection step. The outcome variable is the logarithm of the ration between the probability to be included in the sample and the probability to be excluded from the sample. The top-left panel represents the marginal contribution of the price per square meter, the top-right panel represents the marginal contribution of the building age, while the bottom-left panel represents the contribution of the surface expressed in square meter. 95% confidence intervals are shown in gray area.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## B.2 Selection Model: Overlap

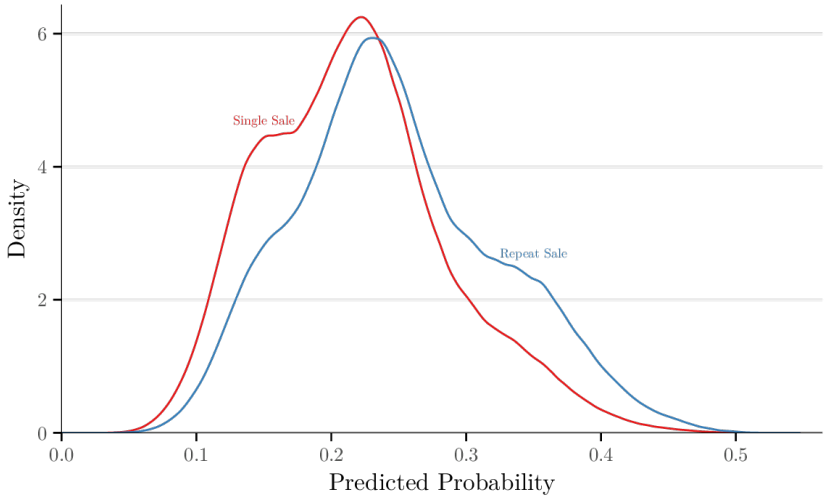


Figure B.2: Probability Distribution According to Sample of Interest

*Notes:* We report the distribution of the estimated probability to be included in the final sample for two population: housing units being sold only once (red line) and housing units being sold at least twice (blue line). The latter observations constitute the final sample of housing transaction. The estimation is performed based on the REML estimation of the logistic regression defined in Equation (2).
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## B.3 Second Step Regression: Control Variables

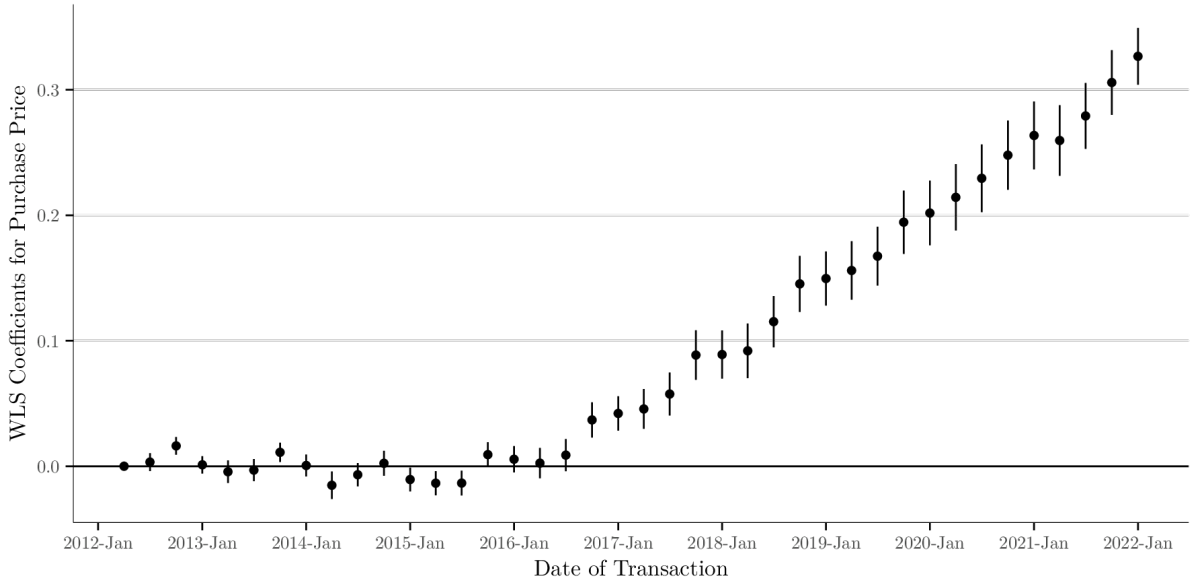


Figure B.3: WLS Results for Date

*Notes:* We report control variables for quarter dummies for the full model (column 6, Table 1). Standard errors at the 95% confidence intervals are also reported. The reference category is the first period of the sample.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## B.4 Second Step Regression: Control Variables

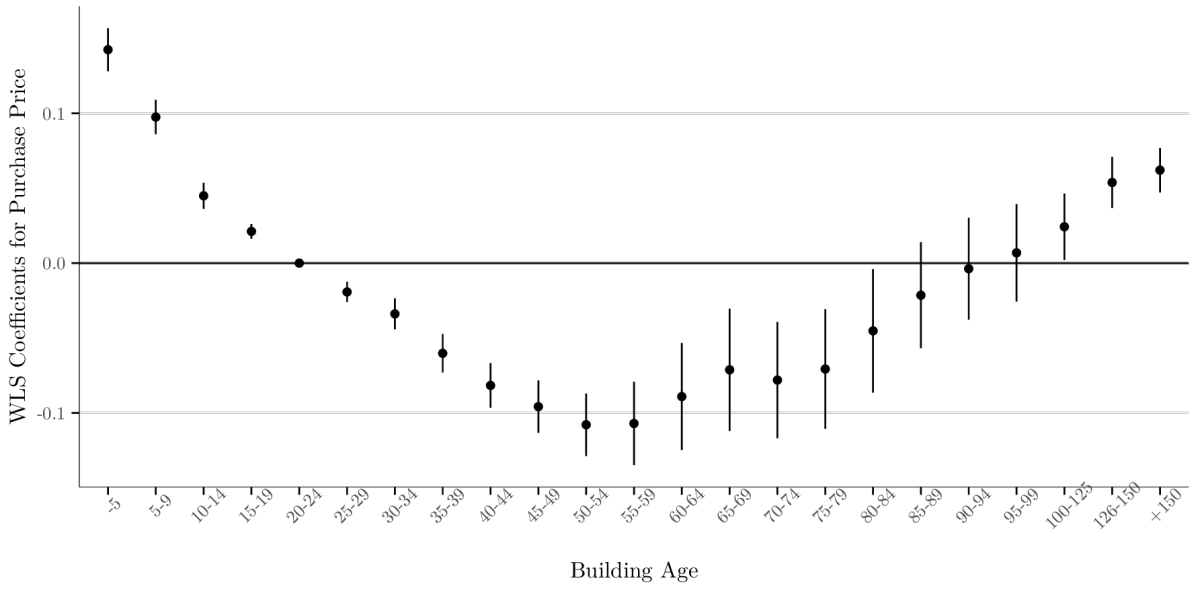


Figure B.4: WLS Results for Building Age Category

*Notes:* We report control variables for building age at the time of transactions for the full model (column 6, Table 1). Standard errors at the 95% confidence intervals are also reported. The reference category is 20–24 years.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## B.5 Second Step Regression: Alternative Results with Distance Control

Table B.5: WLS Results for Purchase Price Differences with Distance Controls

| | Dependent variable: Purchase price (log) | | | | |
|---|---|---|---|---|---|
| Covariate<br>Category (ref. Simple Purchaser-Seller) | (1) | (2) | (3) | (4) | (5) |
| Multi Purchaser-Seller | – | – | – | 0.015*** | 0.006*** |
| | | | | (0.001) | (0.001) |
| FTO | – | – | – | 0.024*** | 0.043*** |
| | | | | (0.002) | (0.001) |
| Small Inv. | – | – | – | 0.029*** | 0.029*** |
| | | | | (0.002) | (0.001) |
| Big Inv | – | – | – | 0.017*** | 0.010*** |
| | | | | (0.004) | (0.002) |
| Consumer-Investor | – | – | – | 0.024*** | 0.009*** |
| | | | | (0.003) | (0.002) |
| Tenure status: Own-occupied (ref.) | | | | | |
| Rent | – | -0.438*** | – | – | -0.036*** |
| | | (0.015) | | | (0.002) |
| Vacant | – | -0.292*** | – | – | -0.039*** |
| | | (0.011) | | | (0.002) |
| Other | – | -0.268*** | – | – | -0.019*** |
| | | (0.029) | | | (0.003) |
| Distance: <5km (ref.) | | | | | |
| 5km-20k | -0.070*** | -0.053*** | 0.019*** | 0.018*** | 0.018*** |
| | (0.013) | (0.010) | (0.002) | (0.002) | (0.001) |
| 21km-50km | -0.160*** | -0.136*** | 0.029*** | 0.028*** | 0.022*** |
| | (0.027) | (0.024) | (0.002) | (0.002) | (0.001) |
| 51km-200km | -0.244*** | -0.223*** | 0.037*** | 0.036*** | 0.027*** |
| | (0.031) | (0.030) | (0.002) | (0.002) | (0.001) |
| >200km | -0.047** | -0.057*** | 0.040*** | 0.039*** | 0.035*** |
| | (0.022) | (0.020) | (0.003) | (0.002) | (0.002) |
| N | 2,006,608 | 2,006,608 | 2,006,608 | 2,006,608 | 2,006,608 |
| $R^2$ | 0.014 | 0.140 | 0.939 | 0.939 | 0.960 |
| Adj. $R^2$ | 0.014 | 0.140 | 0.910 | 0.911 | 0.941 |
| Within $R^2$ | – | – | 0.004 | 0.006 | 0.345 |
| Housing Control | No | No | FE | FE | FE |
| Quarter Control | No | Yes | Yes | Yes | Yes |
| Selection Bias Control | No | No | No | No | Heckman |
| SE Clustered | Two-Way | Two-Way | Two-Way | Two-Way | Two-Way |

*Notes:* We report the results from the estimation of Equation (1) with the purchase price as main dependent variables. Compared to the main results, it includes a distance parameter to account for information asymmetries based on distance. Results are estimated using WLS with standard errors being two-way clustered (municipality and quarterly). Standard error are reported between parentheses.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.
*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$

## B.6 Robustness Checks: Duration Approach

Table B.6: OLS Results with Restricted Sample According to Duration between Two Sales

| | *Dependent Variable:* Price (log transformation) | | | |
|---|---|---|---|---|
| Covariate | 2 year | 3 year | 4 year | 5 year |
| Category (ref. Simple Buyer–Seller) | | | | |
| Multi Purchaser-Seller | -0.010*** | 0.001 | 0.004*** | 0.006*** |
| | (0.002) | (0.002) | (0.002) | (0.001) |
| FTO | 0.046*** | 0.056*** | 0.055*** | 0.052*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| Small Inv. | 0.022*** | 0.036*** | 0.038*** | 0.038*** |
| | (0.002) | (0.002) | (0.001) | (0.001) |
| Big Inv | -0.002 | 0.012*** | 0.017*** | 0.017*** |
| | (0.004) | (0.004) | (0.003) | (0.003) |
| Consumer-Investor | -0.007 | 0.000 | 0.006* | 0.006** |
| | (0.004) | (0.004) | (0.003) | (0.003) |
| Tenure status: Own-occupied (ref.) | | | | |
| Rent | -0.043*** | -0.043*** | -0.039*** | -0.036*** |
| | (0.005) | (0.004) | (0.003) | (0.002) |
| Vacant | -0.046*** | -0.047*** | -0.041*** | -0.038*** |
| | (0.005) | (0.004) | (0.003) | (0.003) |
| Other | -0.021** | -0.017** | -0.012** | -0.011** |
| | (0.009) | (0.007) | (0.005) | (0.005) |
| N | 375,969 | 650,891 | 1,001,103 | 1,358,339 |
| $R^2$ | 0.969 | 0.966 | 0.965 | 0.964 |
| Adj. $R^2$ | 0.957 | 0.952 | 0.950 | 0.949 |
| Within $R^2$ | 0.244 | 0.273 | 0.301 | 0.320 |

*Notes:* We report results for WLS estimation of Equation (1) with restricted samples as robustness checks. We restrict our sample to housing transactions for which the duration is lower than respectively to two, three, four and five years. We report only difference between purchaser category in comparison to the category of reference, being simple purchaser sellers. We report in parentheses two-way clustered standard errors.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## B.7 Robustness Checks: Price Variation Approach

Table B.7: OLS Results with Restricted Sample According to Price Evolution between Two Sales

| | *Dependent Variable:* Price (log transformation) | | | |
|---|---|---|---|---|
| Covariate | Top 1% | Top 5% | Top 10% | Top 20% |
| Category (ref. Simple Buyer–Seller) | | | | |
| Multi Purchaser-Seller | 0.007*** | 0.006*** | 0.006*** | 0.004*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| FTO | 0.042*** | 0.034*** | 0.031*** | 0.031*** |
| | (0.002) | (0.001) | (0.001) | (0.002) |
| Small Inv. | 0.031*** | 0.025*** | 0.023*** | 0.022*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Big Inv | 0.014*** | 0.012*** | 0.013*** | 0.012*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| Consumer-Investor | 0.008*** | 0.010*** | 0.008*** | 0.007*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| Tenure status: Own-occupied (ref.) | | | | |
| Rent | -0.031*** | -0.024*** | -0.019*** | -0.015*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| Vacant | -0.027*** | -0.013*** | -0.005** | 0.004** |
| | (0.003) | (0.002) | (0.002) | (0.002) |
| Other | -0.012*** | -0.011*** | -0.007** | -0.006** |
| | (0.004) | (0.003) | (0.003) | (0.003) |
| N | 2,219,527 | 2,116,484 | 1,996,649 | 1,763,076 |
| $R^2$ | 0.966 | 0.972 | 0.975 | 0.978 |
| Adj. $R^2$ | 0.951 | 0.960 | 0.965 | 0.968 |
| Within $R^2$ | 0.349 | 0.344 | 0.310 | 0.223 |

*Notes:* We report results for WLS estimation of Equation (1) with restricted samples as robustness checks. We restrict our sample to housing transactions for which the average price increase annually by less than 2.5%, 5%, 10%, 15%. We report only difference between purchaser category in comparison to the category of reference, being simple purchaser sellers. We report in parentheses two-way clustered standard errors.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.

## B.8 Capital Gains: Results with Level Difference as Alternative Outcome

Table B.8: WLS Results for Capital Gains Heterogeneity According to Each Group with Level Outcome

| | Difference | | | | | | |
|---|---|---|---|---|---|---|---|
| Covariate | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Multi Purchaser-Seller | 9,219.6*** | 9,562.8*** | 3,296.8*** | – | 5,650.2*** | – | 724.1** |
| | (2,555.4) | (1,502.5) | (431.0) | | (1,271.9) | | (361.4) |
| FTO | -7,645.7*** | -7,211.0*** | -2,743.7*** | – | -11,537.1*** | – | -7,227.3*** |
| | (384.3) | (935.5) | (186.4) | | (1,312.8) | | (432.5) |
| Small Inv. | -6,858.0*** | -1,708.8*** | 65.6 | – | -8,346.2*** | – | -5,539.1*** |
| | (954.2) | (480.2) | (735.2) | | (744.7) | | (442.8) |
| Big Inv | -2,786.6 | 4,650.2*** | 3,387.1*** | – | -3,192.0*** | – | -2,850.5*** |
| | (2,047.7) | (793.3) | (854.8) | | (576.8) | | (669.3) |
| Consumer-Investor | 14,590.7*** | 12,686.5*** | 5,406.1*** | – | 8,651.1*** | – | 3,097.2* |
| | (4,463.8) | (3,552.2) | (1,713.5) | | (3,303.5) | | (1,650.9) |
| Individual Deviation | – | – | – | -3,011.7*** | -3,037.5*** | -2,781.2*** | -2,804.7*** |
| | | | | (220.2) | (223.6) | (149.2) | (150.5) |
| Held Duration (years) | – | 1,146.8*** | 337.9*** | 1,931.6*** | 2,237.7*** | 1,191.6*** | 1,379.3*** |
| | | (200.3) | (107.3) | (164.8) | (186.1) | (63.6) | (62.9) |
| Tenure status: Own-occupied (ref.) | | | | | | | |
| Rent | – | -8,806.5*** | -2,697.7*** | -12,665.0*** | -13,036.1*** | -6,286.5*** | -6,583.7*** |
| | | (1,591.0) | (280.0) | (1,655.4) | (1,730.8) | (233.6) | (223.7) |
| Vacant | – | -1,457.3** | 1,190.0 | -4,800.5*** | -5,199.0*** | -1,646.3** | -1,852.1*** |
| | | (608.1) | (775.3) | (492.1) | (432.1) | (716.0) | (656.2) |
| Other | – | -3,471.6*** | -3,857.7*** | -5,929.7*** | -6,886.9*** | -5,322.1*** | -5,690.3*** |
| | | (673.8) | (545.8) | (651.1) | (568.7) | (431.7) | (528.0) |
| N | 1,354,529 | 1,354,529 | 1,354,529 | 1,354,529 | 1,354,529 | 1,354,529 | 1,354,529 |
| $R^2$ | 0.008 | 0.105 | 0.395 | 0.285 | 0.294 | 0.535 | 0.538 |
| Adj. $R^2$ | 0.008 | 0.105 | 0.383 | 0.285 | 0.294 | 0.526 | 0.529 |
| Within $R^2$ | – | 0.073 | 0.184 | 0.259 | 0.268 | 0.373 | 0.377 |
| Housing Control | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Time Control | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Muni FE | No | No | Muni | No | No | Muni | Muni |
| Heckman Correction | No | Yes | Yes | Yes | Yes | Yes | Yes |

*Notes:* We report results for WLS estimation of Equation (3) with level differences as alternative outcome. The first three columns only include housing characteristics and purchaser categories, whereas the last two include the individual deviation (namely the residuals of Equation (1)). We report only the difference between the purchaser category in comparison to the category of reference, being simple purchaser sellers and the individual deviation. We report in parentheses standard errors using a 1,000-iteration procedure. The estimation is performed on paired housing transactions about housing being sold at least twice between 2012 and 2021 at the country level to compute capital gains.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.
*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$

## B.9 Capital Gains: Results with Removal of Likely Renovation Works

Table B.9: WLS Results for Capital Gains Heterogeneity According to Each Group with Restriction of Sample

| | Relative Difference | | | | | | |
|---|---|---|---|---|---|---|---|
| Covariate | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Multi Purchaser-Seller | 0.002 | 0.001 | -0.004*** | – | -0.017*** | – | -0.018*** |
| | (0.001) | (0.001) | (0.001) | | (0.001) | | (0.001) |
| FTO | 0.001 | 0.005*** | 0.006*** | – | -0.018*** | – | -0.020*** |
| | (0.002) | (0.001) | (0.001) | | (0.001) | | (0.001) |
| Small Inv. | 0.003 | 0.001 | -0.001** | – | -0.032*** | – | -0.034*** |
| | (0.002) | (0.001) | (0.001) | | (0.001) | | (0.001) |
| Big Inv | 0.014*** | 0.008*** | 0.002 | – | -0.021*** | – | -0.025*** |
| | (0.003) | (0.002) | (0.002) | | (0.001) | | (0.001) |
| Consumer-Investor | 0.001 | 0.002 | -0.006* | – | -0.019*** | – | -0.020*** |
| | (0.003) | (0.003) | (0.003) | | (0.002) | | (0.002) |
| Individual Deviation | – | – | – | -0.015*** | -0.015*** | -0.015*** | -0.015*** |
| | | | | (0.000) | (0.000) | (0.000) | (0.000) |
| Held Duration (years) | – | -0.018*** | -0.017*** | -0.004** | -0.003* | -0.004* | -0.003 |
| | | (0.001) | (0.001) | (0.002) | (0.002) | (0.002) | (0.002) |
| Tenure status: Own-occupied (ref.) | | | | | | | |
| Rent | – | -0.006*** | -0.011*** | -0.024*** | -0.022*** | -0.031*** | -0.030*** |
| | | (0.002) | (0.001) | (0.002) | (0.002) | (0.001) | (0.001) |
| Vacant | – | -0.002 | -0.005*** | -0.015*** | -0.014*** | -0.018*** | -0.017*** |
| | | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Other | – | -0.005*** | -0.012*** | -0.018*** | -0.016*** | -0.021*** | -0.019*** |
| | | (0.002) | (0.001) | (0.002) | (0.002) | (0.001) | (0.001) |
| N | 1,097,823 | 1,097,823 | 1,097,823 | 1,097,823 | 1,097,823 | 1,097,823 | 1,097,823 |
| $R^2$ | 0.000 | 0.082 | 0.189 | 0.512 | 0.518 | 0.572 | 0.578 |
| Adj. $R^2$ | 0.000 | 0.082 | 0.171 | 0.512 | 0.518 | 0.563 | 0.568 |
| Within $R^2$ | – | 0.075 | 0.078 | 0.509 | 0.514 | 0.513 | 0.520 |
| Housing Control | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Time Control | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Muni FE | No | No | Yes | No | No | Yes | Yes |
| Heckman Correction | No | No | No | No | No | No | No |

*Notes:* We report results for WLS estimation of Equation (3) while restricting the sample of interest. We remove observations likely to be subjected to renovation works, using first residuals under -0.2 as criteria. The first three columns only include housing characteristics and purchaser categories, whereas the last two include the individual deviation (namely the residuals of Equation (1)). We report only the difference between the purchaser category in comparison to the category of reference, being simple purchaser sellers and the individual deviation. We report in parentheses standard errors using a 1,000-iteration procedure. The estimation is performed on paired housing transactions about housing being sold at least twice between 2012 and 2021 at the country level to compute capital gains.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.
*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$

## B.10 Capital Gains: Results with Removal of Likely Renovation Works

Table B.10: WLS Results for Capital Gains Heterogeneity According to Each Group with Restriction of Sample

| | Relative Difference | | | | | | |
|---|---|---|---|---|---|---|---|
| Covariate | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Multi Purchaser-Seller | 0.006*** | 0.001 | -0.007*** | – | -0.021*** | – | -0.023*** |
| | (0.002) | (0.002) | (0.002) | | (0.001) | | (0.001) |
| FTO | 0.011*** | 0.006*** | 0.008*** | – | -0.023*** | – | -0.024*** |
| | (0.003) | (0.001) | (0.001) | | (0.002) | | (0.001) |
| Small Inv. | 0.012*** | 0.001 | -0.003*** | – | -0.040*** | – | -0.041*** |
| | (0.002) | (0.002) | (0.001) | | (0.001) | | (0.001) |
| Big Inv | 0.026*** | 0.012*** | 0.004 | – | -0.026*** | – | -0.030*** |
| | (0.004) | (0.004) | (0.003) | | (0.002) | | (0.002) |
| Consumer-Investor | 0.007** | 0.003 | -0.007* | – | -0.022*** | – | -0.023*** |
| | (0.003) | (0.003) | (0.004) | | (0.002) | | (0.002) |
| Individual Deviation | – | – | – | -0.018*** | -0.018*** | -0.018*** | -0.018*** |
| | | | | (0.000) | (0.000) | (0.000) | (0.000) |
| Held Duration (years) | – | -0.009*** | -0.008*** | 0.002 | 0.003 | 0.002 | 0.003 |
| | | (0.001) | (0.001) | (0.002) | (0.002) | (0.002) | (0.002) |
| Tenure status: Own-occupied (ref.) | | | | | | | |
| Rent | – | -0.008*** | -0.014*** | -0.028*** | -0.026*** | -0.037*** | -0.035*** |
| | | (0.003) | (0.001) | (0.002) | (0.002) | (0.002) | (0.002) |
| Vacant | – | -0.001 | -0.004*** | -0.017*** | -0.016*** | -0.020*** | -0.019*** |
| | | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Other | – | -0.005** | -0.014*** | -0.021*** | -0.018*** | -0.023*** | -0.021*** |
| | | (0.003) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| N | 1,239,549 | 1,239,549 | 1,239,549 | 1,239,549 | 1,239,549 | 1,239,549 | 1,239,549 |
| $R^2$ | 0.001 | 0.070 | 0.198 | 0.596 | 0.602 | 0.649 | 0.655 |
| Adj. $R^2$ | 0.001 | 0.070 | 0.182 | 0.596 | 0.601 | 0.642 | 0.648 |
| Within $R^2$ | – | 0.039 | 0.049 | 0.583 | 0.588 | 0.584 | 0.590 |
| Housing Control | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Time Control | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Muni FE | No | No | Yes | No | No | Yes | Yes |
| Heckman Correction | No | No | No | No | No | No | No |

*Notes:* We report results for WLS estimation of Equation (3) while restricting the sample of interest. We remove observations likely to be subjected to renovation works, using first residuals under -0.2 as criteria. The first three columns only include housing characteristics and purchaser categories, whereas the last two include the individual deviation (namely the residuals of Equation (1)). We report only the difference between the purchaser category in comparison to the category of reference, being simple purchaser sellers and the individual deviation. We report in parentheses standard errors using a 1,000-iteration procedure. The estimation is performed on paired housing transactions about housing being sold at least twice between 2012 and 2021 at the country level to compute capital gains.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.
*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$

## B.11 Capital Gains: Results with Removal of Likely Renovation Works

Table B.11: WLS Results for Capital Gains Heterogeneity According to Each Group with Restriction of Sample

| | Relative Difference | | | | | | |
|---|---|---|---|---|---|---|---|
| Covariate | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Multi Purchaser-Seller | 0.010*** | 0.001 | -0.008*** | – | -0.025*** | – | -0.026*** |
| | (0.002) | (0.002) | (0.002) | | (0.001) | | (0.002) |
| FTO | 0.018*** | 0.007*** | 0.008*** | – | -0.027*** | – | -0.030*** |
| | (0.003) | (0.002) | (0.001) | | (0.002) | | (0.001) |
| Small Inv. | 0.020*** | 0.001 | -0.004*** | – | -0.047*** | – | -0.048*** |
| | (0.002) | (0.002) | (0.001) | | (0.001) | | (0.001) |
| Big Inv | 0.044*** | 0.021*** | 0.010*** | – | -0.029*** | – | -0.033*** |
| | (0.003) | (0.003) | (0.003) | | (0.002) | | (0.002) |
| Consumer-Investor | 0.008* | 0.001 | -0.009** | – | -0.027*** | – | -0.027*** |
| | (0.004) | (0.004) | (0.004) | | (0.002) | | (0.002) |
| Individual Deviation | – | – | – | -0.021*** | -0.022*** | -0.021*** | -0.021*** |
| | | | | (0.001) | (0.000) | (0.000) | (0.000) |
| Held Duration (years) | – | -0.002 | -0.002 | 0.006** | 0.007** | 0.006** | 0.007** |
| | | (0.002) | (0.002) | (0.003) | (0.003) | (0.003) | (0.003) |
| Tenure status: Own-occupied (ref.) | | | | | | | |
| Rent | – | -0.006** | -0.016*** | -0.031*** | -0.028*** | -0.041*** | -0.039*** |
| | | (0.003) | (0.001) | (0.002) | (0.002) | (0.002) | (0.002) |
| Vacant | – | 0.003* | -0.002* | -0.018*** | -0.016*** | -0.021*** | -0.020*** |
| | | (0.002) | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) |
| Other | – | -0.004 | -0.015*** | -0.021*** | -0.018*** | -0.025*** | -0.022*** |
| | | (0.003) | (0.002) | (0.003) | (0.003) | (0.002) | (0.002) |
| N | 1,310,395 | 1,310,395 | 1,310,395 | 1,310,395 | 1,310,395 | 1,310,395 | 1,310,395 |
| $R^2$ | 0.002 | 0.083 | 0.220 | 0.672 | 0.677 | 0.716 | 0.720 |
| Adj. $R^2$ | 0.002 | 0.083 | 0.205 | 0.672 | 0.677 | 0.710 | 0.715 |
| Within $R^2$ | – | 0.035 | 0.050 | 0.655 | 0.660 | 0.653 | 0.659 |
| Housing Control | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Time Control | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Muni FE | No | No | Yes | No | No | Yes | Yes |
| Heckman Correction | No | No | No | No | No | No | No |

*Notes:* We report results for WLS estimation of Equation (3) while restricting the sample of interest. We remove observations likely to be subjected to renovation works, using first residuals under -0.2 as criteria. The first three columns only include housing characteristics and purchaser categories, whereas the last two include the individual deviation (namely the residuals of Equation (1)). We report only the difference between the purchaser category in comparison to the category of reference, being simple purchaser sellers and the individual deviation. We report in parentheses standard errors using a 1,000-iteration procedure. The estimation is performed on paired housing transactions about housing being sold at least twice between 2012 and 2021 at the country level to compute capital gains.
*Sources:* Authors' Calculation from DV3F and Property Tax Files.
*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$