Winners and Losers: The Individual Consequences of Price Deviation on Capital Gains

Rémi Lei*†
remi.lei@psemail.eu

March 2025

Abstract

We empirically assess how purchase price deviations affect the formation of housing capital gains. Making the best of the universe of transactions in France between 2012 and 2022, we first estimate an individual price deviation through a repeat sales approach with *Heckman* procedure to account for potential selection bias, controlling for experience in the housing market. Then, we regress realized capital gains on the individual deviation to assess the substitution between purchase price deviation and capital gains. Our results highlight that a €1 increase in purchase price decreases capital gains by €0.51. We expect part of these individual deviations to affect the local housing market through future purchasers and seller's beliefs, leading to a mitigation of the capital gains differences.

JEL classification: R2 ; R3 ; D1

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

*EU Tax Observatory, Paris School of Economics, F-75014 Paris, France
†UMR CNRS 7300 ESPACE, Avignon Université, F-84029 Avignon, France

1 Introduction

The formation of capital gains in the housing market is a major determinant of wealth accumulation and related inequality. The uneven distribution of local housing market dynamics through space generates sizeable heterogeneity in realized capital gains (Eggum and Røed Larsen, 2024). The most expensive housing units have benefitted over the last decades from the greater price appreciation while providing greater capital returns (Fagereng et al., 2020; Bach, Calvet and Sodini, 2020). Consequently, homeowners are considered winners and losers in the housing market according to their assets' location, regardless of how they behave through either the purchase or the selling processes. It appears that wealth accumulation is an underlying process in which homeowners have no or little impact on their accumulation.

Yet, this omits individual behaviour that may affect the realization of capital gains and the local housing market dynamics. Through their purchase price, new homeowners will likely affect the local housing market dynamics through a signalling effect. Future purchasers and real estate will use this transaction as observations to set their beliefs and expectations about the local housing market. Hence, an overpayment might fuel the local housing market dynamics and therefore affect the wealth of local homeowners. The current literature has focused on the role of the wealthiest homeowners (Aiello, Kotter and Schubert, 2022) and demonstrates that the overpayments are capitalized in the local housing market dynamics. In this paper, we fill the gap by estimating the substitution between price deviation at the time of the purchase and the realization of capital gains for the entire universe of homeowners.

We exploit a dataset that registers all housing transactions for France that occurred between 2012 and 2022. It derives from administrative data with characteristics about housing units such as location with high-granularity. In addition, we make the best of Lei et al. (2024) improvement of French administrative data about homeowners to qualify purchasers according to their portfolio size (both through number of properties and gross housing wealth), municipality of residence or age. Finally, 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 aim to proxy for experience and bargaining power that may affect purchase price, and hence capital gains. The invariant housing ID allows us to compute capital gains for housing units being sold twice between 2012 and 2022 and observe homeowners that realize it.

The main threat for 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), it assumes that unobservable characteristics are randomly distributed across purchaser groups. In practice, it assumes implicitly 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. Indeed, first-time owners smooth their consumption to access homeownership (Waxman et al., 2020) considering the housing affordability crisis, hence they probably downgrade the housing quality they purchase, especially trading off unobservable characteristics. Hence, we need a price deviation that is orthogonal to housing characteristics, including non-observables, and purchasers characteristics.

We handle the potential omitted variable bias to recover an individual price deviation by introducing housing fixed effects (for application, see e.g. Dorsey et al., 2010) in line with the repeat sales approach. It releases the assumption that unobservable characteristics are evenly distributed between purchasers category and provide individual deviation being orthogonal to housing characteristics. Nonetheless, we assume that variation in unobservable characteristics over time is randomly distributed between categories. Even if by nature the assumption is not verifiable, robustness checks do not challenge it. We then regress the realized capital gains on the individual price deviation, to assess how it affects the formation of capital gains. The drawback of the introduction of housing fixed effects to recover individual deviation is the restriction of the transaction sample to units being sold at least twice between 2011 and 2022, which might raise some sample selection issues. We handle this potential threat through the estimation of the probability of each housing transaction to be sold at least twice and thus included in the sample. The estimated probability is introduced 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 coordinates to account for spatial heterogeneity including non-observable one (Dupont, Marques and Kneib, 2023; Gilbert et al., 2024). Finally, to proxy for purchaser characteristics, we introduce group-specific effects that account for experience and financial constraints that may affect purchase price through heterogeneous bargaining power.

Performance on the housing market for purchasers generates sizeable differences in capital gains. We estimate that a €1 increase in price deviation decreases the realized capital gains by €0.51, making the substitution effect incomplete. We expect part of the overpayment to fuel the local housing market dynamics, following a signalling effect affecting the price expectation of future purchasers and real estate agents. Alternative mechanisms such as better skills leading to higher selling prices are ruled out by empirical analysis. In addition to the individual effect, the less experienced and financially constrained pur-

chasers (i.e. the first-time owners) experience the highest group-specific positive deviation. When turning to capital gains, they also benefit from lower capital gains, even if the loss is not as important as the purchase price deviation. These results support the presence of a spillover mechanism of price deviation of the local housing market through a signalling effect. Similar results are found for alternative groups of purchasers. Finally, change in the unobservable housing characteristics is not likely to drive these results, according to robustness checks.

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 it affects the realization of capital gains. The wealthiest homeowners better anticipate housing market downturns (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 demonstrate that all purchasers are likely to affect their capital gains and disturb the local housing market dynamics. 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 by focusing on price deviation according to purchasers' 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 the entire population of homeowners. Our classification is the first to include the less experienced purchasers (namely the first-time owners), which in addition is 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. We expect mechanisms such as lower bargaining power and inattention to drive these results. 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 lower price compared to owner-occupied ones (-2.7% for rental purchase compared to owner-occupied ones). Our estimations are opposed to those obtained by Biagi et al. (2021). We expect the fact that owner-occupiers are a highly heterogeneous category 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 an increase in interest to pay. 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 individual price deviation 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.

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 areas, type of housing, spatial coordinates or the presence of certain facilities (basement, parking space, etc.). An invariant housing ID is also available, that enables tracking multiple sales for similar housing units over time. In addition, characteristics of the transaction are provided, such as the agreement date and the price. This is particularly convenient to compute the realized capital gains at the housing level.

However, the dataset lacks information about purchaser and seller characteristics. The only available information to qualify both is whether they are legal or physical persons. Hence, this information is not sufficient to control for experience and financial constraints, two variables likely to significantly affect purchase price through bargaining power.

2.2 Property Tax Files

To address the lack of information about purchasers in the transaction dataset, we utilize French property tax files. These data link housing units to their current owners annually since 2011. By leveraging recent advancements that transform 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 for each housing transaction. With this information, we also track the overall housing portfolio over time for both purchasers and sellers.

This 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 a purchases a home). Hence, as wealth becomes more individualized over time (Frémeaux and Leturcq, 2020),the number of properties is not always homogeneous within the purchasers' group. We choose to use the relationship between purchasers and housing units as the finest level, and weight according to the inverse of the number of owners. By doing so, we make the best of the individual level and do not introduce additional assumptions about the purchaser structure.

2.3 Portfolios' Based Categories

To proxy for experience and bargaining power, we leverage the set of information about the housing portfolio over time for purchasers. Our category definition aims to overlap with common ones in the existing literature to provide results as comparable as possible for group-specific deviation. Indeed, our classification includes the small and medium real estate investors, owner-occupiers and rental investment companies. We nonetheless consider these categories to be highly heterogeneous within their group, especially for owner-occupiers. This category aggregates purchasers such as first-time owners, homeowners achieving residential mobility or even purchases for secondary homes. Similarly, we expect purchasers within the Small and Medium Real Estate Investors (SMREI in the remainder of the paper) group to be highly different, especially in regard to their portfolio size.

Hence, we extend common classification in the literature about homeowners and purchase price differences in two ways. First, we account for the portfolio size to proxy for experience in the housing market. The underlying assumption is that more experienced purchasers are more likely to reduce housing prices during the bargaining process. In addition, it aims to differentiate small real estate investors from big ones, which might differentiate according to their knowledge level and aggressiveness in the bargaining process. Second, we account for whether purchasers benefit from a simultaneous sale of a

housing unit previously held. It includes homeowners achieving residential mobility or real investors who sell part of their portfolio to finance their new purchase. It aims to proxy for the financial constraint, as individual who finance their purchase without a joint selling are more likely to borrow. Hence, we consider that the financial constraint might be greater for these purchasers.

We finally obtained eight purchaser categories: first-time owners, single and multiple purchaser-sellers which are assimilated as homeowners achieving residential mobility, small and big investors i.e. purchasers with respectively less than five and more than five housing units and increased their portfolio size, rental investment companies (named *Société Civile Immobilière*), and consumer-investor, i.e. purchaser that sells more properties than purchase. Although these categories aim to control for experience in the housing market and financial constraints, the group-specific deviations are relevant to understanding the wealth accumulation mechanism as well. We also test as an alternative a continuous classification which does not affect the results.

2.4 Descriptive Statistics

l Our overall dataset about transactions is composed of 4,807,190 observations from 2012 to 2021 being purchased either by individuals or rental investment companies. 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 Section A.7). However, their choices differ significantly according to purchase price and unit size.

First-time owners and investors, who are likely to borrow to purchase their housing units, purchase at a lower price than those who benefit jointly from housing sales. These differences in purchase price result mainly from heterogeneity in housing choice, including unit size. First-time owners and investors purchase smaller units than individuals who benefit from a sale at the time of their purchase.

However, 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 facts. 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 1). On the opposite, 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 in Section A.9). Second, housing choices might differ according to unobservable characteristics. Investors might sort into higher qual-

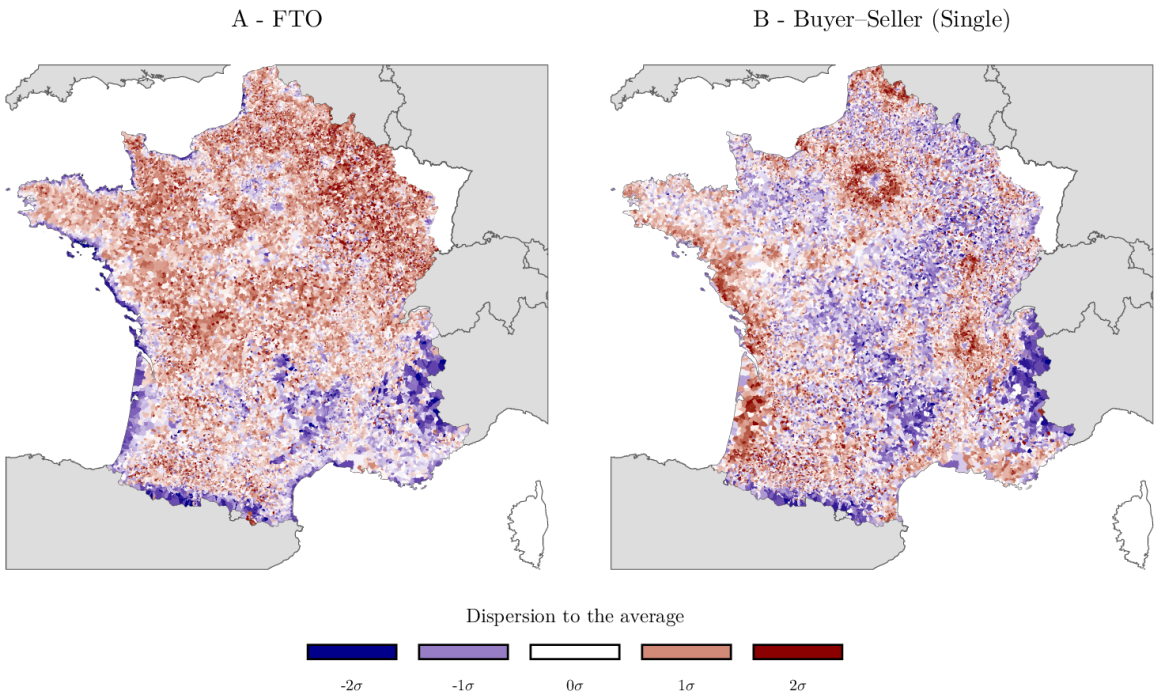


Figure 1: Preferred Location According to Purchaser Category

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.

ity housing units, whereas first-time owners might smooth consumption of unobservable characteristics (Halket, Nesheim and Oswald, 2020).

3 Identification Strategy

The estimation of an individual deviation requires controlling for housing heterogeneity and purchaser characteristics. Whereas the hedonic approach is widely used to control for observable heterogeneity in housing characteristics, we expect it to be inefficient in regards to the implicit assumptions. Consequently, we rather adopt a repeat sales approach to introduce housing fixed effects. We consider that it fits best in comparison to the hedonic method to control for omitted variable bias, which is the main threat we face. We first present a selection model to account for potential selection bias resulting from the restriction of the transaction sample to housing units being sold at least twice (Dorsey et al., 2010).

3.1 Selection Model

The repeat sales approach imposes restrictions on transaction samples to units being sold at least twice. It raises concern about selection bias as these observations might differ significantly from the full sample, either on observable or unobservable characteristics. To correct for potential selection bias, we then estimate the probability for each observation to be included in the final sample, i.e. being sold at least twice.

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

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

with p probability to be sold at least twice, $\mathbf{X}$ observable characteristics of the housing, and ε 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 Restricted Maximum Likelihood method (hereafter REML), the functional form for each variable is endogenously determined using penalized regressions that aim trade-off between under-fitting and over-fitting. In addition, we make the best of 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. Hence, in comparison to spatial fixed effects, the spatial contribution varies smoothly over space, without strong discontinuity. Similarly to univariate variables, the spatial contribution is estimated using the REML approach, with penalized regression to determine the appropriate degrees of freedom for transformation. The set of univariate variables includes the surface, the price per square meter, 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 (1), we estimate for each housing unit j in the transaction sample the probability $\widehat{p}_j$ to be sold at least twice between 2011 and 2023 and consequently to be included in the final sample.

3.2 Repeat Sales Approach

To control for both observed and unobserved characteristics in the regression, the repeat sales method represents an appealing alternative (Dorsey et al., 2010), widely used in the

housing economics literature to control for unobservable characteristics (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

$$y_{ijt} = \eta_j^{RS} + \beta_1^{RS} C_i + \beta_2^{RS} \widehat{p}_{jt} + \mathbf{X}\gamma^{RS} + \mathbf{T}\delta^{RS} + \zeta_{ijt}^{RS} \tag{2}$$

with y_{ijt} outcome of interest for housing j purchased by individual j at time t, C_j purchaser category for individual i, $\widehat{p}_{jt}$ estimated probability to be sold at least twice, η_j housing fixed effects, $\mathbf{T}$ time dummies, and ζ_{ijt} the error term. The housing fixed then 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 a strong determinant of housing price to affect significantly the price and bias the identification of an individual deviation. In addition, unobservable characteristics are likely to be different across the group of purchasers, providing bias estimation. The only remaining threat for the identification of individual and group-specific price deviation is thus characteristics that vary over time (e.g. gentrification or renovation works), which are included in the error term. To lower this threat, we also 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 tenure status premium (Biagi et al., 2021, see e.g.).

Then, we assume that the error term ζ_{ijt}^{RS} includes two terms. First, an individual deviation that is unexplained. It might introduce individual shock, unobservable preferences for specific housing assets or inattention. 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 being captured by the housing fixed effect. Empirically, we expect renovation works to be marginal and assume that the error term mainly aggregates individual shock. Robustness checks removing observations that are the most likely to be concerned by renovation works do not affect our conclusions.

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 long-term. Indeed, while we expect positive price deviations to lower realized capital gains, two mechanisms might be in place. First, positive price deviation might spill to the entire local market through a signalling effect. Second, purchasers associated with a positive price deviation might sell at a higher price than their counterparts. Consequently, we investigate whether there is a correlation between the purchase price

and the final realized capital gains, both at the individual and the group level. We assume that it exists a negative correlation between price deviation and realized capital gains. 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 ξ_{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 (2). This equation is estimated through WLS, with standard errors being clustered at the municipality level. Finally, we introduce ζ^{RS}, the individual unexplained deviation from Equation (2) 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 3.2), 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 modification of housing characteristics are likely to bias our estimation. Yet, considering the rate of renovation works, we assume this to be marginal. 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 present purchase price deviation results following the repeat sales method. 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 2011 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 for a housing unit to be 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 function form. We report the contribution for price per square meter and location in Figure 2.

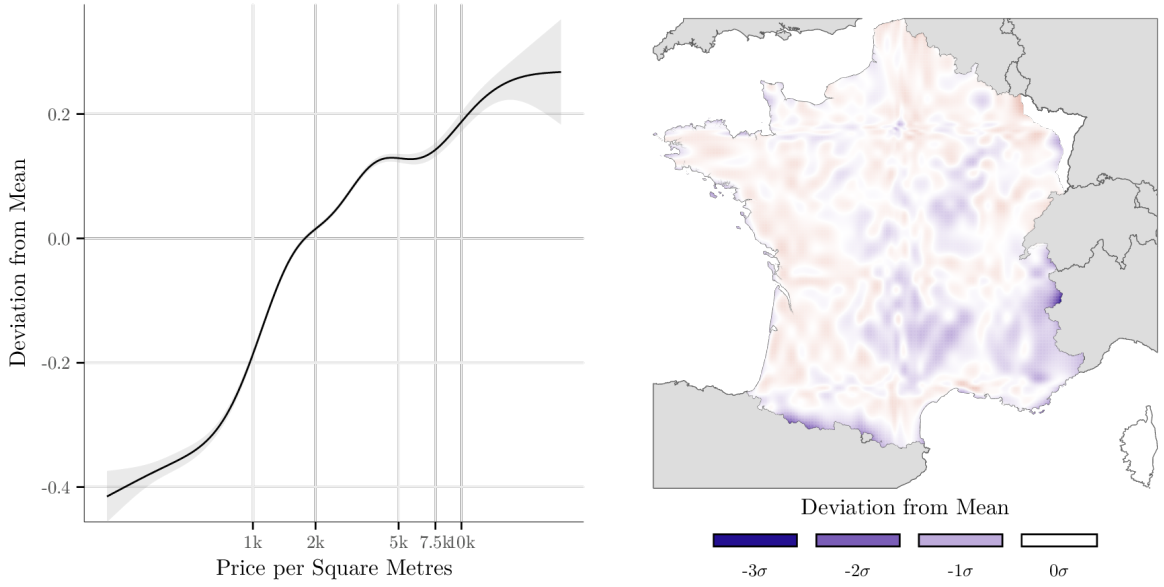


Figure 2: Distribution of Probability to Achieve a Repeat Sales per Category

Notes: The left panel represents the distribution of probability to purchase repeat sales within each purchaser category. These results come from the estimation of Equation (1). Based on this model, we estimate for each housing transaction the probability to be included in the sample. The right panel represents the spatial smoothing splines included in Equation (1). Positive values indicate that housing located in these areas is more likely to be included in the sample.
Sources: Authors' Calculation from DV3F and Property Tax Files.

As expected based on descriptive statistics, the sample restricted to housing units being sold at least twice differs significantly from the universe of housing transactions. Housing units with higher prices per square meter are more likely to be included in the sample (Figure 2), whereas observations located in tourist areas are less likely to be included. Nonetheless, the overlap for the distributions of probability to belong to the final sample for both groups is large, which supports the fact that controlling for probability might reduce the likelihood of having selection bias. The only remaining threat lies in unobservable variables.

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 to be included in the final sample as control vari-

ables *à la Heckman* to correct for potential selection bias. The estimation procedure is performed using Weighted Least Squares, with two-way clustered standard errors. 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 Category (ref. Simple Purchaser-Seller)	(1)	(2)	(3)	(4)	(5)	(6)
Multi Purchaser-Seller	0.056*	0.067***	0.014***	0.014***	0.013***	0.011***
	(0.030)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
FTO	-0.170***	-0.026***	0.035***	0.035***	0.038***	0.035***
	(0.013)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Companies	-0.132**	0.082***	0.020***	0.020***	0.021***	0.013***
	(0.052)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Small Inv.	-0.255***	0.011***	0.032***	0.032***	0.034***	0.031***
	(0.027)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Big Inv	-0.309***	0.007***	0.015***	0.015***	0.016***	0.014***
	(0.044)	(0.002)	(0.001)	(0.003)	(0.003)	(0.003)
Consumer-Investor	0.136***	0.085***	0.016***	0.016***	0.015***	0.015***
	(0.025)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Tenure status: Own-occupied (ref.)						
Rent	–	-0.132***	-0.007***	-0.007***	-0.008***	-0.027***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Vacant	–	-0.088***	-0.005***	-0.005**	-0.007***	-0.030***
		(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
Other	–	-0.042***	0.013***	0.013***	0.011***	-0.004
		(0.002)	(0.002)	(0.004)	(0.004)	(0.004)
N	2,306,015	2,306,015	2,306,015	2,306,015	2,306,015	2,306,015
R^2	0.026	–	0.957	0.957	0.958	0.961
Adj. R^2	0.026	0.715	0.939	0.939	0.941	0.945
Within R^2	–	–	0.290	0.290	0.268	0.352
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 (2) 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.
Sources: Authors' Calculation from DV3F and Property Tax Files.
*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$

The use of housing fixed effects has a dramatic 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 appear to underpay by 2.6% in comparison to homeowners that achieve residential mobility. Yet, introducing housing fixed effects improves the explained variance by more than 20pp and affects significantly group-specific deviation. First-time owners represent the category that pays the most for housing controlling for observable and non-observable characteristics (+3.5%). In addition, small and big investors also overpay by respectively 3.1% and 1.4%. Hence, it appears that purchasers with a lower level of

experience and being more financially constrained have the highest price deviations.

In addition to significant changes when accounting for unobservable characteristics, price differences based on tenure status change dramatically as well. The introduction of housing fixed effects lowers these differences (from -12.4% to -0.7%, 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. It nonetheless highlights that unobserved quality varies significantly according to tenure status.

The correction based on probability to be a repeat sales have little impact on our results (columns 4, 5 and 6, Table 1). It nonetheless affects the significance for consumers (i.e. individuals that sell more housing than they purchase). 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). We compute the distance between the purchaser's main residence (available in the fiscal files) and the housing location. It removes rental investment companies from our transaction sample. The introduction of these control variables does not affect our results, whereas we recover the expected sign for distance (purchase price increases with distance, see in Section B.4).

In addition, we assess whether our results are robust to variation in unobservable housing characteristics, e.g. renovation works. First, we remove observations for which the duration between the two sales is superior to respectively two, three or four years. The assumption is that the lowest the duration between two sales, the more likely the unobservable characteristics are similar. Second, we remove from our sample housing units that experience an important variation of housing per square meter during the two sales. If some renovation works are achieved, then the price per square meter is likely to increase. We retain arbitrary values (respectively 2.5%, 5%, 10% and 15% of increase per year) as the threshold for removing specific transactions. In both cases, the hierarchy between our purchaser categories remains unaffected. Although it affects the magnitude of price difference (see in Section B.5 and Section B.6), it does not negate our main results.

4.3 Consequences on realized capital gains

Finally, we close the results section with the differences based on purchasers' categories of 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 selling as control variables, using quarter dummies. Results are reported in Table 2, whereas we provide similar results with alternative outcomes in Section B.7 (level of capital gains) and Section B.8 (selling price).

Table 2: WLS Results for Capital Gains Heterogeneity According to Each Group

	Relative Difference						
Covariate	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Multi Purchaser-Seller	0.008***	0.008***	0.008***	0.002	-0.008**	–	-0.029***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)		(0.001)
FTO	-0.007	-0.008*	-0.016***	-0.016***	-0.015***	–	-0.047***
	(0.005)	(0.004)	(0.003)	(0.003)	(0.002)		(0.001)
Companies	0.028***	0.025***	0.029***	0.011*	-0.001	–	-0.050***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)		(0.003)
Small Inv.	-0.004	-0.009**	-0.012***	-0.022***	-0.026***	–	-0.063***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)		(0.001)
Big Inv	0.042***	0.037***	0.037***	0.022***	0.008	–	-0.039***
	(0.006)	(0.006)	(0.007)	(0.006)	(0.005)		(0.002)
Consumer-Investor	-0.001	0.001	0.002	-0.003	-0.013**	–	-0.029***
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)		(0.002)
Individual Deviation	–	–	–	–	–	-0.025***	-0.025***
						(0.000)	(0.000)
N	1,378,158	1,378,158	1,378,158	1,378,158	1,378,158	1,378,158	1,378,158
R^2	0.001	0.003	0.046	0.065	0.176	0.826	0.829
Adj. R^2	0.001	0.003	0.046	0.065	0.159	0.823	0.826
Within R^2	–	–	0.004	0.025	0.039	0.797	0.801
Housing Control	No	Yes	Yes	Yes	Yes	Yes	Yes
Time Control	No	No	Yes	Yes	Yes	Yes	Yes
Muni FE	No	No	No	No	Yes	Yes	Yes
Heckman Correction	No	No	No	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 (2)). 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

The introduction of the unexplained individual deviation in purchase price (namely $\widehat{\zeta}_{ijt}^{RS}$) increases dramatically the share of explained variance (column 5, Table 2). An increase of 1pp in purchase price leads to a decrease of capital gains deviation on average by 2.5pp. When converting this into euro, the substitution is above, as €1 increase in purchase price decreases the capital gains by €1.25. However, considering that the realized capital gains are also affected by the selling price, we estimate that the selling price is also positively correlated with the individual deviation. We estimate that a 1pp of increase in the individual deviation decreases the purchase price by 0.7pp. This relationship is likely to be supported by the fact that individual who overpay their housing units due to low experience are also less likely to sell at lower prices. These results are consistent with Biagi et al. (2021) results. Hence, from the €1.26 decrease in capital gains, €0.74 are explained by the selling, leading to a €0.51 loss, which is lower than one.

When turning to capital gains, the group-specific deviations also change significantly. Categories that experience positive price deviation do not experience necessarily negative

capital gains deviation. On the one hand, rental investment companies and big investors benefit from higher capital gains in comparison to homeowners achieving residential mobility (respectively +2.5pp and +1.7pp). On the other hand, small investors and first-time owners do not achieve the same level of capital gains compared to homeowners achieving residential mobility. Yet, turning these differences in price euros, the difference in capital gains is lowered than the difference for the first purchase.

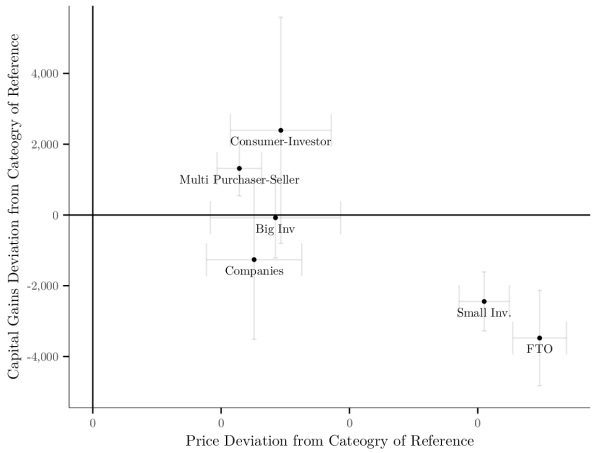


Figure 3: Bivariate Results for Purchase Price and Capital Gains

Notes: We report the difference in euro for each purchaser category for the purchase price difference (Equation (2)) and the capital gains (Equation (3)). The category of reference is homeowners achieving a residential mobility. We report standard errors in light grey for both equations.
Sources: Authors' Calculation from DV3F and Property Tax Files.

As for individuals, the substitution is not complete for group-specific estimation. Reported in Figure 3, we nonetheless have two types of groups. First, despite having a positive price deviation, they realize on average the same capital gains as homeowners achieving residential mobility. Moreover, for individuals with large portfolios, they are even more likely to generate greater capital gains, which is the case of multiple buyer-seller. Second, the less experienced purchasers according to their portfolio size realize lower capital gains than the category of interest. Yet, the loss is not as important as the deviation at the time of the purchase.

5 Discussion

We close the article by discussing our results in regard to the underlying mechanisms in place, the consequences on wealth accumulation, and the sorting of non-observable housing characteristics.

5.1 Underlying Mechanisms

Our results conclude that the substitution between individual price deviation and realized capital gains is not equal to one in absolute value. According to our expectations, a €1 increase should transit into a €1 loss in capital gains, conditionally on the market dynamics and selling skills. Hence, we expect the housing market dynamics or price beliefs to drive this imperfect substitution between purchase price and capital gains.

We expect the individual deviation to affect the local housing market dynamics. The individual deviation is at least partially transmitted to the local housing market dynamics and mitigates the loss in capital gains on the long term. By purchasing at a higher price, is likely to affect the price belief (and thus price listings) of current homeowners, potential purchasers and real estate agents. With the recent development of open-data transactions, citizens are likely to form their price expectations based on previous housing sales (Ben-Shahar and Golan, 2019). Hence, this might affect the future price expectations in the local housing market, and thus the value of all housing units. Therefore, a positive price deviation might fuel the local housing market through future purchasers and sellers' expectations. We expect the upper bound of the magnitude of this spillover effect to reach 49%, i.e. the case where the entire price deviation is capitalized in local values.

However, we might expect a change in unobservable housing quality to bias our estimation, leading to an overestimation of the spillover mechanism. Indeed, the common pattern for housing renovations is a low purchase price, whereas the second purchase price will be higher, considering the renovation works. In this case, the housing fixed effect is unable to capture the variation in housing characteristics, which induces a negative deviation for the first purchase and a positive one for the second (considering that change in quality is included in the error term). Consequently, it fosters the relationships between individual deviation and realized capital gains. As robustness checks, we restrict observation with strong negative residuals.[1] The results remain consistent with our main results, which might indicate that renovation works are not likely to bias our substitution effect. Nonetheless, we consider our substitution measure as the upper bound.

5.2 Consequences on Wealth Accumulation

Our results support that individuals affect their wealth accumulation according to their performance in the housing market. Indeed, from an individual perspective, we demonstrate that positive deviation at the individual lower significantly the capital gains. Yet, when analysing the group-specific deviation, we remark heterogeneity according to the difference between deviation in capital gains and deviation in price difference. First-

[1] We remove first purchase residuals that are under respectively 20%, 15%, 10%, 5%.

time owners and small investors are those who can be considered the biggest losers in the housing market as they experience a positive price deviation and a negative capital gains deviation. On the opposite, big investors and rental investment companies generate sizeable capital gains, despite their positive price deviation. These results are consistent with previous results indicating higher anticipation levels for the wealthiest individuals (Martínez-Toledano, 2022), better financial literacy or bargaining intensity, although we do not disentangle the role of each channel.

Beyond the simple performance, wealth accumulation might be affected by the signalling effect of these price deviations. Indeed, a concentration of local positive price deviation might raise price expectations for future purchasers and sellers locally, thus strengthening local price dynamics as outlined by Aiello, Kotter and Schubert (2022). Hence, in addition to the direct effect on realized capital gains, it likely affects the realized capital gains of local homeowners as well. We expect a concentration of positive to fuel housing price dynamics, making the homeownership transition more difficult for the next generations or exacerbating spatial sorting. The consequences on the affordable housing supply must be considered in future research, as it is one of the main drivers of inequality between homeowners and renters.

Finally, we omit our analysis of the leveraging effect of interest rates. A positive price deviation will also affect households' budgets through an increase of interest associated with the mortgage. Assuming that most housing purchases are financed through a mortgage, the differences in price deviations will exacerbate overall purchase cost between categories and then wealth accumulation. Under the assumption that all housing units are financed through purchase at 80% to comply with the French rule of 20% down payment constraint, the price deviation between first-time owners and homeowners achieving residential mobility raises by €100 monthly payments for the first category in a 2% interest rates context. In addition, the assumption that all categories use mortgages homogeneously is likely to be imprecise. We might expect homeowners to achieve residential mobility and consumer investors to finance their purchases on their own, using simultaneous sell as a financial resource. Consequently, we provide a lower bound of the impact of price deviation on wealth accumulation as we omit the indirect consequences on disposable income.

5.3 Sorting Into Non-Observable Variables

Finally, we provide insightful results about sorting into non-observable between purchaser categories. We infer sorting into non-observable based on differences between the model with housing fixed effects (column 6, Table 1) and 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 the 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 often considered winners or losers in the housing market. Yet, we assess how price differences result from purchaser behaviour. We exploit a new category definition that is based on portfolio size before the purchase (to account for experience) and delta in property owned (to account for financial constraints). Considering significant differences in both observable and non-observable characteristics, we control for heterogeneity in housing choices by introducing housing fixed effects. This approach improves significantly explained variance and affects price differences. Yet, it requires restricting transaction samples to housing being sold at least twice between 2012 and 2021, which introduces potential selection bias. We then estimate the probability of each housing to be sold at least twice, and correct in the price regression using the Heckman procedure.

We demonstrate that the substitution between price deviation and realized capital gains is imperfect, as only €0.51 are transmitted on average. To explain this difference, we expect

the price deviation to affect the price beliefs of current homeowners, future purchasers and real estate agents. Hence, positive price deviations are likely to fuel the local housing market dynamics, leading to a mitigation of the impact of these deviations on realized capital gains. In addition, we highlight that the less experienced purchasers experience the highest price deviation. Finally, we provide insights that the most financially constrained purchasers (the first-time owners) choose a housing unit with a low level of quality for unobservable characteristics. We expect characteristics such as thermal insulation or construction quality to be the main driver.

Further research might focus on third specific questions. First, although we provide some insights that support the underlying mechanism we assume, uncovering how the capital gains difference is mitigated over time is crucial. Whereas it relates to wealth inequality understanding, it might also provide new insights into the housing affordability crisis. Second, understanding the redistributive effect of mispricing in the housing market is of interest. Whereas it might mitigate the wealth inequality if the poorest homeowners benefit from these price deviations, it might also raise difficulty in transitioning to homeownership. Third, while we highlight that first-time owners are likely to trade off housing quality, it might generate a loss for society when lowering gas emissions. Indeed, if financially constrained households live in housing units with low quality, especially with thermal insulation, it represents an important friction for reducing our gas emissions as a society.

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.

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

Ben-Shahar, Danny and Roni Golan (2019), 'Improved Information Shock and Price Dispersion: A Natural Experiment in the Housing Market'. *Journal of Urban Economics* 112, pp. 70–84.

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.

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.

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.

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.

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 Purchase Price Evolution per Category

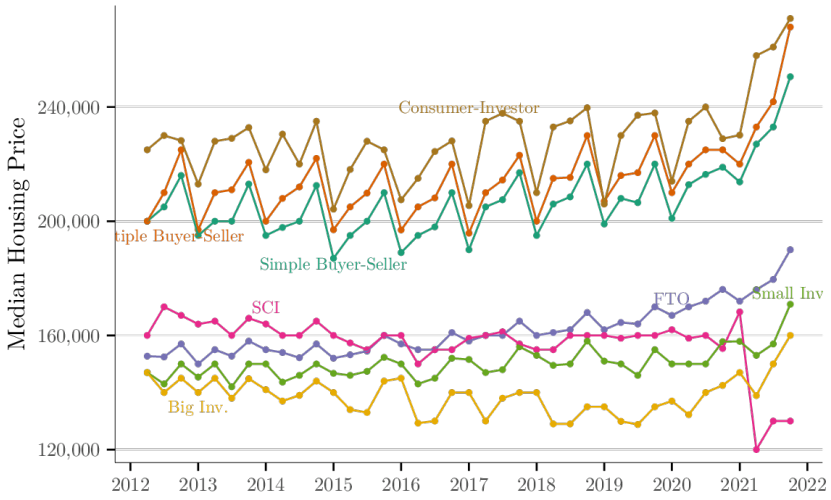


Figure A.1: Median Price per Purchaser Category

Notes: For each purchaser category, we represent the median price of housing transaction per quarter. The median housing price is expressed in euros.
Sources: Authors' Calculation from DV3F and Property Tax Files.

A.2 Relative Importance of Each Category on the Transaction Sample

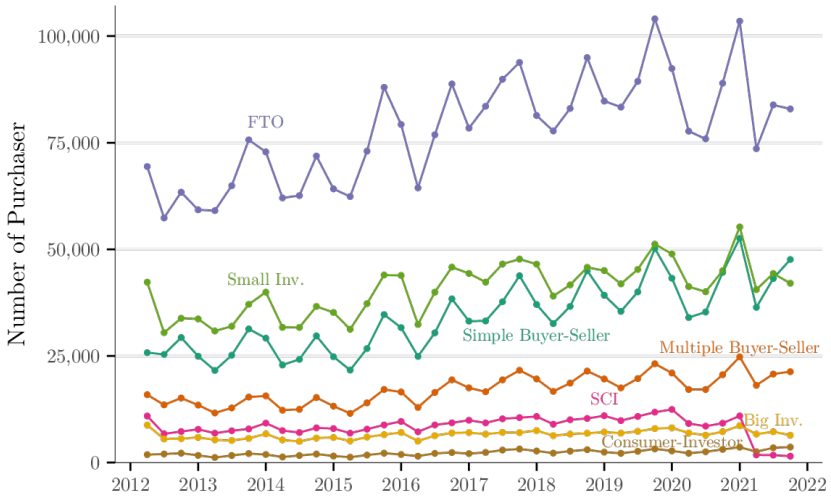


Figure A.2: Number of Purchaser per Quarter in France

Notes: For each purchaser category, we represent the number of housing transaction per quarter.
Sources: Authors' Calculation from DV3F and Property Tax Files.

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

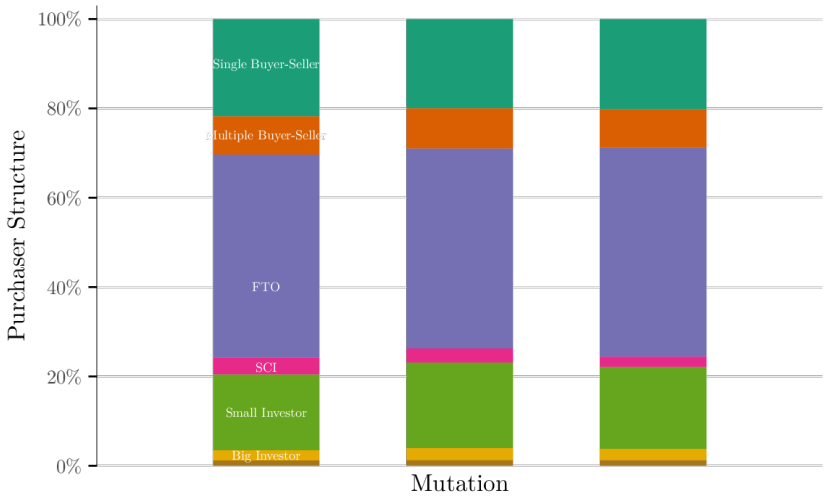


Figure A.3: Purchaser Structure for Each Mutation Number for Repeat Sales

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 Number of Housing Units According to the Number of Sales

Table A.4: Number of Housing Units According to the Number of Sales

Number of Sales	Number of Housing	Share of Transactions
1	5,766,392	88.21%
2	726,012	11.11%
3	44,748	0.68%

Notes: We report the number (and the share) of housing units being sold according to the number sales between 2012 and 2018.
Sources: Authors' Calculation from DV3F and Property Tax Files.

A.5 Distribution of Duration Between Two Sales

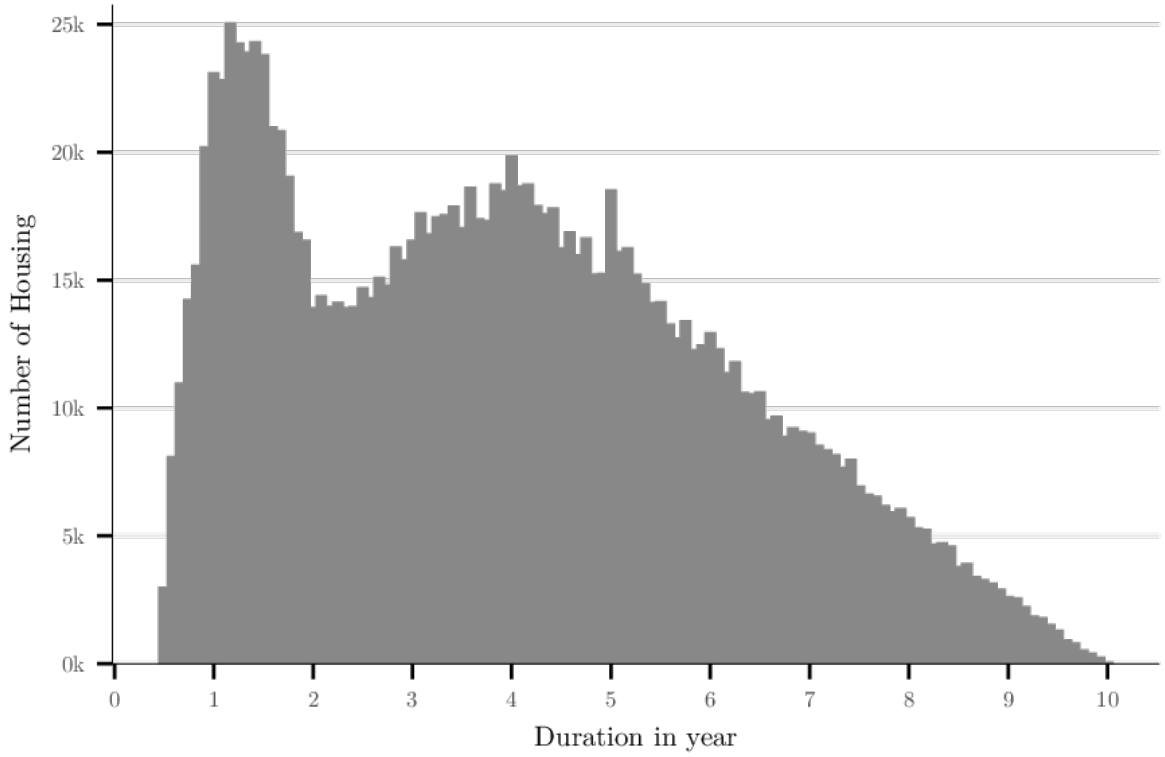


Figure A.5: Distribution of Duration between Two Sales

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.6 Distribution of Relative Capital Gains

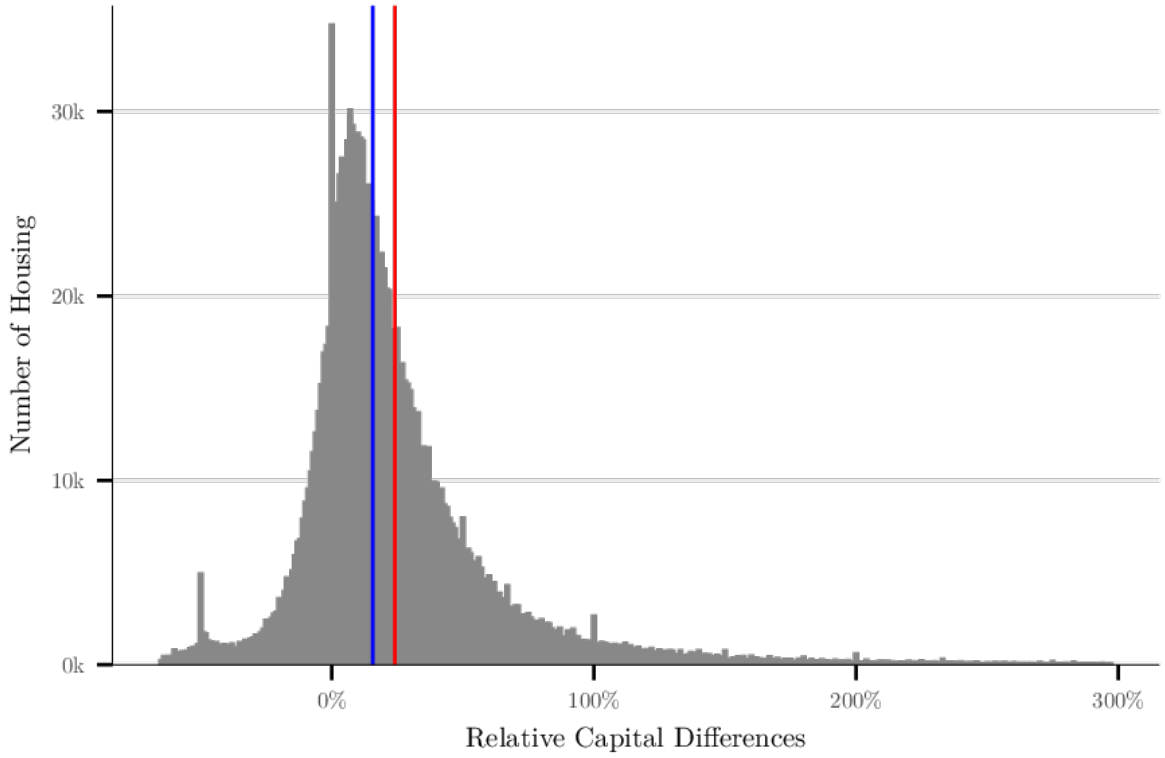


Figure A.6: Distribution of Realized Capital Gains

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.7 Descriptive Statistics about Full Transaction Sample

Table A.7: Descriptive Statistics about Purchase by Category

Category	N	Mean	Std. Dev.	Median	Q1	Q3
			A - Purchase Price			
Simple Purchaser-Seller	1,335,056	236,230	164,400	200,000	134,900	290,500
Multi Purchaser-Seller	663,718	262,780	213,730	208,060	129,000	329,080
FTO	3,041,768	179,410	125,550	152,500	105,000	218,200
Companies	316,691	224,980	240,980	155,500	90,000	269,000
Small Inv.	1,572,677	183,700	155,200	147,000	87,000	228,450
Big Inv	250,639	184,670	181,570	135,000	76,000	227,000
Consumer-Investor	89,228	277,190	221,820	220,000	140,000	342,330
			B - Housing Size			
Simple Purchaser-Seller	1,335,056	95.9	40.0	90.0	70.0	115.0
Multi Purchaser-Seller	663,718	91.0	48.0	84.0	59.0	115.0
FTO	3,041,768	80.5	35.8	76.0	57.0	98.0
Companies	316,691	79.4	50.8	68.0	45.0	99.0
Small Inv.	1,572,677	71.7	41.6	64.0	41.0	91.0
Big Inv	250,639	67.6	44.3	59.0	36.0	85.0
Consumer-Investor	89,228	97.9	48.1	90.0	66.0	120.0
			C - Price per square metres			
Simple Purchaser-Seller	1,335,056	2,615	1,648	2,216	1,536	3,209
Multi Purchaser-Seller	663,718	3,083	1,974	2,600	1,719	3,855
FTO	3,041,768	2,481	1,689	2,026	1,372	3,055
Companies	316,691	3,058	2,226	2,420	1,456	3,929
Small Inv.	1,572,677	2,840	1,884	2,396	1,500	3,641
Big Inv	250,639	2,994	2,079	2,500	1,480	3,866
Consumer-Investor	89,228	3,002	1,966	2,489	1,667	3,718

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

A.8 Descriptive Statistics about Repeat Sales Sample

Table A.8: Descriptive Statistics about Purchase by Category

Category	N	Mean	Std. Dev.	Median	Q1	Q3
			A - Purchase Price			
Simple Purchaser-Seller	289,268	224,300	151,040	190,000	130,000	274,550
Multi Purchaser-Seller	119,899	250,930	199,300	200,000	125,880	312,000
FTO	625,946	183,680	117,950	158,800	113,000	221,000
Companies	45,330	219,690	226,240	156,000	94,000	260,000
Small Inv.	248,389	185,700	151,910	148,000	90,000	231,300
Big Inv	33,455	187,380	179,720	135,740	79,000	232,000
Consumer-Investor	16,728	264,620	203,550	212,800	140,000	325,100
			B - Housing Size			
Simple Purchaser-Seller	289,268	90.1	37.8	85.0	65.0	108.0
Multi Purchaser-Seller	119,899	85.3	45.0	80.0	54.0	107.0
FTO	625,946	76.3	32.9	72.0	55.0	92.0
Companies	45,330	73.9	48.2	64.0	40.0	92.0
Small Inv.	248,389	69.0	39.8	63.0	39.0	89.0
Big Inv	33,455	64.5	42.4	55.0	33.0	83.0
Consumer-Investor	16,728	92.1	44.4	85.0	63.0	114.0
			C - Price per square metres			
Simple Purchaser-Seller	289,268	2,672	1,688	2,250	1,578	3,236
Multi Purchaser-Seller	119,899	3,167	2,025	2,631	1,783	3,905
FTO	625,946	2,690	1,785	2,200	1,533	3,225
Companies	45,330	3,231	2,269	2,571	1,617	4,071
Small Inv.	248,389	2,976	1,969	2,451	1,633	3,671
Big Inv	33,455	3,171	2,178	2,585	1,625	4,000
Consumer-Investor	16,728	3,082	2,002	2,530	1,741	3,776

Notes: We report descriptive statistics about purchase price (top panel), housing size (middle panel) and price per square metres (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.9 Spatial Distribution of Tenure Status

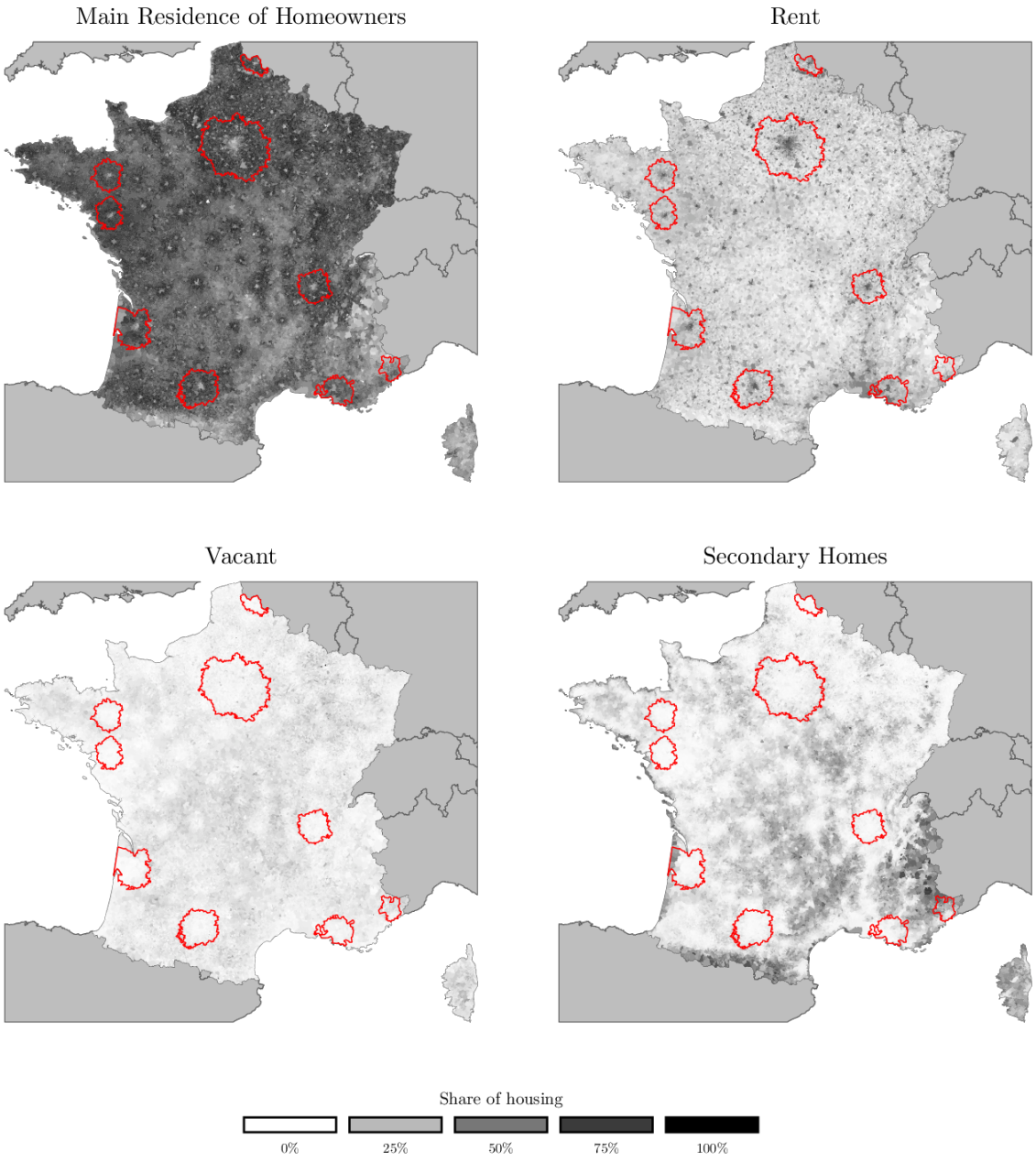


Figure A.9: Share of Housing per Tenure Status (2018)

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.10 Spatial Dispersion for Alternative Categories

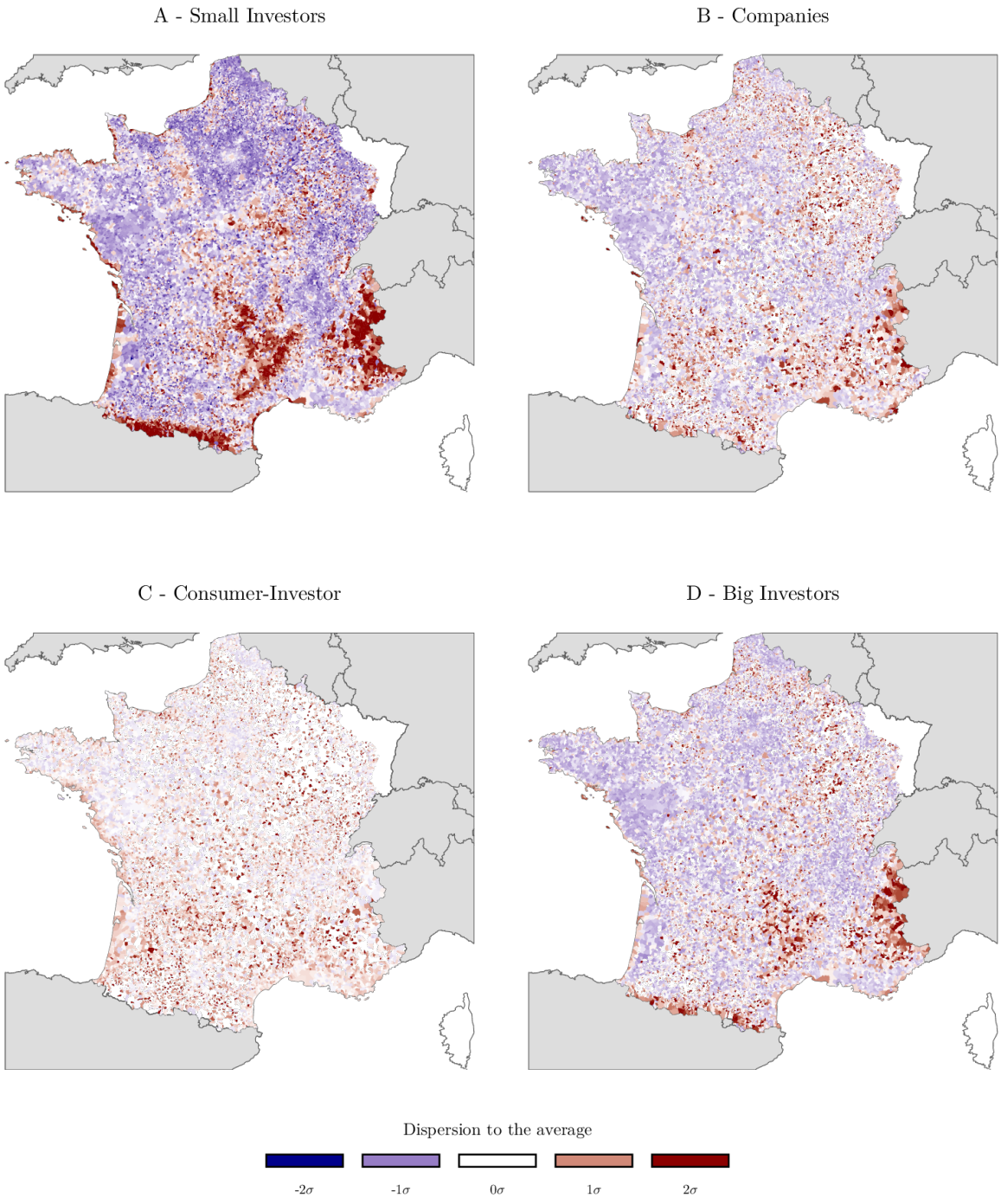


Figure A.10: Spatial Dispersion for Alternative Categories

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: Overlap

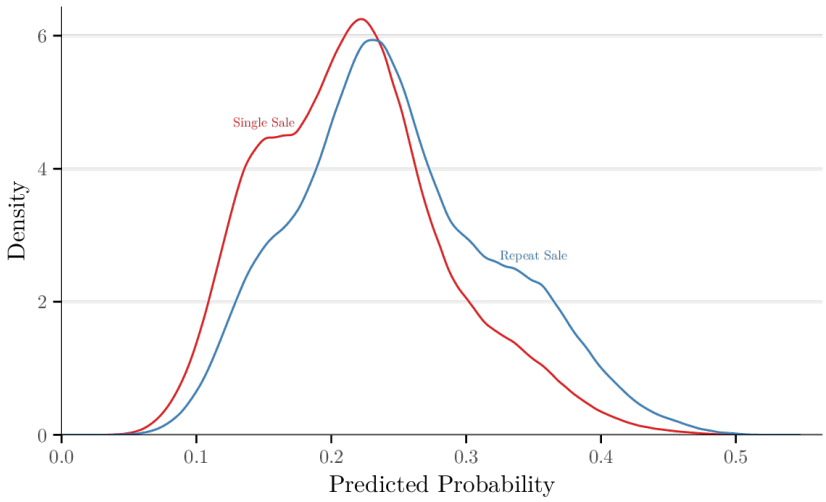


Figure B.1: Probability Distribution According to Sample of Interest

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 (1).
Sources: Authors' Calculation from DV3F and Property Tax Files.

B.2 Second Step Regression: Control Variables

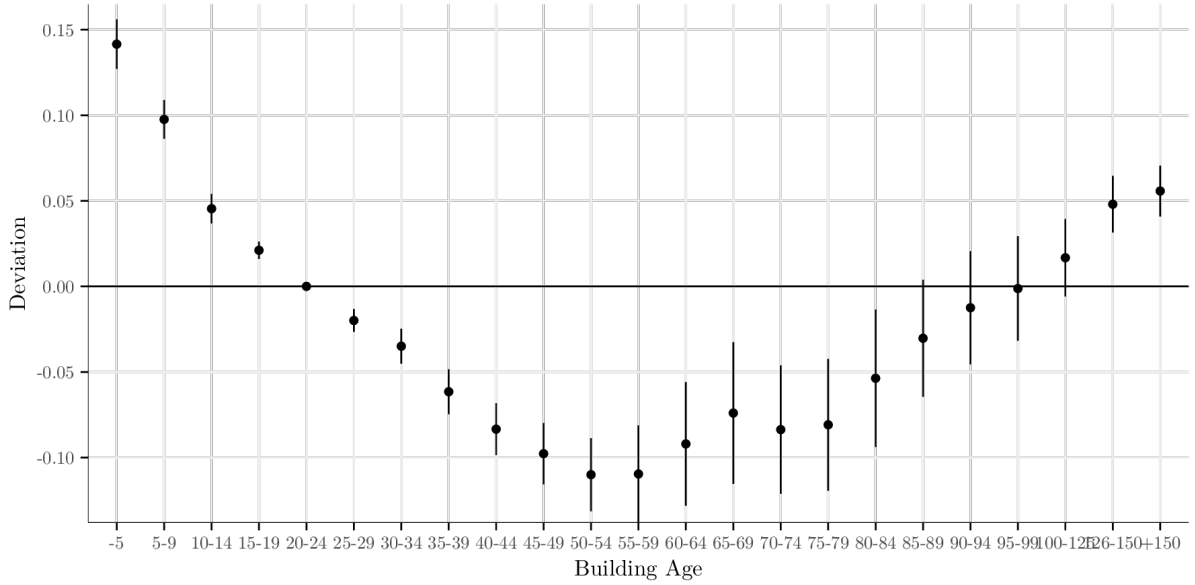


Figure B.2: WLS Results for Building Age Category

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.3 Second Step Regression: Control Variables

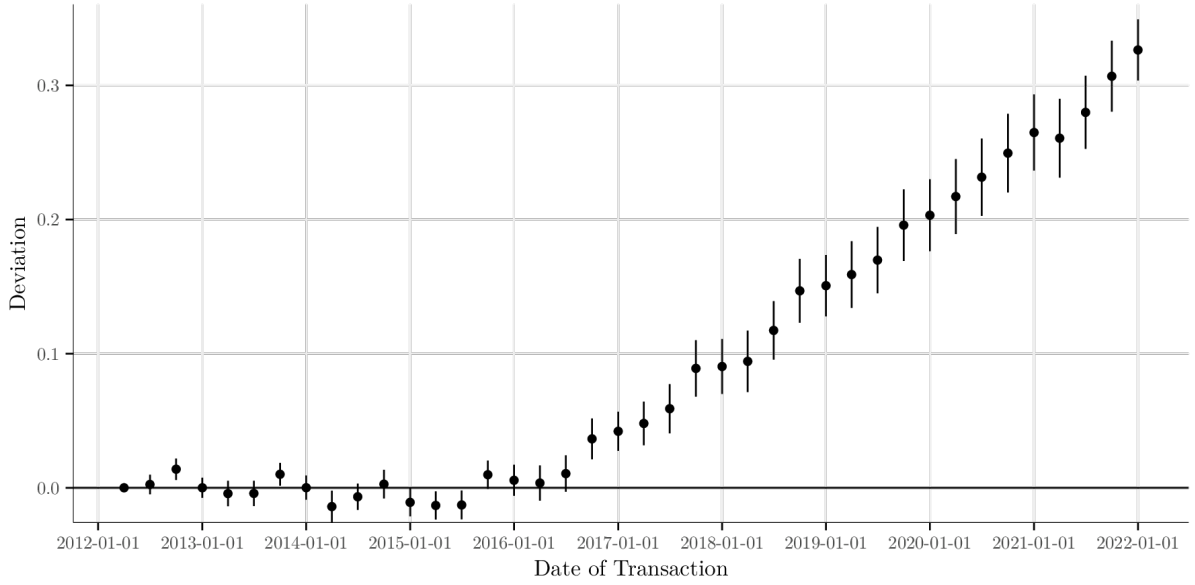


Figure B.3: WLS Results for Date

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: Alternative Results with Distance Control

Table B.4: WLS Results for Purchase Price Differences with Distance Controls

	Dependent variable: Purchase price (log)				
Covariate Category (ref. Simple Purchaser-Seller)	(1)	(2)	(3)	(4)	(5)
Multi Purchaser-Seller	–	–	–	0.015***	0.009***
				(0.001)	(0.001)
FTO	–	–	–	0.024***	0.034***
				(0.002)	(0.001)
Companies	–	–	–	-0.008***	0.010***
				(0.002)	(0.002)
Small Inv.	–	–	–	0.030***	0.027***
				(0.002)	(0.001)
Big Inv	–	–	–	0.016***	0.009***
				(0.004)	(0.002)
Consumer-Investor	–	–	–	0.023***	0.014***
				(0.003)	(0.002)
Tenure status: Own-occupied (ref.)					
Rent	–	-0.407***	–	–	-0.032***
		(0.017)			(0.002)
Vacant	–	-0.275***	–	–	-0.037***
		(0.012)			(0.002)
Other	–	-0.189***	–	–	-0.012***
		(0.035)			(0.003)
Distance: <5km (ref.)					
5km-20k	-0.073***	-0.062***	0.018***	0.017***	0.016***
	(0.012)	(0.009)	(0.002)	(0.002)	(0.001)
21km-50km	-0.163***	-0.137***	0.028***	0.027***	0.022***
	(0.025)	(0.023)	(0.002)	(0.002)	(0.001)
51km-200km	-0.238***	-0.199***	0.037***	0.036***	0.029***
	(0.030)	(0.028)	(0.002)	(0.002)	(0.001)
>200km	-0.036	-0.018	0.039***	0.039***	0.038***
	(0.022)	(0.024)	(0.003)	(0.003)	(0.002)
N	2,047,504	2,047,504	2,047,504	2,047,504	2,047,504
R^2	0.013	0.130	0.938	0.938	0.959
Adj. R^2	0.013	0.130	0.910	0.910	0.940
Within R^2	–	–	0.004	0.006	0.339
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 (2) 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.5 Robustness Checks: Duration Approach

Table B.5: 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	-0.006**	0.005**	0.009***	0.010***
	(0.002)	(0.002)	(0.002)	(0.001)
FTO	0.043***	0.049***	0.047***	0.043***
	(0.002)	(0.002)	(0.002)	(0.002)
Companies	-0.009	0.007*	0.015***	0.018***
	(0.007)	(0.004)	(0.003)	(0.003)
Small Inv.	0.024***	0.035***	0.037***	0.037***
	(0.002)	(0.002)	(0.001)	(0.001)
Big Inv	-0.005	0.009*	0.016***	0.016***
	(0.006)	(0.005)	(0.004)	(0.003)
Consumer-Investor	0.000	0.007*	0.013***	0.012***
	(0.005)	(0.004)	(0.003)	(0.003)
N	333,465	623,480	991,211	1,366,190
R^2	0.956	0.958	0.959	0.959
Adj. R^2	0.944	0.943	0.943	0.943
Within R^2	0.008	0.010	0.009	0.008

Notes: We report results for WLS estimation of Equation (2) 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.6 Robustness Checks: Price Variation Approach

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

	Dependent Variable: Price (log transformation)			
Covariate	2.5 %	5 %	10 %	15 %
Category (ref. Simple Buyer–Seller)				
Multi Purchaser-Seller	0.004***	0.006***	0.010***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)
FTO	0.031***	0.023***	0.021***	0.024***
	(0.002)	(0.002)	(0.001)	(0.001)
Companies	0.033***	0.027***	0.021***	0.019***
	(0.002)	(0.002)	(0.002)	(0.002)
Small Inv.	0.026***	0.021***	0.021***	0.023***
	(0.002)	(0.002)	(0.001)	(0.001)
Big Inv	0.019***	0.015***	0.013***	0.013***
	(0.003)	(0.002)	(0.003)	(0.002)
Consumer-Investor	0.011***	0.011***	0.014***	0.016***
	(0.003)	(0.002)	(0.002)	(0.002)
N	1,048,852	1,571,390	2,017,581	2,159,256
R^2	0.976	0.977	0.973	0.969
Adj. R^2	0.966	0.967	0.962	0.956
Within R^2	0.008	0.005	0.003	0.003

Notes: We report results for WLS estimation of Equation (2) 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.7 Capital Gains: Results with Level Difference as Alternative Outcome

Table B.7: 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	6,919.0**	8,179.7***	7,507.6***	6,896.3***	1,314.9***	–	-1,085.6**
	(2,726.8)	(2,122.9)	(1,890.1)	(1,683.7)	(396.3)		(480.1)
FTO	-5,830.1***	-3,928.2***	-6,871.0***	-6,891.4***	-3,477.6***	–	-7,215.9***
	(833.3)	(828.9)	(1,429.1)	(1,418.7)	(687.4)		(806.6)
Companies	3,567.4	8,475.9**	8,063.5**	6,272.1**	-1,262.4	–	-6,862.0***
	(3,947.8)	(3,771.1)	(3,569.0)	(2,916.8)	(1,148.3)		(707.0)
Small Inv.	-6,373.2***	-742.3	-2,762.9***	-3,739.0***	-2,444.4***	–	-6,730.2***
	(925.7)	(589.6)	(592.2)	(802.1)	(425.2)		(238.3)
Big Inv	-2,240.5*	4,824.3***	3,759.9***	2,195.1**	-77.8	–	-5,592.8***
	(1,250.0)	(757.5)	(704.7)	(831.0)	(579.4)		(585.5)
Consumer-Investor	10,074.8**	9,499.9***	9,194.0**	8,599.6***	2,390.9	–	563.3
	(3,984.8)	(3,498.0)	(3,400.3)	(3,133.2)	(1,631.5)		(1,259.5)
Individual Deviation	–	–	–	–	–	-2,853.4***	-2,863.2***
						(216.0)	(216.7)
N	1,378,158	1,378,158	1,378,158	1,378,158	1,378,158	1,378,158	1,378,158
R^2	0.004	0.032	0.075	0.080	0.294	0.535	0.537
Adj. R^2	0.004	0.032	0.075	0.080	0.280	0.526	0.527
Within R^2	–	–	0.032	0.038	0.094	0.404	0.406
Housing Control	No	Yes	Yes	Yes	Yes	Yes	Yes
Time Control	No	No	Yes	Yes	Yes	Yes	Yes
Muni FE	No	No	No	No	Yes	Yes	Yes
Heckman Correction	No	No	No	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 (2)). 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.8 Capital Gains: Results with Future Price as Alternative Outcome

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

	Future Selling Price						
Covariate	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Multi Purchaser-Seller	0.034	0.088***	0.083***	0.091***	0.020***	–	0.014***
	(0.030)	(0.014)	(0.014)	(0.012)	(0.002)		(0.002)
FTO	-0.163***	-0.092***	-0.109***	-0.109***	-0.040***	–	-0.050***
	(0.015)	(0.008)	(0.008)	(0.008)	(0.002)		(0.003)
Companies	-0.078*	0.087***	0.083***	0.107***	0.013**	–	-0.002
	(0.045)	(0.029)	(0.029)	(0.023)	(0.005)		(0.005)
Small Inv.	-0.272***	-0.090***	-0.103***	-0.090***	-0.035***	–	-0.046***
	(0.024)	(0.010)	(0.010)	(0.014)	(0.003)		(0.003)
Big Inv	-0.298***	-0.069***	-0.078***	-0.056***	-0.038***	–	-0.052***
	(0.037)	(0.011)	(0.011)	(0.011)	(0.004)		(0.005)
Consumer-Investor	0.114***	0.108***	0.104***	0.112***	0.033***	–	0.028***
	(0.026)	(0.020)	(0.021)	(0.019)	(0.004)		(0.004)
Individual Deviation	–	–	–	–	–	-0.007***	-0.008***
						(0.000)	(0.000)
N	1,378,158	1,378,158	1,378,158	1,378,158	1,378,158	1,378,158	1,378,158
R^2	0.025	0.237	0.250	0.261	0.810	0.827	0.829
Adj. R^2	0.025	0.237	0.250	0.261	0.806	0.824	0.825
Within R^2	–	–	0.237	0.249	0.660	0.691	0.694
Housing Control	No	Yes	Yes	Yes	Yes	Yes	Yes
Time Control	No	No	Yes	Yes	Yes	Yes	Yes
Muni FE	No	No	No	No	Yes	Yes	Yes
Heckman Correction	No	No	No	Yes	Yes	Yes	Yes

Notes: We report results for WLS estimation of Equation (3) with selling price 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 (2)). 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

	Difference						
Covariate	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Multi Purchaser-Seller	119	15	-1,048	-1,589**	-3,011***	–	-3,939***
	(1,239)	(1,057)	(837)	(733)	(353)		(358)
FTO	-2,527***	-3,180***	-6,579***	-6,702***	-5,286***	–	-6,720***
	(887)	(1,007)	(1,438)	(1,459)	(836)		(854)
Companies	-12,109***	-10,414***	-11,803***	-13,320***	-15,104***	–	-14,000***
	(1,226)	(1,321)	(1,142)	(1,266)	(1,589)		(1,480)
Small Inv.	-5,553***	-5,057***	-7,554***	-8,496***	-7,186***	–	-8,721***
	(871)	(704)	(961)	(1,169)	(748)		(724)
Big Inv	-5,678***	-4,794***	-6,366***	-7,710***	-8,009***	–	-8,457***
	(596)	(641)	(752)	(1,073)	(1,468)		(1,252)
Consumer-Investor	-3,135**	-3,091**	-3,681***	-4,155***	-5,430***	–	-5,496***
	(1,477)	(1,279)	(1,088)	(1,007)	(699)		(616)
Individual Deviation	–	–	–	–	–	-2,800***	-2,805***
						(214)	(214)
N	656,507	656,507	656,507	656,507	656,507	656,507	656,507
R^2	0.004	0.018	0.113	0.122	0.240	0.482	0.488
Adj. R^2	0.004	0.018	0.113	0.122	0.211	0.462	0.468
Within R^2	–	–	0.021	0.031	0.032	0.340	0.347
Housing Control	No	Yes	Yes	Yes	Yes	Yes	Yes
Time Control	No	No	Yes	Yes	Yes	Yes	Yes
Muni FE	No	No	No	No	Yes	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 (2)). 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.
*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$