Spatial Sorting and Housing Wealth Persistence: Insights from the Universe of French Homeowners

Rémi Lei[*,†]
remi.lei@univ-avignon.fr

Jean-Sauveur Ay[†]
jean-sauveur.ay@inrae.fr

Olivier Dupré[‡]
olivier.dupre@cerema.fr

Julie Le Gallo[†]
julie.le-gallo@agrosupdijon.fr

January 31, 2024

Abstract

This article studies the relationship between initial levels and subsequent changes in gross housing wealth. Using administrative microdata on housing ownership and transactions in France over the period 2011-2019, we impute a market value for all housing portfolios held by private homeowners. We document strong housing wealth persistence (HWP) for this population, particularly with respect to the location of the main residences. We then examine HWP for three sub-populations identified from the panel structure of our data. HWP appears to be mainly driven by homeowners who change their main residence or receive a housing inheritance, while it is less pronounced for homeowners with a constant housing portfolio. The spatial sorting operated by the location choices of main residences (both between and within commuting zones) appears to be a key determinant of HWP, rather than capital gains from price variations of given housing portfolios.

JEL classification: R31 ; R12 ; D31 ; C21

Keywords: Private ownership ; individual data ; housing portfolios ; capital gains ; housing transfers ; spatial inequality.

*ESPACE UMR 7300, Université d'Avignon, F-84000 Avignon
†CESAER UMR 1041, INRAE, Institut Agro, F-21000 Dijon.
‡CEREMA Hauts-de-France, F-59000 Lille

1 Introduction

Since it was recognized as the highest source of monetary inequality in rich countries, wealth and its distribution has received large attention since the papers by (Piketty and Zucman, 2014; Benhabib et al., 2017). Because housing is the most widespread asset in individual portfolios (Jordà et al., 2019), housing wealth persistence (HWP) has been given a particular focus in subsequent studies (Saez and Zucman, 2016; Alvaredo et al., 2018; Blanco et al., 2021; Garbinti et al., 2021).

From a financial perspective, HWP may result from increasing returns to scale of housing investments (Bach et al., 2020) and, consequently, higher returns for better endowed homeowners (Fagereng et al., 2020). In addition, access to homeownership is a significant determinant for HWP (Pfeffer and Waitkus, 2021), stemming both from monetary (Gabriel and Painter, 2020) and non-monetary transfers such as inheritances (Boserup et al., 2016). These differences of initial endowments would persist and condition inequality over the life-cycle (Huggett et al., 2011).

However, housing is also an illiquid durable consumption good (Grossman and Laroque, 1987) used by most individuals when making their location choices (Ortalo-Magné and Prat, 2016). Indeed, the location of the main residence determines the access to local amenities (Roback, 1982) and more generally, life perspectives (Bilal and Rossi-Hansberg, 2021) that in turn are capitalized in housing prices (Glaeser et al., 2005). As a spatially-fixed asset, cross-sectional variations of housing prices then induce homeowners to sort themselves across neighbourhoods (Kuminoff et al., 2013).

In this article, we provide first insights on HWP in France, by studying the relationship between initial levels and subsequent changes of gross housing wealth. For the 2011–2019 period, we leverage a new data-set about the population of private homeowners in order to trace back both the initial levels and the variations of gross housing wealth. Wealth includes rented accommodations, second homes, and inherited dwellings in addition to main residences. By estimating partial correlations between cross-sectional and longitudinal wealth variations, we document the interactions between the financial and the consumption sides of housing wealth that, taken together, are first-order drivers of HWP.

In analysing this relationship, we particularly consider the spatial dimension of HWP. Indeed, it is well known that housing derives most of its value from its location at different spatial scale (Kiel and Zabel, 2008). Considering jointly that rent-to-price ratio has decreased in land-constrained areas (Hilber and Mense, 2021) and that housing values diverge from income (Albouy et al., 2016), the location of main residences might be increasingly driven by wealth over income. Because of the larger moving cost associated to ownership (Van Ommeren and Van Leuvensteijn, 2005), the spatial sorting of households in their location choices could amplify HWP through spatially differentiated dynamics of housing prices. For instance, as wealthier individuals are more likely to invest in local public goods (Hilber and Mayer, 2009), the attractive features of their neighbourhood are strengthened through prices capitalisation (see, e.g., the numerous works about the capitalization of local amenities that followed Black, 1999; Chay and Greenstone, 2005; Banzhaf and Farooque, 2013; Diao et al., 2017). Finally, spatial sorting across neighbourhoods would produce externalities that strengthen the housing price dynamics of the most attractive locations (Guerrieri et al., 2013).

Despite the potential high effect of spatial sorting on HWP, few empirical papers simultaneously consider both processes. Using both spatial equilibrium and asset pricing models, Ortalo-Magné and Prat (2016) aim to provide "a first step" in this direction and derive some important propositions. Firstly, the decreasing marginal utility of consumption, a determinant mechanism in spatial sorting models, is not sufficient to explain location decisions when investment in housing is considered. Indeed, as tenure choices derive partly from the balance between risks associated with expected rent variations and potential capital gains (Sinai and Souleles, 2005), portfolio considerations from the whole individual wealth also affect location choices. Secondly, spatial sorting affects overall asset pricing (including bounds stocks) through differentiated investment capabilities related to housing wealth. Thirdly, transaction and moving costs do not affect the preference for local investment.

Nonetheless, the lack of individual data about housing wealth limits the empirical approach despite recent development (Eggum and Larsen, 2021). The administrative data used in this paper come from French fiscal sources to construct a panel of the universe of French homeowners with both the locations of their main residences and their whole housing portfolio. Firstly, using property tax records (*Fichiers Fonciers*), we identify unique private homeowners and their detailed housing portfolio. Note that, independently of our work, André and Meslin (2021) also use similar data sources to provide the cross-sectional distribution of housing wealth between private owners. Our cross-sectional results for the year 2017 are consistent with theirs. Secondly, we appraise housing wealth to owners using housing value imputation model making the best of exhaustive geocoded transaction datasets (DV3F). Our method captures spatial heterogeneity both in cross-section and longitudinal way using bivariate smoothing splines in interaction with time (Wood, 2017). Thirdly, we improve current transaction datasets derived from tax sources by identifying changes in ownership due to inheritance and first-time ownership.

The definition of the spatial dimension to provide empirical insights on the HWP is based on the commuting zone. Commuting zones are commonly assumed to be a consistent spatial unit for local housing markets as decision for residence places and job opportunities are closely related (Zabel, 2012). To draw partial correlation of interest, we choose to rank commuting zone according to a continuous variable over the introduction of fixed effects. Our choice is motivated by the tractability it provides, especially for interactions specification. We therefore use built-up area in 2000 as a proxy for the economic status of commuting areas. The underlying assumption is that built-up area is correlated with the level of amenities and productivity (Saiz, 2010, Proposition 2). In addition, we treat heterogeneity within commuting zones using distance from the Centre Business District (CBD), following the usual monocentric approach (Duranton and Puga, 2015). Remark that the simplicity of the spatial dimensions enables to interact both, without loss of generality. Empirically, we define exogeneously commuting zones and their CBD using commuting zones supplied by the INSEE.[1]

[1] French National Institute of Economic Studies and Statistics

Our contribution concerns both the data we bring, and the empirical results we derive from it. The new exhaustive dataset on homeowners and their entire housing portfolios for the period 2011-2019 in France is a major contribution, and is likely to feed further empirical work. From their use, we provide first insights based on partial correlations about the HWP through the spatial and individual dimensions. Our four empirical findings concern four different populations that suggest potential drivers of HWP. First, wealth accumulation that stem from the housing market dynamics is correlated with the initial level of wealth at the individual level and, more importantly, with the location of the main residence. Second, HWP is likely to result from housing choices for the main residence. Both transaction price and spatial sorting are correlated with the initial level of housing wealth. Third, housing transfers mainly benefit the wealthiest homeowners, as the share received increases with the initial level of wealth. Moreover, the value of the transfer increases with the economic status of the commuting zone in which the recipient lives. Fourth, the ownership of long-distance rental investments mitigate the HWP through the spatial dimension.

Our paper is structured as follows. We detail the administrative data sources we gather to build exhaustive panel data about owners over the 2011–2019 period (Section 2). It includes methodological choices associated to the main extension we make. Then, we provide our empirical analysis in Section 3. Conclusions and possible extensions are presented in Section 4.

2 Data Processing

We first present the administrative data sources we use, and how we construct panel data about owners and housing properties over the 2011–2019 period. In addition, we detail our housing market imputation (Section 2.2), which is required to assess both initial levels and trends for the HWP, the identification of transfers (Section 3.4) and the identification of first-time owners (Section 2.4). Our results in the cross-sectional dimension are close to those obtained by André and Meslin (2021).

2.1 Raw Administrative Sources

Our database exploits three raw data sources i) a housing stock dataset with matched owners ii) a housing transaction dataset iii) commuting zone perimeter. Except from the commuting zone perimeter, all data sources are derived from fiscal sources.

Housing Stock Database This database, entitled *Fichiers Fonciers* provided by the CEREMA,[2] contains information about housing stock using to property tax collection. For each January 01^{st}, we observe the housing stock in France and detailed observations about current owners. Data sources are exhaustive about both owners and housing properties. Private owners are defined according to their civil state, with the date of birth, gender and current address, while legal persons are classified according to the structure type (public, private, social landlords) and identified with national ID. Furthermore, each housing is identified by a unique national ID and its structural features, including all relevant characteristics for tax assessment such as surface, housing type, building years or presence of particular facilities (e.g. swimming pool, cellar, parking lot). We expect these characteristics to be precisely reported as they condition the property tax value. In addition, we observe the housing location using the centroid of the parcel where the housing belongs. As parcel sizes are tiny,[3] observations are precisely located. Finally, a property right table is available to assign properties to owners each year. Each property right is described according to its type, following the French law definition (full owners, usufruct, bare owner). We provide detailed statistics about housing transaction dataset in Appendix A.1 including a description of the housing stock through detailed housing characteristics and spatial distribution.

Housing Transaction Dataset We also access administrative files registering all housing transactions in France since 2010 (DV3F).[4] Housing is defined with similar variables as

[2] *Centre d'Expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement.*

[3] We estimate that the average size for a parcel in France about 0.57 hectares. It represents the area covered by a 75-metre square.

[4] Only the region Alsace-Moselle is not available due to historical reasons. These areas have an alternative system, named as *Livre Fonciers*, inherited from a German law in 1896. Recall that Alsace-Moselle

the housing stock database (as they both derive from fiscal sources), including structural characteristics, housing location and unique identifier. Furthermore, additional information with respect to transactions, such as dates, nature (sale, exchange, expropriation, etc.) and the purchase price are available. Nonetheless, we do not observe the financial conditions related to the operation such as equity and mortgage reimbursement conditions. Finally, the data source provides information on the type of seller and buyer according to whether they are private individuals or legal entities. Transaction datasets contain 9,158,323 transactions about 10,774,349 housings over the 2010–2020 period. We report additional descriptive statistics are available in Appendix A.2 about transaction price, housing characteristics and spatial distribution.

Commuting Zones We adopt commuting zone as stable unit for local housing market. We exploit the zoning supplied by the INSEE for 2010. Commuting zones are defined based on working and residence place for most inhabitants, leading to stable and consistent geographical units for both housing and labour markets. We adopt as the centre of the commuting zone, the chief town of the municipality with the highest density within the area. Instead of introducing commuting zone fixed effects, we classify the commuting zone using a continuous variable. This allows us to interact both dimensions of spatial heterogeneity (within and between commuting zones) with tractable and transparent results that do not result from the fixed effects approach. While the fixed effects approach is more precise, it requires a univariate analysis that does not allow for interactions. Nevertheless, we expect heterogeneity within commuting zones to vary with their characteristics. Our continuous classification for commuting zone proxy for economic status with the built-up area in 2000, in line with Saiz (2010). We assume that top commuting zones in terms of economic status have high built-up areas. Empirically, we consider as built-up parcels, land with at least one construction built prior to 2000 using the housing stock database. We report additional statistics about the spatial distribution of built-up area in 2000 in Appendix A.3.

belonged to Germany between 1870 and 1918.

Making the best of administrative data sources, we construct panel data about French homeowners over the 2011–2019 period. We introduce a new homeowner ID that is based on civil status rather than address or department. This allows the ID to be considered invariant over time, which is an appealing property for building panel data. Our ID construction take account for potential misspecification in civil state using names truncature.[5] Consequently, we overcome the main shortcoming of the current ID, which is based on French departmental boundaries.[6]. Then, the current ID is irrelevant to study housing wealth both in cross-section (e.g. homeowners can have assets in distinct departments) and longitudinal (e.g. individuals can change their residence place) dimensions.

Empirically, we retain as civil state variables birth name (being more stable over time), first name and day of birth. The improvement is sizeable: while we estimate the number of unique private owners to 41.6M over the 2011–2019 period based on the current ID, this estimation decreases to 30.7M using the unified ID.

2.2 Housing Value Imputation

Beyond the construction of panel data about private owners and their properties, we improve fiscal data through the imputation of housing market value to appraise owners' housing wealth. Although the number of properties is a good proxy for wealth, the valuation of housing allows the derivation of the gross housing wealth for owners and its evolution required for the HWP. Our imputation method to estimate unitary housing price for each observation account specifically for spatially heterogeneous trend of the housing market over the 2011–2019 period. These approaches are mainly used for tax property purposes as it aims to retrieve transparent and fair values for tax purpose that overcome main issues related to self-reported values (Tur-Sinai et al., 2020). In addition, researches have focus on appraisal models to understand the determinants of housing or rent prices especially for the spatial dimension (Ahlfeldt et al., 2023).

[5] We provide an error assessment using type I and type II distinction according to civil state truncating to find the option that fits the best.

[6] Departments are administrative boundaries for France introduced in 1789 following the *French Revolution* It corresponds to NUTS-3 level in Nomenclature of Territorial Units for Statistics.

The housing value imputation procedure must account for various sources of heterogeneity. Apart from housing characteristics, the location is determinants for multiple spatial scale (Lee and Myers, 2003; Kiel and Zabel, 2008). While controlling for housing characteristics such as surface or housing type is most of the time straightforward, spatial heterogeneity is more difficult to model. Our imputation model introduces spatial coordinates using bivariate smoothing splines, in line with the Generalised Additive Models framework (Wood, 2017), to capture spatial heterogeneity. It takes advantage of the continuity of spatial coordinates available in administrative sources, with data-driven definition. Indeed, the effective degree of freedom for the spatial contribution is endogenously defined to fit the best to the data. The intuition is simple, as models with higher degree of freedom are more likely to capture the local singularity, but decrease the precision.

Our imputation model has many advantages for an implementation at the country level compared to common approaches to mass appraisal, such as a spatial fixed effects approaches estimated by Ordinary Least Squares (OLS) or a Geographically Weighted Regression (GWR Brunsdon et al., 1996; Fotheringham et al., 2002). Firstly, it relaxes the need to find the most appropriate spatial unit for the fixed effects approach based on the trade-off between statistical power and granularity. Secondly, it relaxes the assumption that price is homogeneous within the spatial unit conditional on housing characteristics for fixed effects as well (McMillen, 2010). Thirdly, although recent developments have addressed computational cost issues (Li et al., 2019; Murakami et al., 2020; Li and Fotheringham, 2020), GWR remains difficult to implement on a national scale. Although it does not allow the marginal contribution of covariates to vary across space, we expect the spatial smoothing approach to be the best compromise between accuracy and tractability. Making an accuracy comparison on smaller dataset, we consider that the GAM approach is more accurate than the OLS, and performs at least as well as the GWR (for more details about the accuracy comparison procedure, see Appendix A.6).

In addition to spatial locations, we introduce variables to control for heterogeneity in housing characteristics. Nonetheless, our variable set is exogenously selected for computational reasons and consistency at a national level. Each housing transaction is defined based on its housing surface, housing type, dependence surface, and building age. We do not introduce additional variables such as the average surface per rooms to reduce computational costs. We introduce the spatial coordinates taking advantage of geocoded nature of the observations, using bivariate smoothing thin plates, in interaction with time dummies. Our objective is to capture both cross-sectional and dynamic heterogeneity. The introduction of additive smoothing splines aims to capture for different scale levels that determine housing prices Kiel and Zabel (2008). The mass appraisal model is reported in Equation 1.

$$y_{it} = \alpha + h(z_i, t) + \sum_{j=1}^{J} f_j(x_{it}) + \mathbf{X}\beta + \varepsilon_{it} \tag{1}$$

where y_{it} is the outcome for observation i at time t (unitary price); z_i are the spatial coordinates for the location of housing i; h represents the spatial smoothing function; x_{it} is the $i-th$ selected variable in our mass appraisal; f_j represents transformations functions using additive splines; and ε_{it} is the idiosyncratic error term. Empirically, we define f_j and h as additive variable transformation. Variance of errors is minimised by penalised least squares, while smoothing parameters are estimated using restricted maximum likelihood approach. Using the estimation of α, f_j, h and β (respectively noted $\widehat{\alpha}$, $\widehat{f}_j$, $\widehat{h}$ and $\widehat{\beta}$), we estimate housing market price for housing asset i at time t

$$\widehat{y}_{it} = \widehat{\alpha} + \widehat{h}(z_i, t) + \sum_{j=1}^{J} \widehat{f}_j(x_{it}) + \mathbf{X}\widehat{\beta} \tag{2}$$

As our dataset is large (more than 7M transactions, 36M of housing), we split the procedure by commuting zone to mitigate computational issues. Our imputation model is then composed of 293 independent models. It also has practical advantages, as it handles geographical discontinuities such as the Mediterranean Sea for a pooled model as Corsican commuting zones are spatially disjoint from others. We also introduce observations located in a 5-km area from the commuting zone area to prevent from border effects. According to Equation 2, we estimate the appraised value of the entire housing stock each January 01^{st} during the 2011–2019 period. We report from Appendix A.6 to Appendix A.10 additional results for the imputation procedure, including partial correlation for independent variables, joint significance for additive smoothing splines, and imputed housing values.

Our housing market price imputation confirms sizeable heterogeneity both in the cross-sectional and the longitudinal dimensions. Although the unitary housing price based on location is left-skewed,[7] unitary housing price ranges from 0.5k to 10k euros per square metres. It reinforces the spatial dimension of housing fundamental values. Moreover, the housing price dynamics also exhibit spatial heterogeneity, as annual gross rate vary from -2.5% to 5.0%.

Our price imputation model confirms the ability of local housing markets to generate heterogeneous capital gains (and loss). In addition, we retrieve sizeable cross-sectional differences both within and between metropolitan areas that support the ability of housing assets to drive inequalities.

2.3 Housing Transfers

We exploit the joint exhaustive feature of the transaction dataset and the panel data to identify housing transfers indirectly. We merge the transaction dataset to housing experiencing a change in their owner's composition making the best of common housing ID.[8] We condition the merging process using the temporal dimension as owners' change must occur in the same year or the year after than the transaction to prevent for potential delay in fiscal source update. We consider our merging process as valid as we match more than 99.15% of housing transactions restricted to private owners, which are those likely to be concerned by housing transfers (for detailed results about methodological choices for the transfers identification, see Appendix A.13). We expect remaining unmatched observations to arise from reverse mortgage operations and transactions for usufruct rights.

[7] For detailed results, see distribution on Appendix A.11

[8] From the panel data with housing as the observation unit, we identify sequentially, housing which experience change in the owner's composition.

Our main contribution to the exhaustive identification of housing transfers with detailed characteristics has two shortcomings resulting from methodological choices. Firstly, as we observe owners' changes sequentially every year, we omit multiple changes within a year. Then, if one transfer occurs the same year as a housing transaction for a particular observation,[9] our method fails to identify the change resulting from housing transfers. Secondly, we only identify housing transfers *between* private owners leaving rental investment companies out of scope. As a consequence, it brings two recommendations for the interpretation of the results. Firstly, our measure of the amount of housing transfers is likely a downward estimation. Secondly, our resulting transfer dataset is not suitable to study housing portfolio choices consecutive to receiving a transfer, as quick resale are not identified. Despite these two shortcomings, the transfers dataset provide detailed information about housing being transferred or characteristics of both legatees and recipients.

2.4 First-time Ownership

Finally, we make the best of the longitudinal feature of the data derived from fiscal sources to identify first-time owners according to the housing policy definition.[10] Empirically, we observe annually whether individuals are owners of their main residence. We consider as first-time owners, new homeowners who have not owned their residence for two years. The housing policy definition has practical advantages in that it does not require tenure status to be considered over the whole life cycle. Consequently, the reduced time span for the panel data is not bounding for the identification.[11]

The lack of common definition for first-time ownership makes the comparison difficult. The housing policy definition is likely to provide an upward estimation of first-time owners in comparison with the statistical definition provided by the INSEE. In addition, despite the growing interest for first-time ownership in academic works according to their impli-

[9] Assuming that housing change of owner from O_1 to O_2, and consecutively O_2 to O_3 within the same year, we only observe change O_1 to O_3 from the housing stock database.

[10] The policy definition is less restrictive in regards with the statistical one. It requires that households have not owned their main residence for a period of two years.

[11] The only limitation is that we precisely estimate the number of first-time owners from 2013, as we require to observe tenure status in 2011 and 2012.

cations on a life-time perspective, our identification is the first to exhaustively identify this population as far as we know. Making the best of the fiscal data, we observe their housing choices, including surface or location choices.

3 Empirical Results

Our panel data set for private homeowners and housing in France covers a 9-year period. We identify 34.35M unique owners with at least one property over the 2011–2019 period. We estimate that the housing stock is composed of 38.17M unique housing, including both new housing and demolished one.

The average gross housing wealth within owners equals in 2019 (respectively 2011) 170.1k euros (respectively 158.4k) per owner, each owner having 1.58 properties (respectively 1.58) on average. Average owner is 55.9 years old in 2019, while it is 54.5 years old in 2011. We estimate that the number of individuals with at least one property in 2019 (respectively 2011) equals 29.17M (respectively 27.67M). In addition, the average housing assets in 2019 is appraised to 176.7k euros (respectively 165.6k). Finally, we estimate that transfer represents 36.2% of owners change over the 2011–2018 period. The mean age recipient is 47.3 years old, while average legatees is 86.6 years old.

We illustrate the HWP (recall Housing Wealth Persistence) at a commuting zone scale. Then, we seek for potential drivers, including the housing market dynamics (Section 3.2), the location decisions for main residence (Section 3.3), the housing transfers (Section 3.4), and the property of rental assets (Section 3.5).

3.1 Housing Wealth Persistence

We first document the HWP for the commuting zone area perimeter. We simply compute the average level of housing wealth for residents in commuting zone in 2011 (Panel A, Figure 1) and the average annual gross rate for wealth (Panel B, Figure 1), without targeting specific population of homeowners. We then regress the gross rate over the

2011–2019 period on the initial level of wealth to provide evidence that the cross-section differences in housing wealth between commuting zones persist over time.

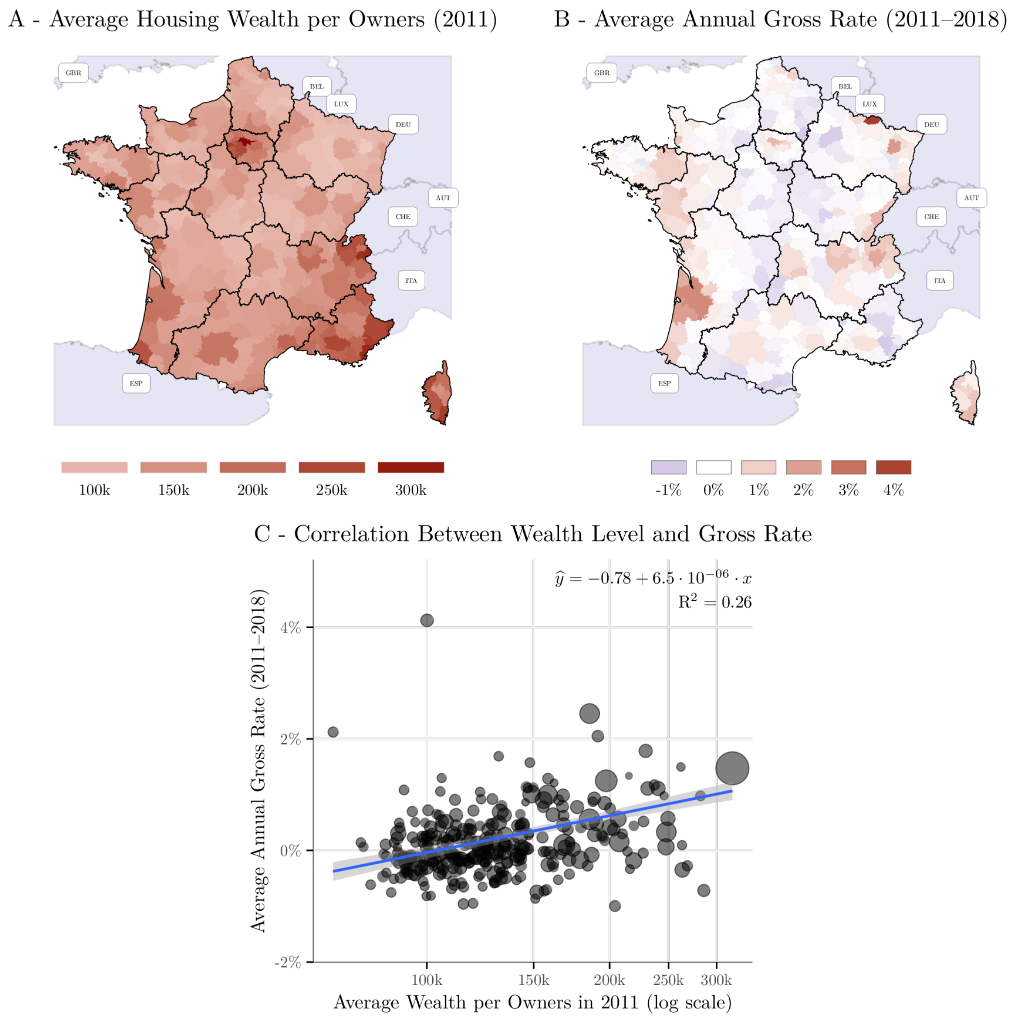


Figure 1: Housing Wealth Persistence between Commuting Zones of homeowners' main residences

Notes: We report the average initial level of housing wealth for homeowners (Panel A) and the average annual gross rate over 2011-2019 (Panel B) according to their place of residence. Our results are reported by commuting zone. We also report the linear relationship between the two variables (Panel C). It represents the simple WLS estimation between the average annual gross rate (y-axis) and the initial level of wealth (x-axis). We weight the commuting zone according to the number of homeowners in 2011 (size of the dot).
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

Although it is not deterministic ($R^2 = 0.26$), we estimate a positive relationship between the gross rate and the initial level of wealth for commuting zones (Figure 1, Panel C). Some commuting areas are an exception, such as the Luxembourg border and the Bordeaux area, which have the highest gross rate. Conversely, the Mediterranean coast, despite its wealthy inhabitants, offers a lower accumulation rate for the 2011-2019 period. Nevertheless, the wealth evolution of housing wealth is more pronounced for the commuting zones with the highest average wealth in 2011.

The HWP can be driven by several factors. On the one hand, the dynamics of the housing market affect the accumulation of housing wealth through differentiated capital gains. Our imputation procedure supports the potential for heterogeneous wealth evolution arising from housing market dynamics. However, the potential spatial differences between the location of assets and the owners' main residence may redistribute capital gains across space. The wealth evolution for homeowners is thus the average of asset evolution based on their locations. On the other hand, changes in the housing portfolio resulting from the moving to another main residence, inheriting a house or acquiring a rental property have a direct influence on housing wealth evolution. We then document the relationships between housing market dynamics, portfolio choices and HWP.

3.2 HWP for homeowners with a constant housing portfolio

While the local dynamics of housing markets are well described in the housing literature, we first document how it contributes to the housing wealth persistence at an individual level. Yet, we expect the housing wealth appreciation to differ at least slightly from local housing markets dynamics. Indeed, we identify two drivers with opposite effects. On the one hand, the prevalence of owner-occupation in individual portfolios suggests that wealth accumulation follows the same pattern as housing market dynamics. On the other hand, as the majority of the French housing stock is held by multiple owners,[12] the locations of owners and assets are likely to differ. It is then mainly the prevalence of local assets that determines the similarities between housing market dynamics and local housing wealth development.

To isolate the correlation between housing market dynamics and HWP, we select owners who do not experience portfolios' change between two consecutive years. By doing so, any variation of individual portfolios arises from changes in housing asset imputation, and therefore market evolution. Every year, we estimate that more than 93.7% of homeowners

[12] We estimate that almost 60% of the housing stock is held by individuals with at least two dwellings. André and Meslin (2021) estimate this share to be two thirds in 2017.

have no changes in their housing portfolios. We annually select nearly 25M of individuals between 2011 and 2018.

Our outcome of interest is the housing wealth variation (noted $w_{it+1} - w_{it}$) between two years. The logarithm transformation normalises the wealth accumulation according to the sum being invested.[13] We introduce as explanatory variable the initial level of wealth and spatial location of residence place to discuss the HWP. The location of residence place is defined according to the economic status of the commuting zone (heterogeneity between) and the distance from the CBD (heterogeneity within). In addition, we control by individuals' age. It yields annually to the estimation of

$$log(w_{it+1} - w_{it}) = \alpha_t + h_t(d_{it}, \ell_{it}) + f_{1t}(w_{it}) + f_{2t}(a_{it}) + \varepsilon_{it} \tag{3}$$

with w_{it} gross housing wealth for individual i at time t; α_t intercept, h_t bivariate smoothing function to account for the distance to the CBD (d_{it}) and built-up area for commuting zone (ℓ_{it}) based on individuals' location; a_{it} individual age; while ε_{it} represents idiosyncratic error term. We do not adopt pooled models to avoid computational issues and estimate Equation 3 separately for each year from 2012 to 2018.

Relationship of interest are respectively f_{1t} and h_t as it relates respectively to the initial level of wealth and spatial locations. We consider that both functions are relevant to the HWP. Firstly, any positive relationship between accumulation rate and initial level of wealth support the HWP individually. Second, spatial heterogeneity can either amplify or mitigate differences in housing wealth. Empirically, both functions are specified using additive spline transformations. The variance of errors is minimised by penalised least squares, while smoothing parameters are estimated using restricted maximum likelihood criteria. For clarity reasons, we report the spatial contribution for three distinct commuting zones based on their economic status. We range built-up area from 30% (low commuting zone) to 70% (top commuting zone) with 50% as intermediate value.

[13] It mainly overcomes the main drawbacks of level variation. For instance, a 10k increase is not as meaningful for a 500k housing wealth than a 50k one.

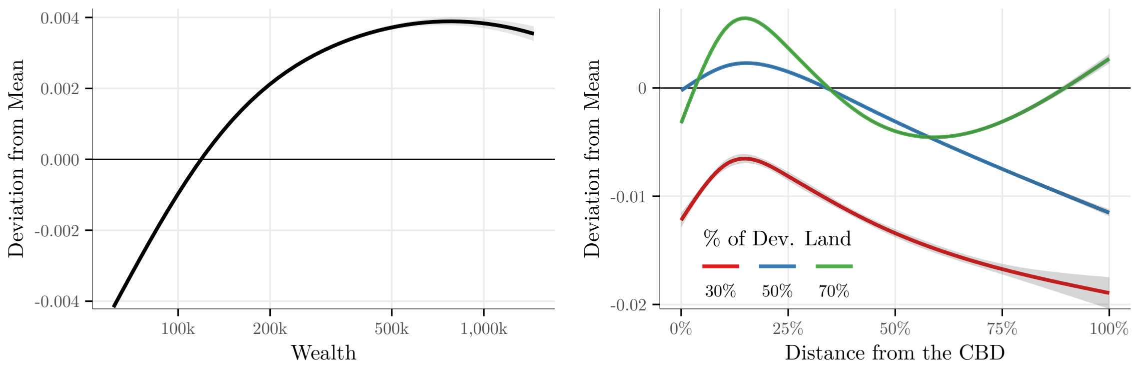


Figure 2: Partial Correlation Between Housing Wealth Variation, Initial Level of Wealth and Location of Main Residence in 2018

Notes: We report the relationship between the initial level of wealth and the variation in housing wealth at the individual level (left panel). We also report the relationship between housing wealth variation and the location of the main residence (right panel). We select three commuting zones based on their economic status to account for potential heterogeneity. These results come from the estimation of Equation 3. Our observation unit is distinct owners who do not experience a portfolio change in 2018. We report confidence intervals at the 95% confidence level. For annual results over the period 2012–2018, see Appendix B.1. We use the `mgcv` R package to implement additive transformations of variable with endogenous definition of degree of freedom.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

We find that portfolio appreciation stemming from housing markets increases with initial level of wealth regardless the studied year (Figure 2, left panel).[14] Thus, housing appears to behave similarly to other assets, by providing larger gains to the most expensive properties. Yet, these greater returns is not the result of higher risk. In a negative economic context (e.g. the 2012-2014 period of falling housing values), the potential capital loss is smoother for the wealthiest, while in a positive economic context (e.g. the 2017-2019 period) the capitalisation is more pronounced. As a result, housing wealth is persistent, as the wealthiest benefit from the highest growth and hold the most secure assets.

In addition, conditional on individual wealth, we find that the gross rate is heterogeneous according to the place of residence. Firstly, the wealth accumulation depends on the characteristics of the commuting zone. The deviation from the mean increases with the economic status of economic zone. This is in line with (Eggum and Larsen, 2021) findings that individuals who have held assets in the top commuting zones (in her case, Oslo) are able to accumulate more wealth over the life cycle. Secondly, the gross rate appears to decrease with distance from the CBD. As a result, owners living in central areas with easier access to desirable amenities benefit from a higher accumulation rate than those

[14]We do not report annual results for clarity reasons. See in Appendix B.1

living on the urban fringe.

The dynamics of housing markets are likely to reinforce the spatial inequalities for wealth. Indeed, while the housing wealth appreciation is increasing with individual wealth, we document spatial heterogeneity according to the residence place. Individuals living in top commuting zones and central areas are less likely to experiment capital losses in non-favourable economic context and more likely to receive higher capital gains in positive one. However, although housing markets generate heterogeneous capital gains, it mainly depends on the residence place locations, and thus housing portfolio choices.

3.3 HWP for homeowners that change their main residences

In line with the importance of the residence place for homeowners, we document spatial sorting for owner-occupiers according to previous housing endowments. As housing prices diverge from income, the financial capabilities are more likely to depend on wealth rather than income. Focusing on purchase achieved for main residence over the 2012-2017 period,[15] we empirically assess the relationship between housing characteristics, including location, and previous housing endowment for owner-occupiers purchases. We motivate our choice to restrict to owner-occupying as it directly relates to the spatial sorting, unlike rental investment.

Our outcome of interest to discuss spatial sorting is the distance from the CBD following the traditional monocentric model. In addition, we use the transaction price to provide partial correlations between transaction price and initial level of housing wealth (in this case, the year prior to the purchase). We enable the wealth relationship to vary with commuting zone characteristics. We select the total housing wealth of buyers rather than the average value as it better reflects financial capabilities, although this specification does not affect the results. Moreover, as the distance from the CBD might also capture heterogeneity in housing characteristics, we control for housing size. Our model specification contains time fixed effects to capture for the general trend and average purchaser

[15] Given that tenure status is provided with a delay of one year, we cannot identify main residences for the 2019 year.

age. It yields to the estimation of

$$y_{ijt} = \alpha_t + h_t(w_{it}, \ell_{jt}) + \mathbf{X}\beta + g(q_{jt}) + \varepsilon_{ijt} \tag{4}$$

with y_{ijt} outcome for purchasers groups i for housing j at time t; α_t time fixed effects; w_{it} sum of initial level of wealth for purchasers group i; ℓ_{jt} built-up area in commuting zone where housing j is located; q_{jt} housing size; $\mathbf{X}$ structural characteristics j including housing type and construction period; while ε_{ijt} represents idiosyncratic error term. Our sample is composed of 3,156,974 observations, including housing purchased by first-time owners. Their wealth is set to zero.

To discuss HWP for homeowners who change their main residence, we focus on the relationship between housing choice (dependent variable) and initial level of wealth (covariate). Indeed, it relates to both the value of the housing stock (for the price outcome) and spatial sorting (for the distance outcome). Any positive relationships between initial level of housing wealth and housing transaction would support the HWP, while the location choices reflect spatial sorting. Given the cross-sectional heterogeneity of local housing markets, we allow the partial correlation with wealth to vary with the economic status of commuting zone. However, we cannot infer a wealth effect on housing decisions for main residence as initial level of housing wealth is endogenous and may reflect unobservable variables such as heterogeneous preferences. Empirically, we define h_t as additive bivariate transformation, with endogenous shrinkage procedures to set effective degree of freedom. We report the relevant contributions for the 2017 year for clarity purposes, but full results are detailed in Appendix B.3 despite similar pattern over time.

The choice of the main residence of homeowners supports the HWP. The transaction price increases continuously with the initial level of wealth, regardless of the type of commuting zone considered (Figure 3, left panel). The difference is particularly pronounced for first-time owners, who buy the cheapest housing within the commuting zone. We expect that the lower level of deposits resulting from the lack of previous housing wealth is at least partly responsible for these differences. Moreover, the cross-sectional differences between

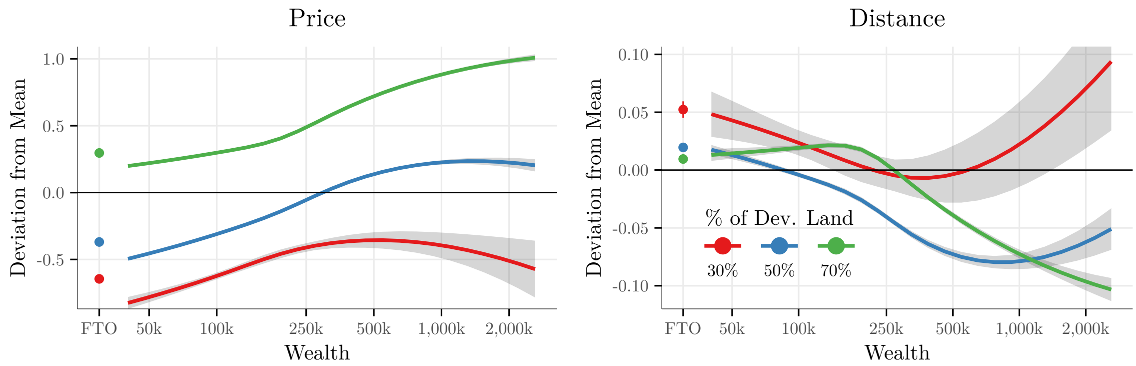


Figure 3: Partial Correlation of Initial Level of Wealth with Housing Decisions within Commuting Zones for Owner-Occupiers for 2017

Notes: We report the partial correlation for the interaction between the economic status of the commuting zone and the initial wealth level of the buyer from Equation 4. The left panel corresponds to the transaction price, while the right panel corresponds to the distance from the CBD. Our results are limited to the year 2017, but full results are available in Appendix B.3. We introduce variation in commuting zone characteristics as we estimate the wealth correlation for commuting with 30%, 50% and 70% of built-up area in 2000. We also introduce first-time homeowners (dot, left side of plot). We use the `mgcv` R package to implement additive transformations of variable with endogenous definition of degree of freedom.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

commuting zones are substantial. For example, we estimate that owners with an initial wealth of 250k purchase housing that is 2.7 times more expensive than owners in the medium commuting zone with a similar initial wealth. This suggests that either debt or capital gains being reinvested are likely to increase with the economic status of the commuting zones, given similar financial capabilities. It reinforces that housing market is local, with no spatial sorting *between* commuting zone according to financial capabilities.

The heterogeneity observed for transaction price stems partly from the location choices (Figure 3, right panel). We observe that on average, distance from the CBD decreases with financial capabilities, except for the bottom commuting zones. Moreover, first-time owners locate themselves at higher distance from the average housing choices, although we find no significant temporal trends that indicate a reinforcement of the fringe locations for this population. Then, the difference in transaction price is at least partly due to spatial sorting according to initial wealth.

Despite a general relationship indicating that distance from the CBD decreases with initial level of wealth, there remains spatial heterogeneity according to the economic status of the commuting zone. In fact, for similar initial level of wealth, individuals in the top commuting zones are located further away than those in the middle zones, despite the more expensive purchase. The difference in location decisions is likely to reflect important heterogeneity in the fundamental value of housing between commuting zones. Finally, low commuting zones exhibit U-shaped relationship. Then, we expect the spatial distribution of amenities to diverge from those in the top commuting zones ones (Brueckner et al., 1999) according to their nature (Lee and Lin, 2018). In addition, the assumption that jobs are concentrated in the centre of the commuting zone may be less credible.

The spatial sorting appears to contribute to the HWP as the wealthiest individuals purchase the most expensive asset within commuting zones, at closer distance from the CBD. Beyond the direct effect of utility derived from the consumption of local amenities, it increases the likelihood to benefit from higher capital gains according to the housing market dynamics. Indeed, the central areas seem to offer better returns in the long term, resulting in the most secure asset in a negative context and more favourable in a positive one. Consequently, the characteristics of main residence for housing consumption, especially the location decision, is likely to drive future wealth accumulation.

3.4 HWP for homeowners that receive a housing inheritance

In addition to the relationship between homeowners decisions for consumption and HWP, non-monetary transfers are commonly assumed to drive wealth inequality on the long-term through intergenerational persistence (De Nardi, 2004; Garbinti et al., 2021). Then, we look for potential persistence in housing wealth resulting from housing transfers. As housing transfers represent nearly one third of owners' change for housing over the 2012–2018 period,[16] their ability to either contribute or mitigate HWP is significant.

The average share being received by individuals is estimated to be close to 100k (detailed results in Appendix B.5). However, the 2012 year is singular with the highest average housing transfers (nearly 120k per recipients) over the studied period. We lay out this singularity by potential adaptation in behaviour due to policy reform introduced in August 2012. The reform raised inheritance tax which might have caused anticipation for legatees

[16] *Sources*: Authors' Calculation based on enhanced property tax files.

to donate prior the introduction of the reform. Agents behaved similarly for tax reforms applied to gasoline price (Coglianese et al., 2017), although economic goods differ significantly. Our results estimate that restricted to housing wealth, average recipient does not have to pay any inheritance tax assuming direct ascending line between recipients and legatees, except for the 2012 year.[17]

We therefore restrict our sample to housing transfers recipients over the 2012–2018 period. For each recipient, we observe the received share derived from our housing value imputation and their residence place. In addition, we observe whether recipients were previously housing owners, and if so, their initial level of housing wealth (i.e. prior to the housing transfers). Additional individual characteristics such as age or gender are introduced as control variables. Consequently, we regress the received share by individual characteristics, including previous housing wealth and individual location:

$$log(s_{it}) = \alpha_t + h_t(d_i, \ell_i) + f_{1t}(w_{it}) + f_{2t}(a_{it}) + \varepsilon_{it} \tag{5}$$

with s_{it} share received by transfer recipients i at time t; α_t time fixed effects; d_i distance from the CBD of individual i; ℓ_i built-up area of commuting zone where individual i is living; w_{it} individual housing wealth prior to the transfers; a_{it} individuals age prior to the transfers; finally ε_{it} corresponds to idiosyncratic error term. As previously, unknown functions h_t, f_{1t} and f_{2t} are empirically specified using additive splines, with smoothing parameters being defined endogenously using restricted maximum likelihood criteria. Our sample is composed of 1,852,126 observations about transfer recipients.

The contribution of housing transfers to the HWP derives from two facts. Firstly, it stems from the individual aspect following the wealth relationship. Indeed, if the wealthiest owners benefit from larger transfers, the housing wealth would persist over time. Secondly, it results from the spatial heterogeneity. Conditionally on recipients' wealth, any positive relationship between housing transfers and economic status of commuting zone would support the HWP spatially. Consequently, the relevant parameters are f_{1t}

[17]Indeed, currently, legatees fees concerns value above 100k.

for the individual effect and h_t for the spatial one. As we allow both heterogeneity to vary over time, we report results for the 2018 year despite no sizeable differences over the 2011–2019 period, only for clarity reasons.[18]

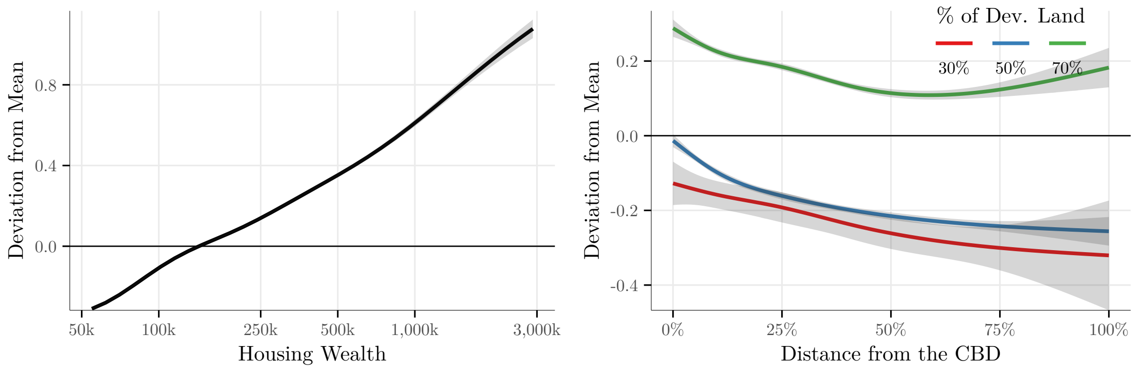


Figure 4: Partial Correlation of Initial Level of Wealth and Individual Locations with Transfer Share in 2017

Notes: Our results are derived from estimating Equation 5 restricted to housing transfer recipients. The left panel reports the partial correlation according to the initial level of wealth. We use a log transformation for the initial level of wealth. The right panel reports the partial correlation for homeowners' place of residence, using heterogeneity both between and within commuting zones. Our observation unit is recipients of housing transfers between 2012 and 2018. We report confidence intervals at the 95% level. We report all our studied period ranging from 2012 to 2017 for clarity reasons in Appendix B.6. We use the `mgcv` R package to implement additive transformations of variable with endogenous definition of degree of freedom.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

However, the housing share being transferred increases with the wealth of recipients (see in Figure 4, left panel), supporting the intergenerational persistence over time (De Nardi, 2004; Garbinti et al., 2021), and thus HWP. Although we report partial correlations for 2018 year, the overall patterns remain unaffected over time (see in Appendix B.6) despite slight differences among the wealthiest individuals.[19]

In addition to the observed heterogeneity according to individual characteristics, housing transfers spatially concentrate housing wealth both within and between commuting zone (see in Figure 4, right panel). Firstly, regardless of the location within the commuting zone, the share received in the top commuting zones is at least 10% higher than the average share. Meanwhile, received shares are lower than the national average for beneficiaries living in the middle and low commuting zones. Secondly, the share received decreases with distance from the CBD, regardless of the economic status of the commuting zone.

[18]We report detailed results per year over the 2012–2018 period in Appendix B.6.

[19]Indeed, while we estimate that the share received by recipients with 1M of housing wealth was 71.6% higher than the mean in 2012, it reaches 78.6% in 2018. The increase is, however, not driven by the singularity of the 2012–year. For instance, the difference in 2014 is estimated to 71.8%.

Recipients living in central areas benefit from larger housing transfers than those living on the urban periphery. The gradient is homogeneous and remains similar over the period 2012-2018.[20]

Both observations support the HWP. While the consequence of increasing the share received with the initial level of housing wealth on the HWP is straightforward, we suggest a potential mechanism for spatial heterogeneity. If housing transfers are capitalised directly (e.g. through sale), the cash flow may be reinvested in the residence place either directly (e.g. through renovation) which capitalise into price, or indirectly by changing the location of the main residence. Recall that the choice of location for main residence is highly correlated with initial level of wealth (Section 3.3), and affects future wealth accumulation (Section 3.2). Finally, if housing transfers are not capitalised, they provide additional insurance for housing wealth, which may influence household behaviour with increased risk aversion.

3.5 HWP for homeowners with rental assets in their portfolios

Finally, we conclude the section on empirical results by focusing on homeowners with rental properties in their portfolio. These homeowners hold most of the housing stock in France and part of the location of their housing wealth is different from their place of residence. To understand how the HWP potentially results from the interaction between both locations, we consider two sub-populations. The first, defined as local investors, concerns individuals whose entire housing portfolio is located within the same commuting zone as their place of residence. Conversely, the latter refers to distant investors who have at least one rental property located in a different commuting zone from their place of residence.

The first sub-population is dominant as almost 80% of rental properties are within the same commuting zone as their owner-occupied home. Given that owner-occupation is widespread among private owners (more than 90% of homeowners are owner-occupiers),

[20] Full results are available in Appendix B.6

it provides a basis for HWP spatially as housing wealth depends on the dynamics of local housing markets. However, rental investment property may have mixed effects on HWP. On the one hand, if homeowners with the highest initial level of wealth hold rental assets in the most dynamic areas,[21] rental facilities would contribute to the HWP. On the other hand, if homeowners with the lowest initial level of housing wealth benefit from a higher gross rate due to rental properties located in different commuting zones, rental properties are likely to spatially redistribute housing wealth and thus mitigate the HWP.

We therefore calculate the average annual gross rate for each sub-population at the commuting zone level and compare it to the local housing market dynamics. We then regress the average annual gross rate for wealth evolution on the gross rate of housing located within the commuting area for both sub-populations.

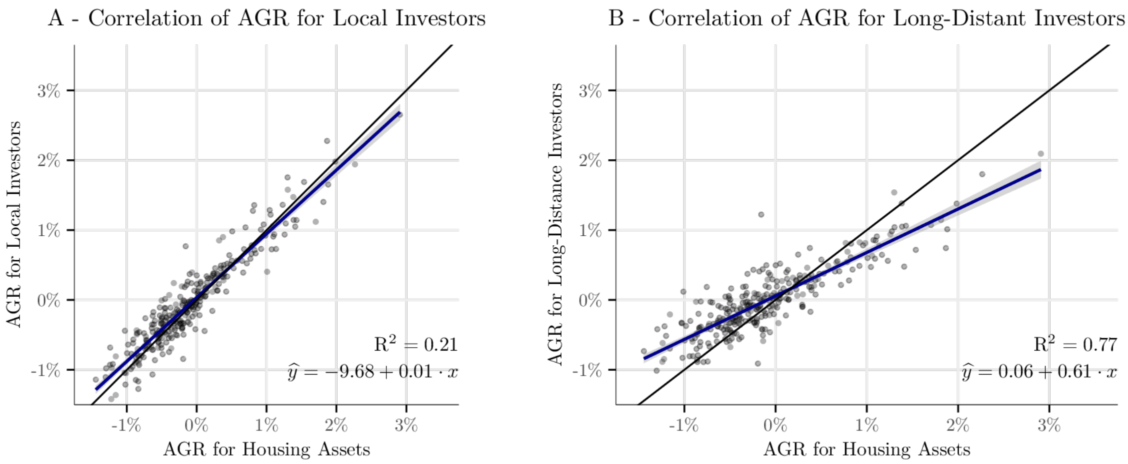


Figure 5: Correlation of Annual Gross Rates Between Wealth and Housing Market Dynamics for Local and Distant Investors

Notes: The left (respectively right) panel is the bivariate distribution of commuting zone according to average asset evolution within the commuting zone (x-axis) and local (respectively long-distance) investors (y-axis). We report the linear regression and the first bisector in addition to the coefficient of determination. Our population of interest is composed of homeowners with at least one rental property in their housing portfolio. AGR is the abbreviation for Average Growth Rate, expressed in percent. We exclude Alsace-Moselle homeowners as we cannot calculate their AGR due to data limitations.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

The long-distance investment weakens the relation between housing market dynamics and wealth evolution over the 2011–2019 period as expected. Nonetheless, the relationship we provide mitigate the HWP, as coefficient of correlation is lower than 1 (Figure 5, right panel). In fact, long-distance investors living in commuting zone with low or negative asset

[21]Recall that, despite the correlation, the relationship between initial level of wealth and average annual gross rate is not deterministic, see in Section 3.1.

evolution benefit from larger evolution. Hence, the long-distance investments smooth the wealth evolution over space, lowering expected results through an increase of the safety. In comparison, the relationship for local investors is strong, despite some deviations resulting from the within location of their rental assets (Figure 5, left panel). Again, the internal location for rental asset mitigate the HWP spatially as we observe similar pattern than long distance investment, despite lower intensity.

It therefore results that rental assets are likely to mitigate the HWP spatially, both within and between commuting zones, despite the fact that the intensity is more pronounced for the first-one. Consequently, the housing market by itself is not a driver for wealth inequality, as we illustrate that it mainly results from decisions investments including location for rental assets.

4 Conclusion

The housing wealth is a good proxy for overall wealth (Garbinti et al., 2021) resulting from its widespread aspect in individuals portfolio (Jordà et al., 2019). In addition, as it defines the main residence for homeowners, it conditions the access to economic opportunities over the life-cycle (Bilal and Rossi-Hansberg, 2021). Nevertheless, following Piketty and Zucman (2014) work, few works have specifically focused on housing assets and individuals' locations. We provide empirical insight supporting the need to consider jointly spatial sorting and wealth inequality to understand the housing wealth persistence (HWP).

Using exhaustive individual data for 2011-2019 on French homeowners and their properties, we document four insights based on different populations. All of our insights seek to understand the persistence of housing wealth, both spatially and individually. Firstly, variation that stem from the housing market mainly depends on the initial level of wealth and residence place. Secondly, the spatial sorting for owner-occupier depends on the initial level of wealth. The wealthiest individuals locate at the closest distance from the CBD. In addition, with similar financial capabilities, the distance from the CBD increases with the economic status of commuting zone despite larger housing investments. Thirdly, housing transfers contribute to the HWP, both individually and spatially. The share received increases with previous housing wealth, while residents of the top commuting zone benefit from a larger amount. Fourthly, the long-distance rental investment tend to mitigate the housing wealth persistence at a commuting zone scale. It allows homeowners with long-distance investments to have higher annual gross rate than housing market dynamics in low commuting areas.

The implications of HWP are numerous, and relates to different topics. Firstly, the correlation between distance from the CBD and previous housing wealth for owner-occupier yields to the observation that financially constrained households are likely to be more sensitive to the evolution of the commuting cost. In light of the evolution of the economic context (rising of gasoline price, land constraints), the attractiveness of fringe locations may decrease with potential consequences on spatial sorting and housing market dynamics. Secondly, the role of transfers, although being commonly discussed in inequality literature, is likely to determine further portfolio choices, including location of the main residence. For instance, the access to unaffordable locations could be made conditional on the amount of housing transfers received. Beyond the direct impact on the tenure status, transfers are likely to affect the location decision for recipients, with sizeable implications on the long-term.

Considering the contribution we make in terms of individual data to study housing wealth and spatial sorting, we consider that numerous contributions should follow and suggest potential leads. Firstly, despite the evidence on the persistence of housing wealth, we provide indications that rental assets are likely to have redistributive effects. Given that rental assets provide additional income flows to owners, it would be interesting to further investigate the potential redistributive effects of rental assets through the spatial dimension. Secondly, we suggest to draw causal relationships between housing transfers and further portfolio choices. Indeed, while most of the existing literature focuses on the impact on tenure decisions, we expect housing transfers to affect housing decisions within beneficiaries. This is an outstanding challenge due to the endogeneity of the share received. Thirdly, although beyond the scope of this paper, we suggest further discussion of the relationship between wealth accumulation and land restrictions. While the existing literature focuses on the price capitalisation that stem from restrictions, we suggest to detail the consequences on capital gains and housing portfolio decisions.

References

Ahlfeldt, Gabriel M., Stephan Heblich, and Tobias Seidel (2023), "Micro-Geographic Property Price and Rent Indices". *Regional Science and Urban Economics* 98.103836.

Albouy, David, Gabriel Ehrlich, and Yingyi Liu (2016), "Housing Demand, Cost-of-Living Inequality, and the Affordability Crisis". *NBER Working Paper* 22816, p. 62.

Alvaredo, Facundo, Anthony B. Atkinson, and Salvatore Morelli (2018), "Top Wealth Shares in the UK over More than a Century". *Journal of Public Economics* 162, pp. 26–47.

André, Mathias and Olivier Meslin (2021), "Et Pour Quelques Appartements de plus : Étude de La Propriété Immobilière Des Ménages et Du Profil Redistributif de La Taxe Foncière". *INSEE Working Paper*, p. 93.

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.

Banzhaf, H. Spencer and Omar Farooque (2013), "Interjurisdictional Housing Prices and Spatial Amenities: Which Measures of Housing Prices Reflect Local Public Goods?" *Regional Science and Urban Economics* 43.4, pp. 635–648.

Benhabib, Jess, Alberto Bisin, and Mi Luo (2017), "Earnings Inequality and Other Determinants of Wealth Inequality". *American Economic Review* 107.5, pp. 593–597.

Bilal, Adrien and Esteban Rossi-Hansberg (2021), "Location as an Asset". *Econometrica* 89.5, pp. 2459–2495.

Black, Sandra (1999), "Do Better Schools Matter? Parental Valuation of Elementary Education". *Quarterly Journal of Economics* 114.2, pp. 577–599.

Blanco, Miguel Artola, Luis Bauluz, and Clara Martínez-Toledano (2021), "Wealth in Spain 1900–2017: A Country of Two Lands". *The Economic Journal* 131.633, pp. 129–155.

Boserup, Simon H., Wojciech Kopczuk, and Claus T. Kreiner (2016), "The Role of Bequests in Shaping Wealth Inequality: Evidence from Danish Wealth Records". *American Economic Review* 106.5, pp. 656–661.

Brueckner, Jan K., Jacques-François Thisse, and Yves Zenou (1999), "Why Is Central Paris Rich and Downtown Detroit Poor?" *European Economic Review* 43.1, pp. 91–107.

Brunsdon, Chris, A. Stewart Fotheringham, and Martin E. Charlton (1996), "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity". *Geographical Analysis* 28.4, pp. 281–298.

Chay, Kenneth Y. and Michael Greenstone (2005), "Does Air Quality Matter? Evidence from the Housing Market". *Journal of Political Economy* 113.2, pp. 376–424.

Coglianese, John, Lucas W. Davis, Lutz Kilian, and James H. Stock (2017), "Anticipation, Tax Avoidance, and the Price Elasticity of Gasoline Demand". *Journal of Applied Econometrics* 32.1, pp. 1–15.

De Nardi, Mariacristina (2004), "Wealth Inequality and Intergenerational Links". *The Review of Economic Studies* 71.3, pp. 743–768.

Diao, Mi, Delon Leonard, and Tien Foo Sing (2017), "Spatial-Difference-in-Differences Models for Impact of New Mass Rapid Transit Line on Private Housing Values". *Regional Science and Urban Economics* 67, pp. 64–77.

Duranton, Gilles and Diego Puga (2015), "Urban Land Use". *Handbook of Regional and Urban Economics*. Vol. 5A. Boston, MA: Elsevier, p. 698.

Eggum, Terje and Erling Røed Larsen (2021), "Is the Housing Market an Inequality Generator?" *Housing Lab Working Paper* 2, p. 43.

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

Fotheringham, A. Stewart, Chris Brunsdon, and Martin Charlton (2002), *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, England ; Hoboken, NJ, USA: Wiley.

Gabriel, Stuart and Gary Painter (2020), "Why Affordability Matters". *Regional Science and Urban Economics* 80.103378.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty (2021), "Accounting for Wealth-Inequality Dynamics: Methods, Estimates, and Simulations for France". *Journal of the European Economic Association* 19.1, pp. 620–663.

Glaeser, Edward L, Joseph Gyourko, and Raven E Saks (2005), "Why Have Housing Prices Gone Up?" *American Economic Review* 95.2, pp. 329–333.

Grossman, Sanford T and Guy Laroque (1987), "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumer Goods". *Econometrica : journal of the Econometric Society* 58, pp. 25–21.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst (2013), "Endogenous Gentrification and Housing Price Dynamics". *Journal of Public Economics* 100, pp. 45–60.

Hilber, Christian A.L. and Christopher Mayer (2009), "Why Do Households without Children Support Local Public Schools? Linking House Price Capitalization to School Spending". *Journal of Urban Economics* 65.1, pp. 74–90.

Hilber, Christian A.L. and Andreas Mense (2021), "Why Have House Prices Risen so Much More than Rents in Superstar Cities?" *Geography and Environment Discussion Paper Series* 30, p. 51.

Huggett, Mark, Gustavo Ventura, and Amir Yaron (2011), "Sources of Lifetime Inequality". *American Economic Review* 101.7, pp. 2923–2954.

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.

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.

Kuminoff, Nicolai V., V. Kerry Smith, and Christopher Timmins (2013), "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets". *Journal of Economic Literature* 51.4, pp. 1007–1062.

Lee, Sanghoon and Jeffrey Lin (2018), "Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income". *The Review of Economic Studies* 85.1, pp. 663–694.

Lee, Seong Woo and Dowell Myers (2003), "Local Housing-Market Effects on Tenure Choice". *Journal of Housing and the Built Environment* 18.2, pp. 129–157.

Li, Ziqi and A. Stewart Fotheringham (2020), "Computational Improvements to Multi-Scale Geographically Weighted Regression". *International Journal of Geographical Information Science* 34.7, pp. 1378–1397.

Li, Ziqi, A. Stewart Fotheringham, Wenwen Li, and Taylor Oshan (2019), "Fast Geographically Weighted Regression (FastGWR): A Scalable Algorithm to Investigate Spatial Process Heterogeneity in Millions of Observations". *International Journal of Geographical Information Science* 33.1, pp. 155–175.

McMillen, Daniel P. (2010), "Issues in Spatial Data Analysis". *Journal of Regional Science* 50.1, pp. 119–141.

Murakami, Daisuke, Narumasa Tsutsumida, Takahiro Yoshida, Tomoki Nakaya, and Binbin Lu (2020), "Scalable GWR: A Linear-Time Algorithm for Large-Scale Geographically Weighted Regression with Polynomial Kernels". *Annals of the American Association of Geographers*, pp. 1–22.

Ortalo-Magné, François and Andrea Prat (2016), "Spatial Asset Pricing: A First Step". *Economica* 83.329, pp. 130–171.

Pfeffer, Fabian T and Nora Waitkus (2021), "The Wealth Inequality of Nations". *American Sociological Review* 86.4, pp. 567–602.

Piketty, Thomas and Gabriel Zucman (2014), "Capital Is Back: Wealth-income Ratios in Rich Countries 1700–2010". *Quarterly Journal of Economics* 129.3, pp. 1255–1310.

Roback, Jennifer (1982), "Wages, Rents, and the Quality of Life". *Journal of political Economy* 90.6, pp. 1257–1278.

Saez, Emmanuel and Gabriel Zucman (2016), "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data". *Quarterly Journal of Economics* 131.2, pp. 519–578.

Saiz, Albert (2010), “The Geographic Determinants of Housing Supply”. *Quarterly Journal of Economics* 125.3, pp. 1253–1296.

Sinai, Todd and Nicholas S. Souleles (2005), “Owner-Occupied Housing as a Hedge Against Rent Risk”. *Quarterly Journal of Economics* 120.2, pp. 763–789.

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

Van Ommeren, Jos and Michiel Van Leuvensteijn (2005), “New Evidence of the Effect of Transaction Costs on Residential Mobility”. *Journal of Regional Science* 45.4, pp. 681–702.

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

Zabel, Jeffrey E. (2012), “Migration, Housing Market, and Labor Market Responses to Employment Shocks”. *Journal of Urban Economics* 72.2-3, pp. 267–284.

A Data

A.1 Descriptive Statistics about Housing Stock Derived From Fiscal Sources

Table A.1: Descriptive Statistics for Housing Stock

Variable	N	Mean	Median	Std Dev	Quantile 1	Quantile 3	Min	Max
House								
Surface	19,960,874	102.3	95.0	45.0	76.0	122.0	11.0	4,122.0
Dep. Surface	19,960,874	14.7	0.0	35.7	0.0	18.0	0.0	3,503.0
Building Year	19,960,874	1941	1971	77	1900	1992	1300	2019
Flat								
Surface	18,206,318	57.8	57.0	28.6	38.0	73.0	11.0	6,646.0
Dep. Surface	18,206,318	7.5	3.0	17.1	0.0	8.0	0.0	6,610.0
Building Year	18,206,318	1923	1970	152	1914	1993	1300	2019

Notes: The top panel reports main descriptive statistics for individual housings (named as house) for housing characteristics (surface, dependence surface, building year). Bottom panel provide similar features for collective housing (named as flat).
Sources: Fichiers Fonciers.

A.2 Descriptive Statistics for Housing Transactions Derived From Fiscal Sources

Table A.2: Descriptive Statistics for the DV3F database

Variable	N	Mean	Median	Std Dev	Quantile 1	Quantile 3	Min	Max
House								
Surface	4,042,638	100.9	94.0	41.5	75.0	120.0	11.0	1,440.0
Dep. Surface	4,042,638	61.2	47.0	57.3	21.0	85.0	0.0	8,749.0
Building Year	4,042,638	1938	1966	74	1900	1988	1300	2019
Price	4,042,638	200,699	167,000	159,252	109,000	249,000	10,001	15,750,000
Price per m^2	4,042,638	2,009.1	1,769.3	1,193.3	1,216.2	2,523.6	17.3	19,949.7
Flat								
Surface	3,407,612	57.9	56.0	27.1	39.0	72.0	11.0	1,500.0
Dep. Surface	3,407,612	7.6	4.0	14.3	0.0	9.0	0.0	3,025.0
Building Year	3,407,612	1927	1973	158	1930	2002	1300	2019
Price	3,407,612	195,529	152,000	184,297	98,500	228,150	10,001	17,865,830
Price per m^2	3,407,612	3,493.2	2,954.5	2,259.0	1,985.3	4,210.5	39.5	19,837.8

Notes: The top panel reports main descriptive statistics for individual housings (named as house), for housing characteristics (surface, dependence surface, building year), and transaction feature including the purchase price and unitary price. Bottom panel provides similar features for collective housing (named as flat).
Sources: DV3F.

A.3 Share of Built-up Area in 2000 Within Each Commuting Area

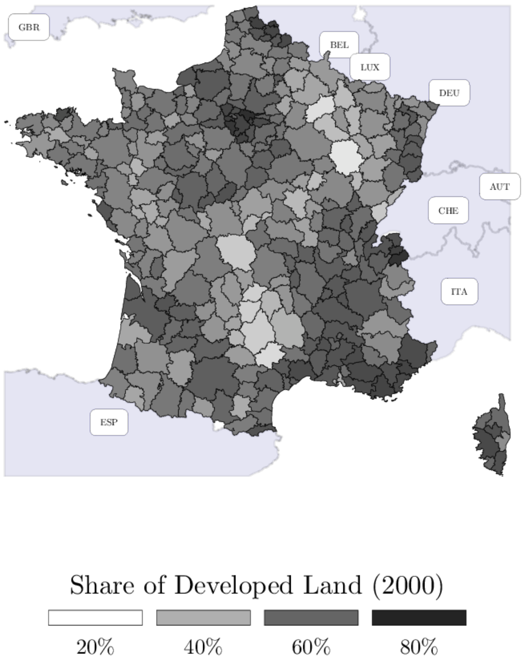


Figure A.1: Share of Developed Land According to Commuting Area

Notes: We report share of developed land in 2000 for each commuting area defined by the INSEE in 2010. We consider as developed parcels, parcel with at least one built premises prior to 2000 using the *Fichiers Fonciers*. We remove from the parcel considered as natural areas.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.Commuting areas are derived from the INSEE.

A.4 ID Construction

We create a national ID to overcome this issue by using the civil state available in the *Fichiers Fonciers*. Our new ID relies on the available civil stat. However, as data include some errors (not filled data, inconsistent date of birth, typographic errors, etc.), we aim to select the best approach to minimize errors (both type I and type II errors, see definition in the following).

Potential IDs First, we review the potential IDs we consider. Our potential IDs aim to reduce the likelihood of typographic errors stemming from misspecification in the fiscal data. Hence, we test 3, 5 and all characters derived from the first part to characterise names. In addition, we introduce variation according to the gender. Finally, we remove particle for names to avoid potential misspecification.

Table A.3: Potential National ID Based on Civil State and Name Truncating

Names	Gender	Place of Birth	Results
Part 1	Yes	No	M-PIERREDUPONT-DDMMYYYY
Part 1	Yes	Yes	M-PIERREDUPONT-DDMMYYYY-75PAR
3 char.	Yes	No	M-PIEDUP-DDMMYYYY
3 char.	Yes	Yes	M-PIEDUP-DDMMYYYY-75PAR
5 char.	Yes	No	M-PIERRDUPON-DDMMYYYY
5 char.	Yes	Yes	M-PIERRDUPON-DDMMYYYY-75PAR
Part 1	No	No	PIERREDUPONT-DDMMYYYY
Part 1	No	Yes	PIERREDUPONT-DDMMYYYY-75PAR
3 char.	No	No	PIEDUP-DDMMYYYY
3 char.	No	Yes	PIEDUP-DDMMYYYY-75PAR
5 char.	No	No	PIERRDUPON-DDMMYYYY
5 char.	No	Yes	PIERRDUPON-DDMMYYYY-75PAR

Notes: Example for a man who its last name is Dupont and its first name Pierre born in Paris (75).

Error Assessment Type I error corresponds to distinct individuals registered under the same national ID whereas type II error corresponds to same individual registered under different national ID. Empirically, we adopt the following definitions for type I and type II errors. A **type I error** is identified when the national ID is equal but there are differences in either last name, first name or date of birth (when they are available). Conversely, **a type II error** is defined by different national ID with similar civil state. We define similarity by enabling slight differences in the character strings that composed civil state. Specifically, the Levenshtein distance[22] must be lower or equal to one for both

[22]Levenshtein distance between two strings of characters corresponds to the number of characters to either remove, add or replace to obtain equivalent character strings.

first name and last name. We also add the condition of the date of birth equality when the date of birth is available.

Results We provide results from type I and type II errors according to the definition we adopt. For each ID, we estimate the overall error resulting from both type I and type II.

Table A.4: Overall Results for Potential Errors Assessments according to Potential ID

			All				Without Incomplete Data	
Names	Gender	Place of Birth	False	Omitted	Errors	Error Rate	Errors	Error Rate
Part 1	Yes	No	241,932	92,076	334,008	0.80%	92,126	0.22%
Part 1	Yes	Yes	203,298	926,014	1,129,312	2.71%	926,024	2.23%
3 char.	Yes	No	2,901,967	53,472	2,955,439	7.10%	2,711,382	6.52%
3 char.	Yes	Yes	266,576	920,024	1,186,600	2.85%	982,991	2.36%
5 char.	Yes	No	379,899	98,181	478,080	1.15%	291,158	0.70%
5 char.	Yes	Yes	177,742	928,664	1,106,406	2.66%	952,451	2.29%
Part 1	No	No	249,514	81,533	331,047	0.80%	81,587	0.20%
Part 1	No	Yes	209,907	922,285	1,132,192	2.72%	922,295	2.22%
3 char.	No	No	3,640,464	32,321	3,672,785	8.83%	3,420,467	8.22%
3 char.	No	Yes	285,328	914,532	1,199,860	2.88%	989,466	2.38%
5 char.	No	No	449,585	78,601	528,186	1.27%	334,623	0.80%
5 char.	No	Yes	189,188	923,657	1,112,845	2.67%	953,311	2.29%

Notes: The ID composed by the first part of names, the gender without the place of birth leads to 241,932 departmental accounts affected by false matching and 92,076 by omitted matchings. The overall error rate is then 0.80%. However, the error rate is only 0.22% for well-coded data.

The ID which minimizes the number of errors is the ID with the first character string without both place of birth and gender. The number of departmental accounts potentially affected by errors is 0.80% (if we consider all accounts). If we remove omitted matchings for observation with incomplete data, the error rate falls to 0.20%. This can be justified as we have not enough selective criteria to say whether or not they are true errors.

Consequently, we adopt the current ID for observation with misspecified civil state. It leads to an increase of type II errors. We rather prefer type II errors than type I to avoid the over-estimation of housing wealth, and consequently wealth inequalities. To close our choice, we select the ID composed of the first part of first and last names, in addition to the date of birth. For civil state with misspecification, we adopt the current ID. Finally, we anonymise the ID using integer trough random assignment. As variables selected in the civil state are invariant, we can apply the national ID from 2011 to 2020.

A.5 Comparison With Existing Database

The administrative data about housing stock we use are similar than those exploited by André and Meslin (2021) to analyse the distribution of housing properties in 2017 within private owners. We both select ownership rights that correspond to wealth definition (rather than income flow), i.e. full and bare ownership. They identify unique household based on address (as they have a cross-sectional approach) and assign detained housing properties.

However, as we extend the study period from 2011 to 2019 (while they solely focus on 2017), place of residence is no longer a determinant in identifying stable households due to moving.. Hence, we adopt unique private owners for stability reasons over time (households might experience separation for instance), being consistent with previous empirical works exploiting panel data (Bach et al., 2020; Fagereng et al., 2020). In addition, we do not impute housing properties detained through rental investment companies due to data limitations. On the contrary, André and Meslin (2021) have partly assigned housing properties detained via rental investment companies to individual owners. However, they have imposed stricter assumptions to deal with missing variables about shareholders. Consequently, our population of interest differs in regards with the housing properties under consideration.

Despite differences in methodological choices to identify unique owners that are the consequences of the differences in pursued objectives, our results with respect to multiownership and gender gaps are close to theirs for 2017. We estimate that nearly 60% of housing stock detained by private persons belongs to multiowners; it reaches two third according to André and Meslin (2021). Moreover, the gender gap is increasing with the number of properties detained by private persons. We report detailed results on Appendix A.5. Hence, although our methodological choices differ from theirs to extend studied period and provide panel data about owners and housing properties from 2011 to 2019, our results are close. It thus reinforces the credibility of both approaches.

Table A.5 reports the main methodological differences between both dataset. Although we focus on methodological differences, we also adopt similar choices, especially for property rights and housing definition.

Using the database provided by André and Meslin (2021) as reference, we compare the results obtained with our database to their main ones for 2017. Our estimation leads to differences in magnitudes of measured effects, but our results about multiownership and gender gaps are consistent with those provided by André and Meslin (2021).

Table A.5: Summary of Methodological Choices

Topic	André and Meslin (2021)	Current Paper
Observation Unit	Household	Individual
Period	2017	2011–2019
Rental Investment Companies	Assign to natural owners	Not assigned
Owners Location	French Resident	All
Spatial Perimeter	Metropolitan France	Metropolitan France
Housing Properties	Ind. and Coll.	Ind. and Coll.
Property Rights	Full owners and Bare-owners	Full owners and Bare-owners

Notes: We report main differences we identify between our methodological choices to overcome current ID issue of administrative data sources.

A.6 Comparison of Methods for Housing Value Imputation

Procedure We choose to adopt the Leave-One Out Cross Validation (LOOCV) procedure to compare methods. Our procedure is as follows: We sample 10% of the transaction datasets and we remove it from the transaction data. Then, we run the calibration step on the remaining transaction and we predict the value of the transaction based on the model. We itemize this process in order to have multiple predicted value. As this procedure is computationally intensive, we work on 50 iterations.

Perimeter We restrict our perimeter of interest to the *Côte d'Or* department. It is composed of metropolitan areas such as Dijon, middle-size cities such as Beaune, and rural areas. Hence, the department provides sufficiently diversified local contexts to assess the accuracy of housing value imputation to local characteristics. For the temporal dimension, we take advantage of the entire period for the dataset, namely 2010 to 2020. We select 39,799 transactions.

Model Specification We perform three distinct imputations through three different specifications. Firstly, we perform the estimation through OLS, with different spatial units for fixed effects. We also test as benchmark the OLS estimation without spatial fixed effects to account for spatial heterogeneities. The GWR is performed through the `scgwr` package (Murakami et al., 2020) which adapt seminal GWR for large datasets. Finally, we perform the estimation through GAM using the `mgcv` package (Wood, 2017) that enables to introduce spline transformations for continuous variable. The effective degree of freedom for the spline transformation is endogenously shrinked using iterated procedure. We model the spatial heterogeneity using bi-variate smoothing function using spatial coordinates of housing transactions.

Selected Variables To perform this comparison, we select housing characteristics to control for their contribution to the price. The following variables are introduced either in linear form or spline transformations in regards with the estimation procedure: housing surface; outbuilding surface (including cellar, garage, etc.); building age; housing type; balcony presence; time fixed effects using quarter aggregation.

Results We report results according to the average relative error resulting from the 50 iterations procedure we perform. It represents 198,995 individual estimations. Firstly, accounting for spatial heterogeneities appears crucial to obtain accurate estimation of housing value. Indeed, the benchmark (OLS with no spatial fixed effects) have poor accuracy. In addition, the ScaGWR and GAM procedures provide more accurate estimation than OLS with spatial fixed effects. Both procedure are close, either for mean or median relative errors. The dispersion is however more important for the ScaGWR.

Table A.6: Distribution of Relative Errors According to the Estimation Procedure

Method	Mean	Median	Std Dev	Quantile 1	Quantile 3	Decile 1	Decile 9
Benchmark	25.44	20.46	20.59	9.54	35.95	3.76	54.71
OLS FE Municipality	19.29	15.51	16.17	7.50	26.17	3.02	39.72
OLS FE IRIS	18.05	14.18	15.76	6.66	24.64	2.62	37.69
ScaGWR	16.59	12.32	15.63	5.64	22.33	2.20	35.81
GAM	16.57	12.76	14.83	6.08	22.41	2.42	35.03

Notes: We report main descriptive statistics resulting from our comparison procedure. Our outcome of interest is the relative error for a prediction. We use similar dataset for each estimation procedure under consideration. The benchmark corresponds to an OLS estimation without spatial fixed effects. The ScaGWR is estimated using the `scgwr` R package, while we use the `mgcv` R package to implement the GAM procedure. Our results stem from 198,995 individuals estimation from 50 iterations. Our perimeter of interest is the *Côte d'Or* department, with transactions from 2010 to 2020.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

Spatial Heterogeneity We also assess the accuracy heterogeneity of each method according to the local characteristics. We exploit the urban areas providing by the INSEE. We qualify municipality within each urban areas according to the distance from the centre. In addition, rural areas are clustered.

Table A.7: Distribution of Relative Errors According to the Estimation Procedure and Housing Locations

Method	Center	1st Ring	2nd Ring	3rd Ring	Rural
Benchmark	26.64	23.16	24.31	27.33	30.40
OLS FE Municipality	19.08	17.67	19.93	21.99	22.02
OLS FE IRIS	17.39	15.86	19.84	21.88	22.02
ScaGWR	14.88	13.76	20.00	22.67	24.47
GAM	15.45	14.87	18.66	20.56	21.31

Notes: We report the average relative error according to the municipality characteristics. Our outcome of interest is the relative error for a prediction. We use similar dataset for each estimation procedure under consideration. The benchmark corresponds to an OLS estimation without spatial fixed effects. The ScaGWR is estimated using the `scgwr` R package, while we use the `mgcv` R package to implement the GAM procedure. Our results stem from 198,995 individuals estimation from 50 iterations. Our perimeter of interest is the *Côte d'Or* department, with transactions from 2010 to 2020.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

Although the GAM and the GWR are close when comparing the entire *Côte d'Or* department, they both recover spatial heterogeneities in their accuracy. Indeed, the accuracy decreases with distance from the CBD. Nonetheless, the GAM approach provides more consistent results according to transaction location, in line with the lower dispersion observed on general results. The GWR performs poorly for rural areas and third ring of urban areas, as OLS with spatial fixed effects provide better results. Finally, computational times are sizeable for the GWR, while the GAM procedure is ten times faster. Hence, we exploit the GAM procedure to implement the imputation model.

A.7 Relationship for 1D Variables with Unitary Housing Price

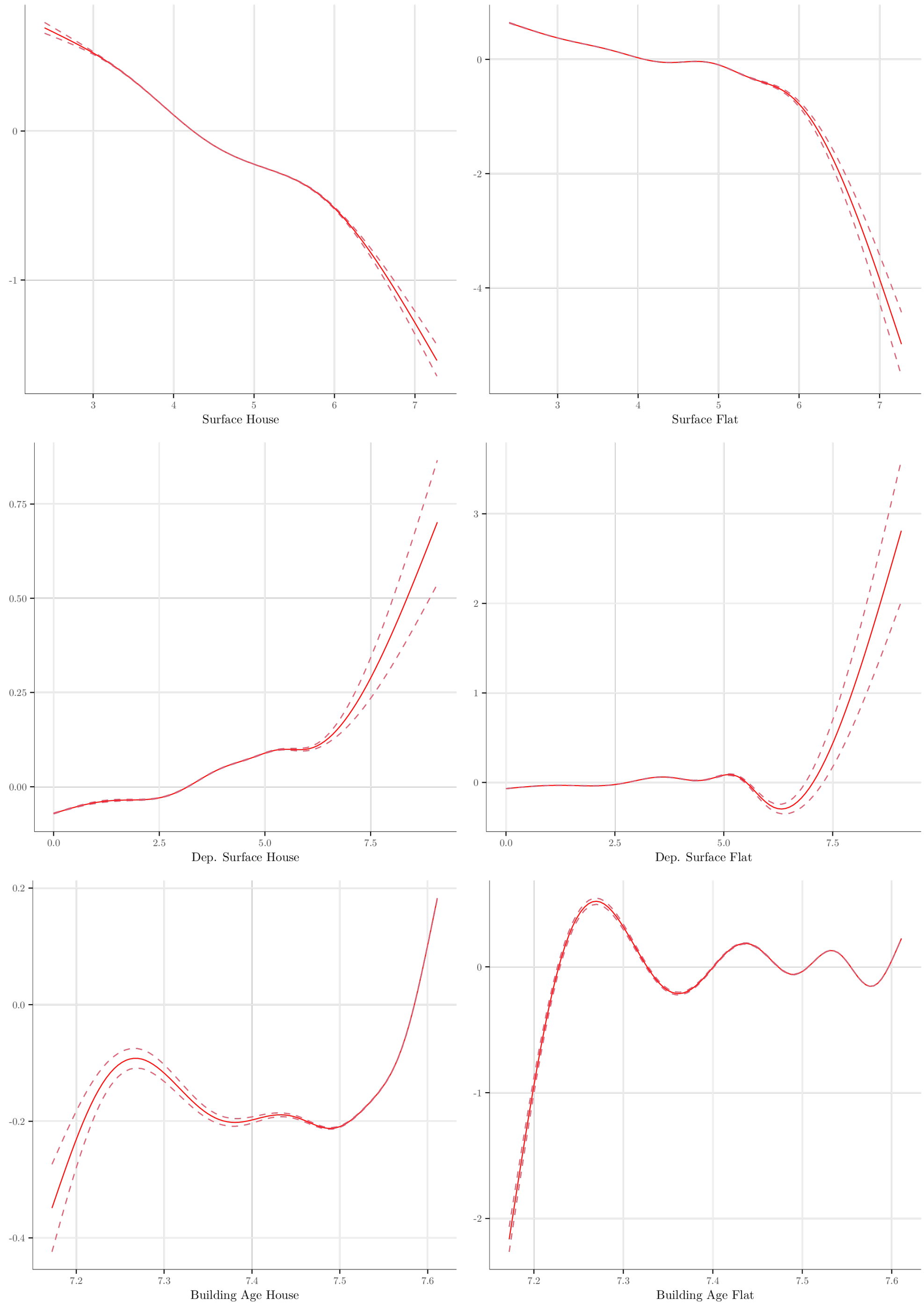


Figure A.2: Relationship for 1D Variables with Unitary Housing Price

Notes: We report the marginal contribution for continuous variables introduced in the housing value imputation model. It stems from the estimation of pooled national model with spatial coordinates using bivariate spline transformation (non reported). The dependent variable is the unitary housing price. We distinguish marginal contribution according to whether it concerns collective or individual housing. The estimation is performed using the `mgcv` R package.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

A.8 Joint Significance for Country Models to Explain Price and Unitary Housing Price

Table A.8: Joint Significance of Covariates in Pooled Models for Auto Valuation Method

	Unit. Price			Price		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Spatial Coordinates		12,149***	8,716***		2,825***	8,663***
		[786.8]	[733.3]		[768.3]	[744.0]
Ind. Housing						
Surface	28,446***		50,344***	145,757***		258,894***
	[7.0]		[6.9]	[7.0]		[7.0]
Age	53,358***		54,191***	53,363***		54,998***
	[7.0]		[7.0]	[7.0]		[7.0]
Dep. Surface	1,069***		7,858***	1,069***		7,668***
	[6.9]		[6.9]	[6.9]		[6.9]
Coll. Housing						
Surface	51,909***		45,529***	157,603***		386,225***
	[7.0]		[7.0]	[7.0]		[7.0]
Age	40,300***		53,235***	40,289***		53,282***
	[7.0]		[7.0]	[7.0]		[7.0]
Dep. Surface	10,510***		2,907***	10,510***		2,951***
	[7.0]		[6.9]	[7.0]		[6.9]
N	7,176,945	7,176,945	7,176,945	7,176,945	7,176,945	7,176,945
McFadden R2	31.44	57.47	65.19	30.90	23.91	64.80
AIC	10,349,171	6,924,063	5,486,871	10,349,089	11,041,679	5,508,224

Notes: We report joint significance estimated using χ^2 for pooled appraisal models for unitary housing price (columns 1 to 3) and overall price (columns 4 to 6). Our pooled model is estimated using endogenous spline transformations for covariates, with GAM. The effective degree of freedom for variable transformations is reported in brackets. Our spline transformation differs between individual and collective housings. The bottom panel reports metrics about regression, including the number of transactions (N), the McFadden R^2 and Aikike Criterion Information.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.
*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$

A.9 Joint Significance for Commuting Areas Models to Explain Unitary Housing Price

Table A.9: Results from Joint Significance of Smoothing Splines Transformation

Covariates	Mean	Q1	Q3	Min	Max
Ind. Housing					
Surface	577.8	215.0	679.8	13.6	7,163.1
	[4.7]	[4.2]	[5.3]	[1.3]	[6.8]
	292/292				
Dep. Surface	164.1	43.2	197.1	8.2	1,644.4
	[4.3]	[3.7]	[5.0]	[1.4]	[6.7]
	292/292				
Building Age	332.9	121.5	394.9	8.0	3,256.9
	[6.3]	[6.2]	[6.7]	[2.1]	[6.9]
	292/292				
Coll. Housing					
Surface	301.3	26.5	219.8	1.7	5,054.2
	[3.5]	[2.1]	[4.8]	[1.0]	[6.9]
	287/292				
Dep. Surface	54.8	3.4	46.0	0.0	955.0
	[3.0]	[1.7]	[4.1]	[1.0]	[6.2]
	227/292				
Building Age	582.7	11.5	322.0	0.0	15,422.0
	[5.2]	[4.1]	[6.7]	[1.0]	[7.0]
	264/292				
Spatial Coordinates	33.6	1.2	18.4	0.1	2,162.8
	[199.2]	[118.8]	[268.7]	[17.8]	[731.3]
	292/292				
R^2	0.42	0.32	0.52	0.18	0.79
N	32,938	7,923	31,298	1,007	910,207
AIC	-51	1,020	4,523	-158,996	21,596

Notes: We report the average joint significance, the average effective degree of freedom and the number of joint significance for our independent mass appraisal models. Unit of observations is the commuting zone. We also report in the bottom panel average statistics about regression (average). Finally, we provide in columns the distribution of statistics for each covariate introduced in imputation model.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

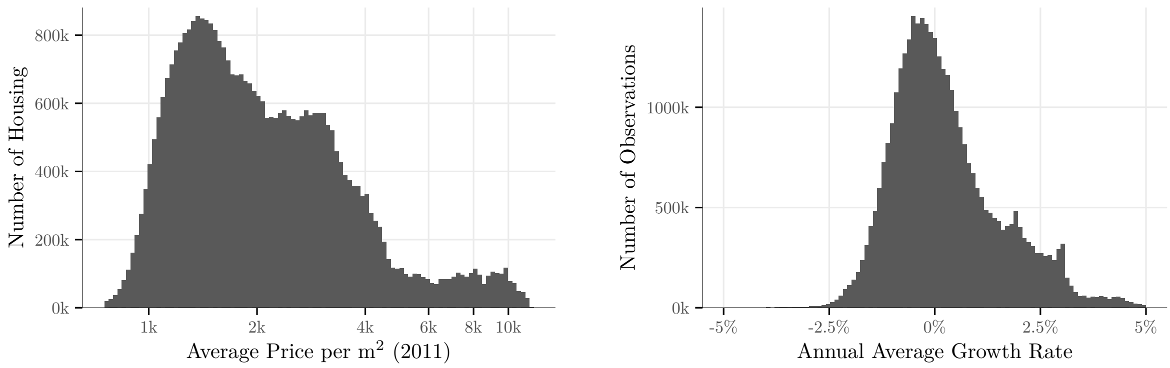
A.10 Descriptive Statistics about Imputed Housing Value

Table A.10: Descriptive Statistics for Housing Value Imputation

Year	N	Mean	Median	Std Dev	Quantile 1	Quantile 3	Min	Max
2011	33,105,876	166,583	138,480	112,829	94,784	203,126	751	10,567,560
2012	33,458,690	168,337	138,750	116,966	94,612	204,768	1,705	10,388,280
2013	33,810,212	167,049	137,390	116,928	93,568	203,040	1,705	10,213,980
2014	34,130,178	164,550	135,548	115,081	92,308	200,016	1,716	42,460,440
2015	34,423,797	163,714	134,910	114,549	91,770	199,064	1,705	37,254,888
2016	34,784,340	165,309	135,892	116,801	92,106	201,110	1,716	34,046,808
2017	35,069,421	168,671	138,208	120,309	93,360	205,248	1,738	31,275,807
2018	35,355,911	172,675	141,120	123,925	94,966	210,315	1,635	30,342,829
2019	35,650,295	177,126	144,432	127,717	96,768	216,087	715	31,036,918

Notes: We report the distribution of imputed housing value through our procedure. It exploits the `mgcv` R package. Imputed values concern the metropolitan France, except from the Alsace-Moselle.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

A.11 Distribution of Imputed Housing Value (Cross-sectional and Evolution)



(a) Housing Prices per m^2 in 2019 (b) Annual Average Growth Rate

Notes: We report the distribution of appraised unitary housing price in 2019 (left panel) and average annual growth rate between 2011 and 2019 (right panel). Our observation unit arises from 200×200 square grid provided by INSEE to provide demographic files.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

A.12 Imputed Housing Value According to Locations (Cross-sectional and Evolution)

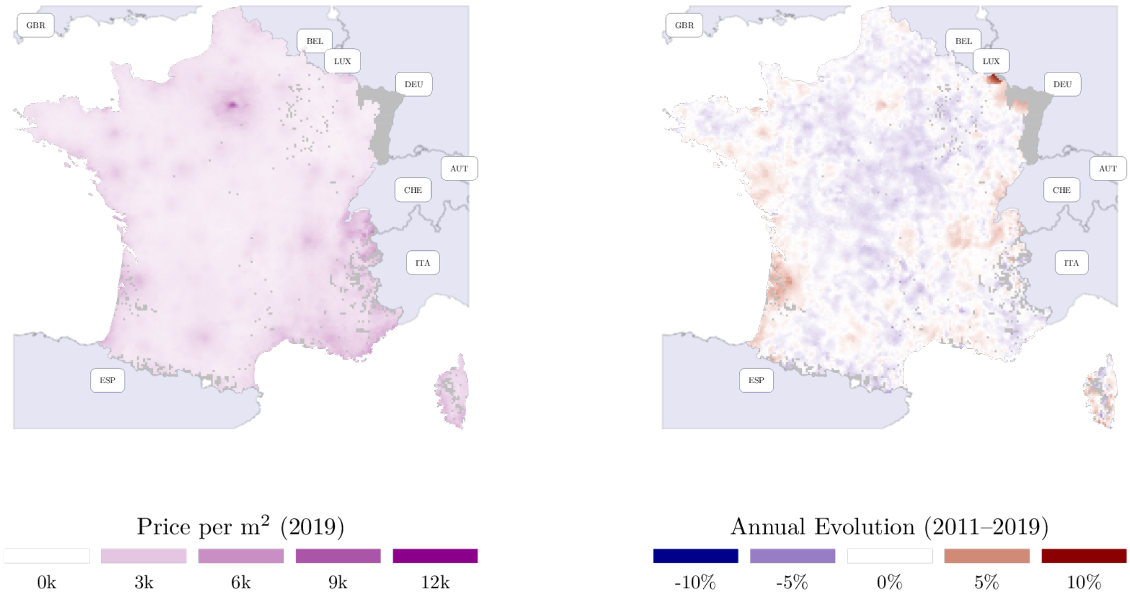


Figure A.4: Descriptive Statistics about Appraised Values According to Location

Notes: We report on the left panel average unitary housing price using 200 × 200 meters square grid. The right panel represents the average annual evolution for unitary housing price between 2011 and 2019.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

A.13 Transfers

To identify transfers, we exploit the exhaustive character of the DV3F registering all transactions for pecuniary purposes, combined with the panel data about housing we construct previously. We indeed identify all housing being concerned by change of owners every year. As both dataset provides unique ID about housing (from fiscal sources), we can merge both dataset and then filter on the change year. That way, we can distinguish change of owners occurring with transfers from purchase. Furthermore, it enables us to distinguish partial changes of owners (*e.g.* consecutive to a divorce widowhood) from full ones.

Our method, to be valid, requires the missing rate, defined as the share of transaction for pecuniary interest associated to no change of owners, to be as low as possible. We then run the merging process as defined previously and estimate the missing rate. Remark that we do not restrict our datasets (for instance remove outliers). However our identification might underestimate the number of transfers as we cannot observe within year change based on the panel data. Hence, if a dwelling is transferred and consecutively sold within the same year, our method qualifies this change as purchase, as we match the change of owners with a transaction for pecuniary purposes. Then, it limits the possibility to exploit these data to study the decision to sell after receiving a transfer.

We report on Table A.11 the missing rate based on the simple merging process between owners change (derived from panel data we construct) and the transaction dataset (derived from the DV3F database). We chose to characterise according to the previous type of owners and the following one, using the moral – natural person distinction. Indeed, as we mainly focus on natural persons, we are more interested to lower the missing rate for natural – natural transactions as they are more likely to be of interest for transfers.

Table A.11: Missing Rate According to Owner Type and Temporal Assumption

	Legal	Natural
Legal	26.40	14.61
Natural	15.17	11.96

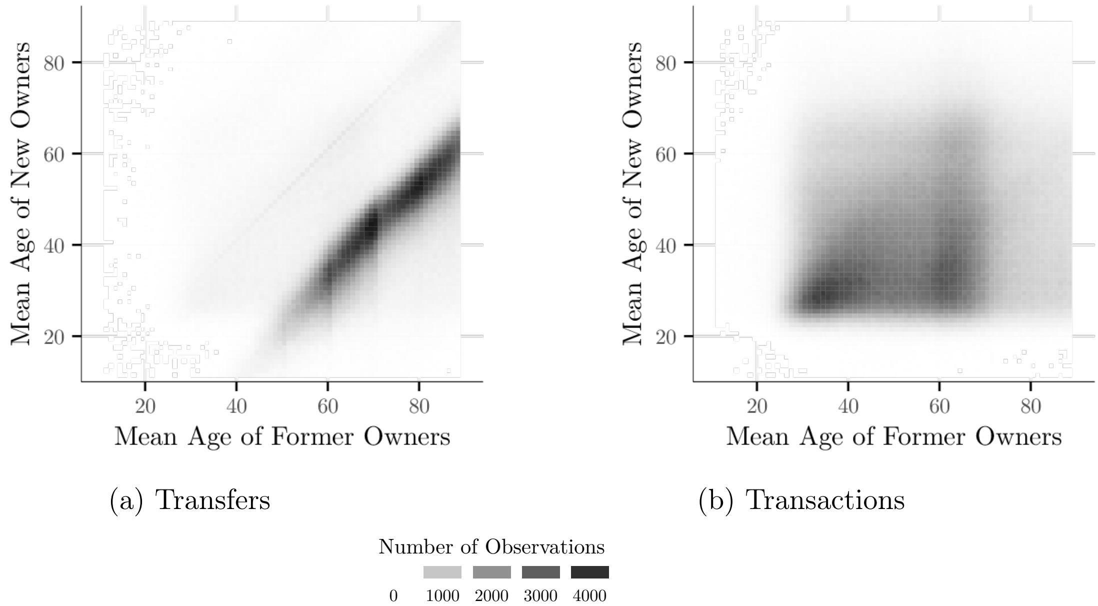
(a) Restricted Temporal Assumption

	Legal	Natural
Legal	9.10	2.03
Natural	3.75	0.85

(b) Loosen Temporal Assumption

Notes: We report the missing rate, defined as the ratio between matched transactions in DV3F and overall transactions in DV3F. Our studied period is 2011–2019. The column represents a type of sellers while the row specification represents purchasers. The left panel provide results with greater restriction for the temporal criterion, while the left panel relaxes the assumption (with one additional year).
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

A.14 Age Profile for Housing Changes



(a) Transfers

(b) Transactions

Figure A.5: Age of New and Former Owners According to Change Type

Notes: We report the number of observations according to the mean age of both former (x-axis) and new owners (y-axis). We distinguish between transfer transaction (figure a) and purchase (figure b). We pool all year from 2012 to 2018. We remove observations with missing observation about the mean age of either former or new owners. *Sources:* Authors' calculation based on DV3F and *Fichiers Fonciers*.

B Additional Material

B.1 Partial Correlation Between Housing Wealth Variation, Initial Level of Wealth and Location of Main Residence

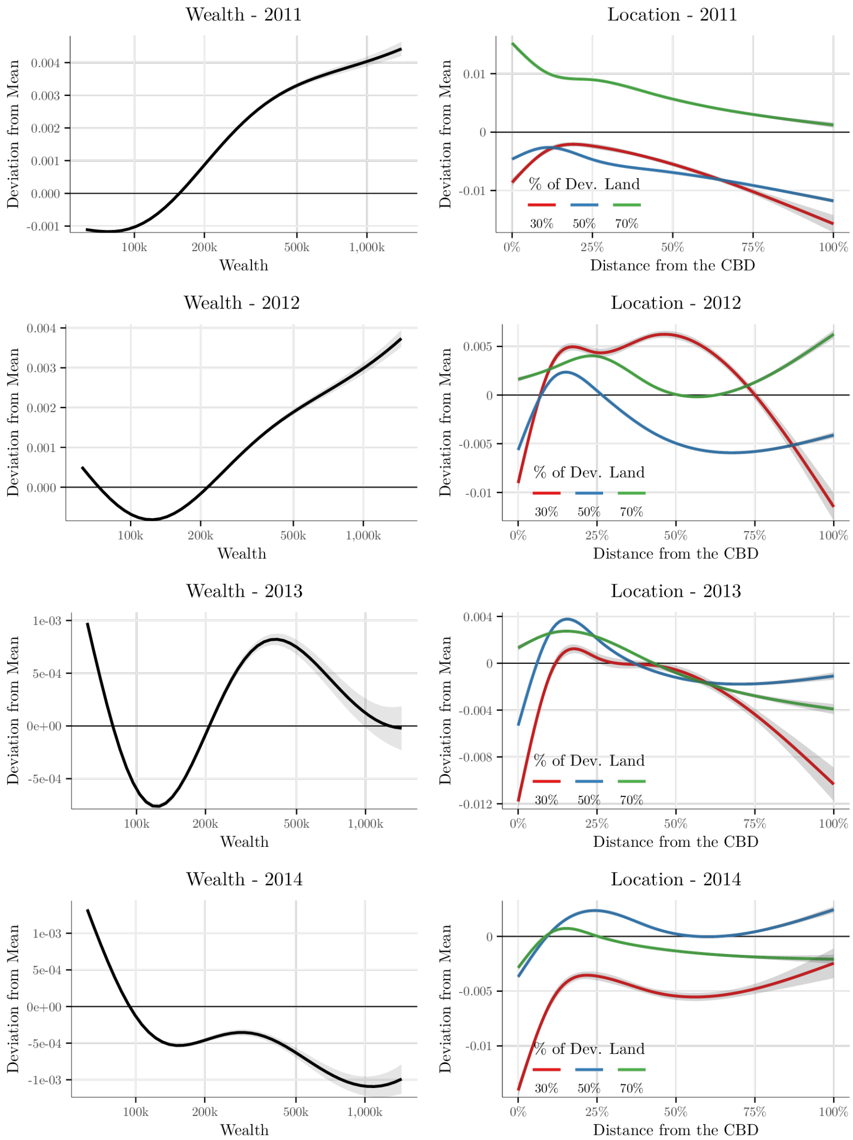


Figure B.1: Partial Correlation Between Housing Wealth Variation, Initial Level of Wealth and Location of Main Residence (2011-2014)

Notes: We report partial correlation with wealth evolution according to the previous of housing wealth (left panel) and locations (right panel). These results stems from the estimation of Equation 3. Our observation unit is distinct owners that does not experience portfolio changes in 2018. We report confidence interval to a 95% confidence interval. We use the `mgcv` R package.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

B.2 Partial Correlation Between Housing Wealth Variation, Initial Level of Wealth and Location of Main Residence

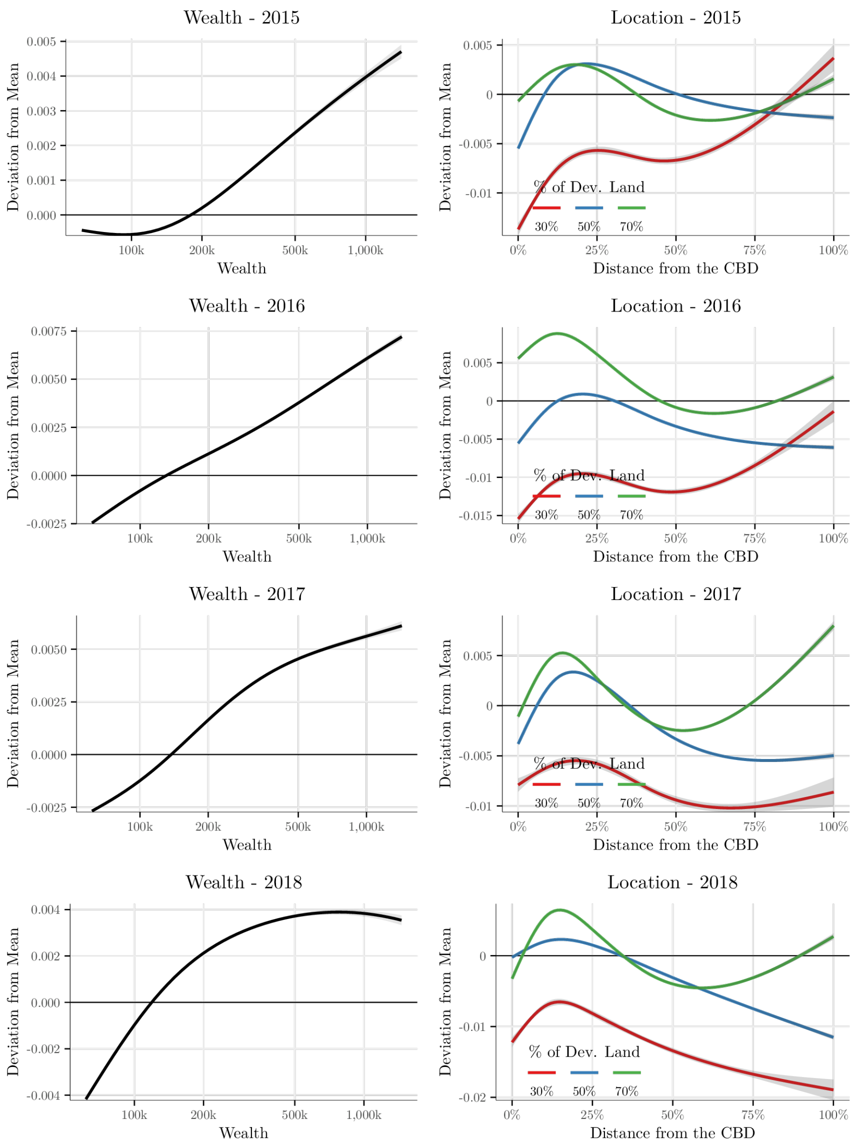


Figure B.2: Partial Correlation Between Housing Wealth Variation, Initial Level of Wealth and Location of Main Residence (2015-2018)

Notes: We report partial correlation with wealth evolution according to the previous of housing wealth (left panel) and locations (right panel). These results stems from the estimation of Equation 3. Our observation unit is distinct owners that does not experience portfolio changes in 2018. We report confidence interval to a 95% confidence interval. We use the `mgcv` R package.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

B.3 Partial Correlation of Initial Level of Wealth with Housing Decisions within Commuting Zones for Owner-Occupiers

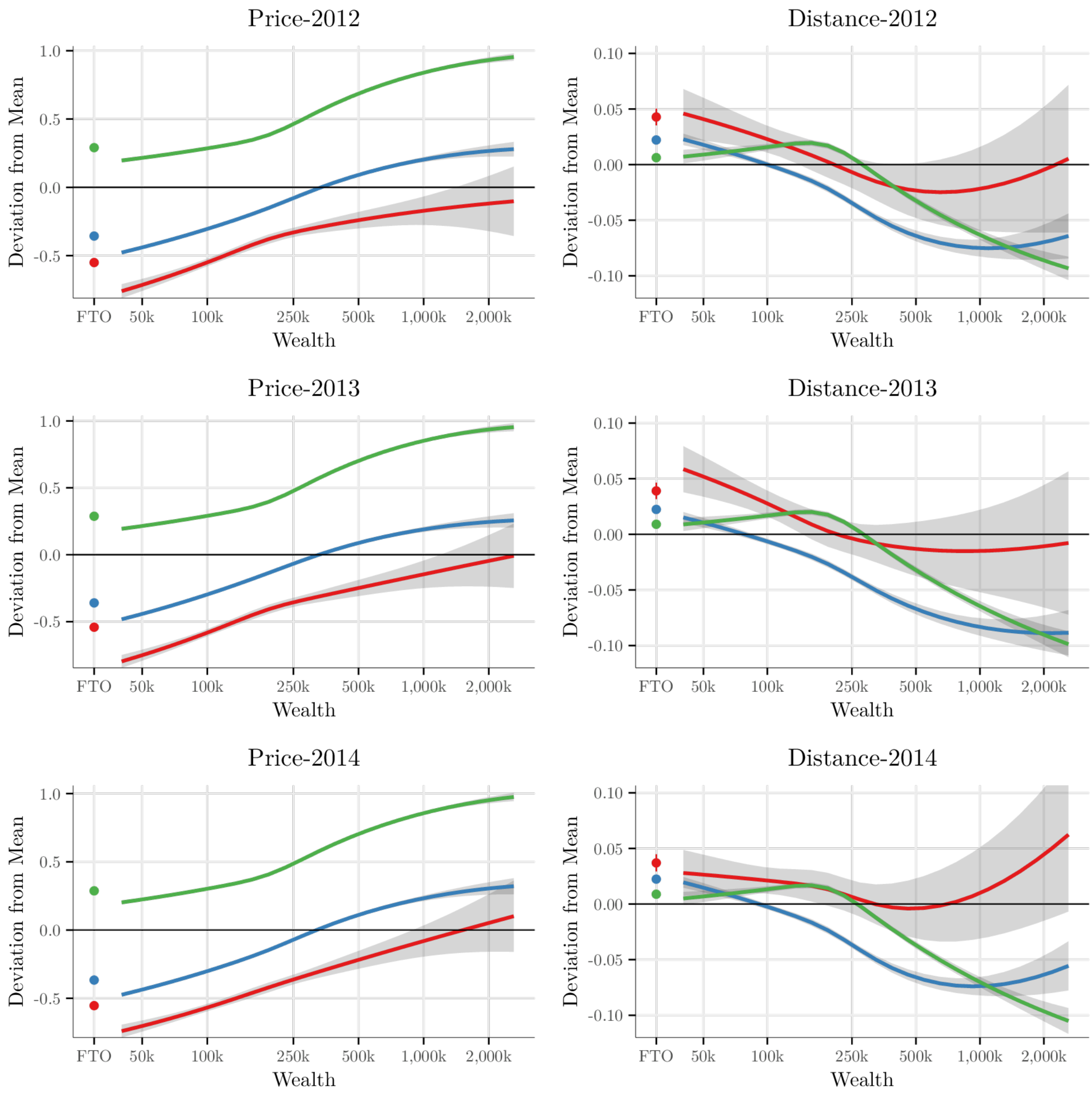


Figure B.3: Correlation between Previous Wealth, Location Choices and Transaction Price for Owner-Occupiers (2012-2014)

Notes: We report the partial correlation for the interaction between the economic status of the commuting zone and the initial wealth level of the buyer from Equation 4. The left panel corresponds to the transaction price, while the right panel corresponds to the distance from the CBD. We introduce variation in commuting zone characteristics as we estimate the wealth correlation for commuting with 30%, 50% and 70% of built-up area in 2000. We also introduce first-time homeowners (dot, left side of plot). We use the `mgcv` R package to implement additive transformations of variable with endogenous definition of degree of freedom.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

B.4 Partial Correlation of Initial Level of Wealth with Housing Decisions within Commuting Zones for Owner-Occupiers (Bis)

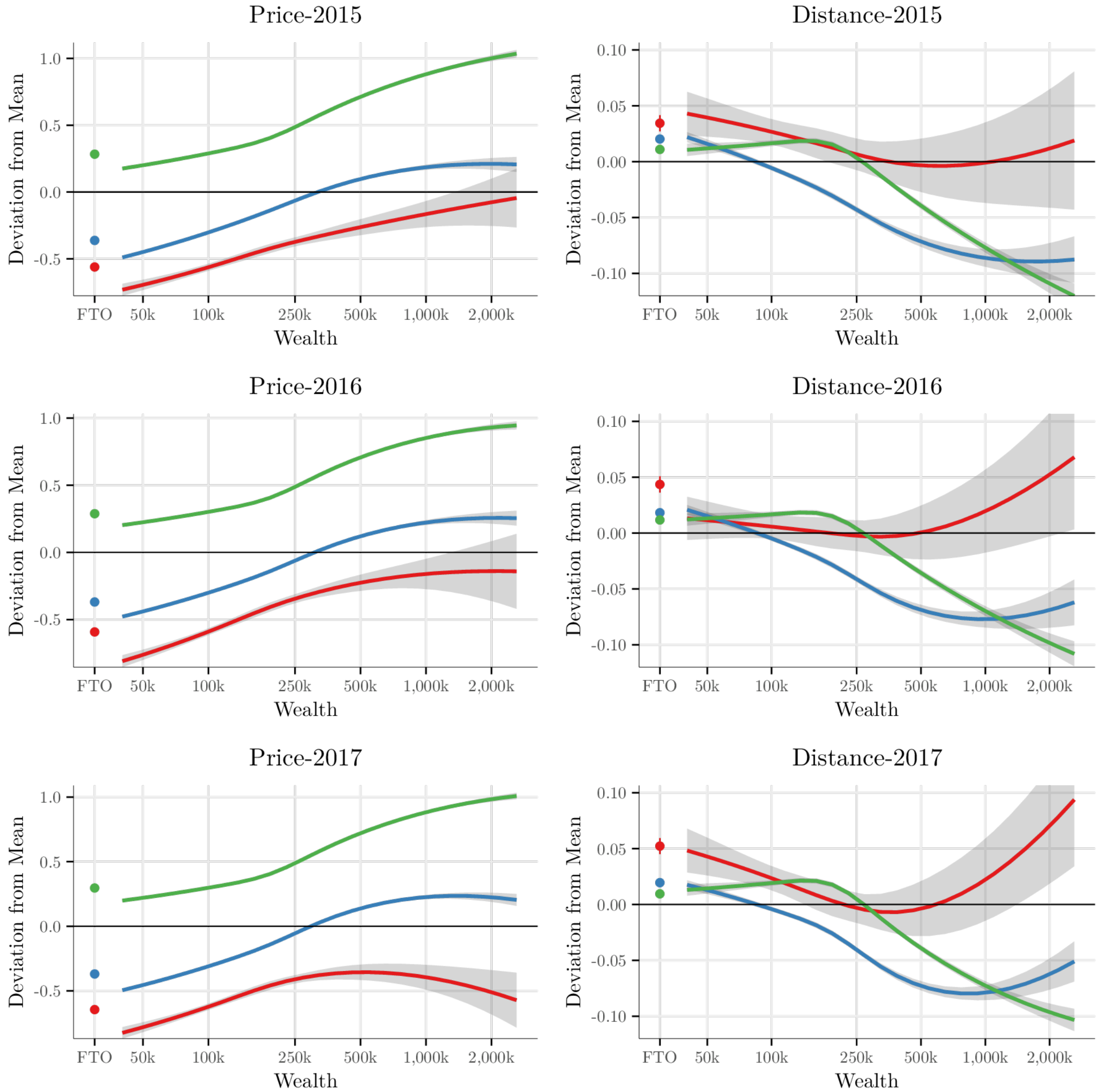


Figure B.4: Correlation between Previous Wealth, Location Choices and Transaction Price for Owner-Occupiers (2015-2017)

Notes: We report the partial correlation for the interaction between the economic status of the commuting zone and the initial wealth level of the buyer from Equation 4. The left panel corresponds to the transaction price, while the right panel corresponds to the distance from the CBD. We introduce variation in commuting zone characteristics as we estimate the wealth correlation for commuting with 30%, 50% and 70% of built-up area in 2000. We also introduce first-time homeowners (dot, left side of plot). We use the `mgcv` R package to implement additive transformations of variable with endogenous definition of degree of freedom.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

B.5 Mean Housing Transfers per Year

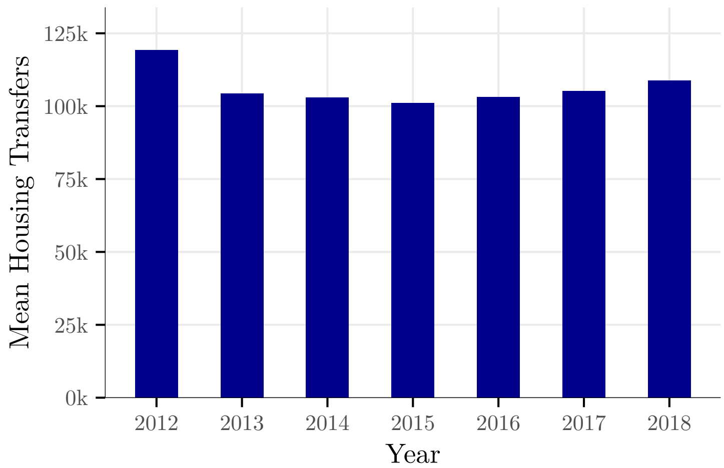


Figure B.5: Mean Housing Transfers per Year per Individuals

Notes: We report the estimation of the average housing transfers per year. The housing price is estimated using our housing market value imputation. If the same housing is split between recipients, we assign equal share of housing value to recipients. Transfers are restricted to housing one.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

B.6 Partial Correlation of Initial Level of Wealth and Residence Place with Housing Transfers

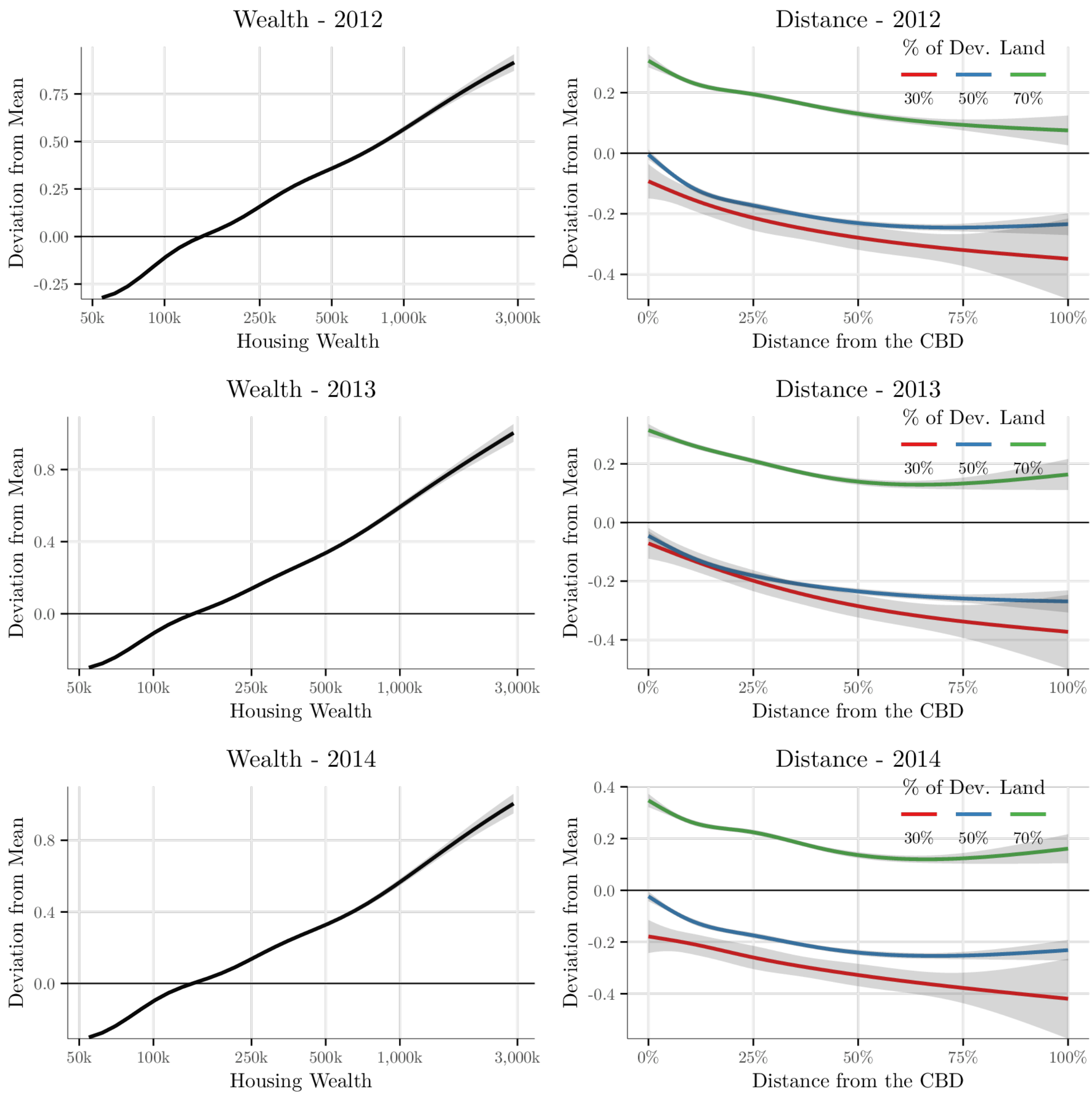


Figure B.6: Partial Correlation of Initial Level of Wealth and Residence Place with Housing Transfers (2012–2014)

Notes: Our results are derived from estimation of Equation 5 restricted to housing transfers recipients. The left panel reports marginal contribution according to previous current wealth. The right panel reports spatial heterogeneity based on recipients location. We report pooled results over year for the spatial contribution. We report confidence interval at a 90% level. We use the `mgcv` R package.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.

B.7 Partial Correlation of Initial Level of Wealth and Residence Place with Housing Transfers (Bis)

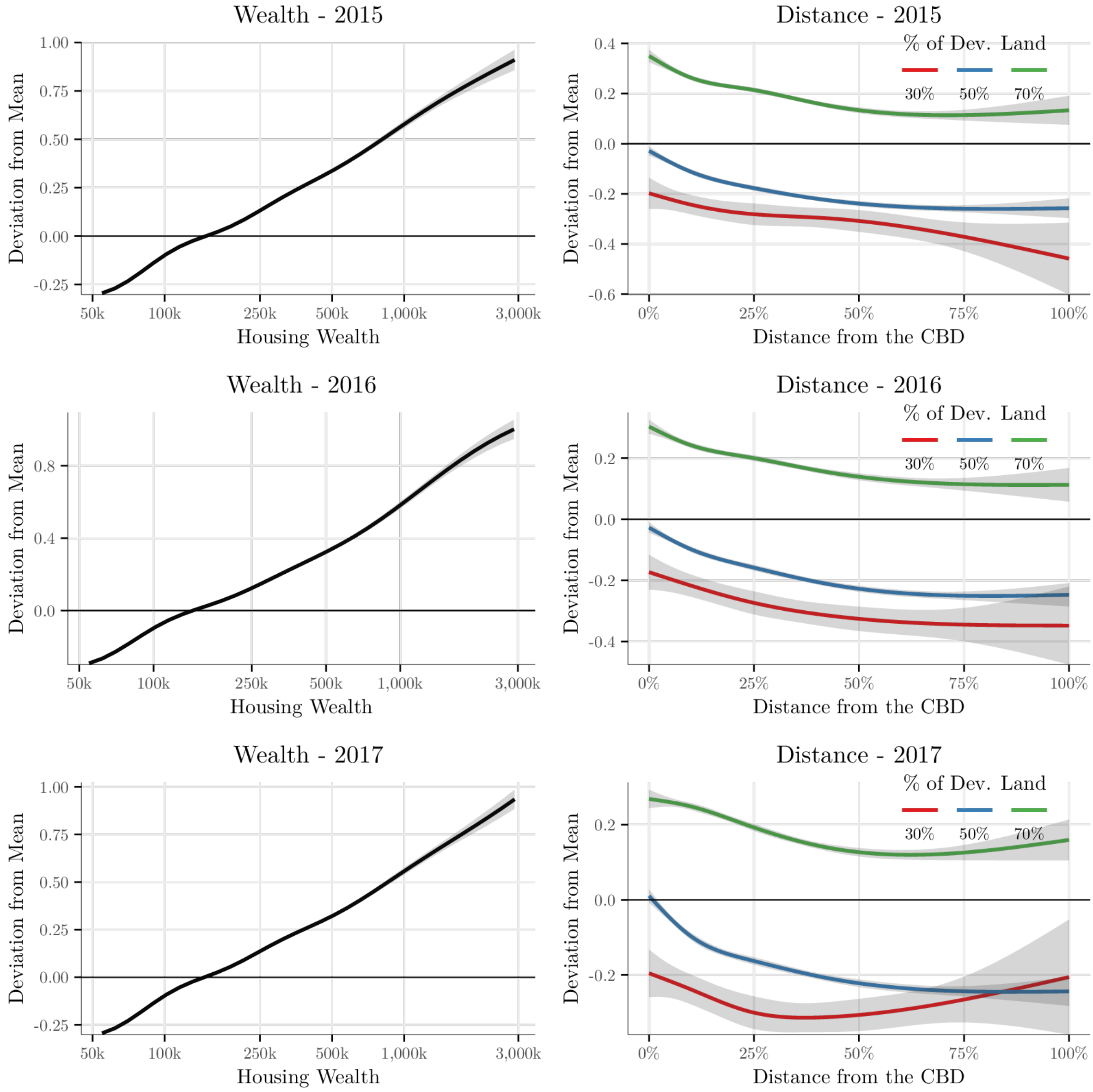


Figure B.7: Partial Correlation of Initial Level of Wealth and Residence Place with Housing Transfers (2015–2017)

Notes: Our results are derived from estimation of Equation 5 restricted to housing transfers recipients. The left panel reports marginal contribution according to previous current wealth. The right panel reports spatial heterogeneity based on recipients location. We report pooled results over year for the spatial contribution. We report confidence interval at a 90% level. We use the `mgcv` R package.
Sources: Authors' calculation based on DV3F and *Fichiers Fonciers*.