UNDERSTANDING REAL ESTATE INVESTMENTS THROUGH BIG DATA GOGGLES

A GRANULAR APPROACH ON INITIAL YIELDS

MARCELO CAJIAS

ABSTRACT:

Initial yields are used by institutional investors and investment managers to assess the pricing conditions of real estate markets. In contrast to commercial real estate, initial yields in the residential sector are hard to quantify, especially due to the lack of comparables. In the era of digitalisation and big data residential assets are mostly brought to the market via digital multiple listing systems. The paper develops semiparametric hedonic models for extracting the implicit information to calculate residential net initial yields for both a buy-to-hold and rental investment strategy based on more than 3 million observations. The results are robust and confirm that the pricing conditions of residential markets are captured by the hedonic approach, enhancing the transparency in real estate markets.

KEYWORDS:

Net initial yields, semiparametric regression, big data, German residential, buy or rent.

ACKNOWLEDGEMENT

The author especially thanks PATRIZIA Immobilien AG for contributing the dataset to conduct this study. All statements of opinion reflect the current estimations of the author and do not necessarily reflect the opinion of PATRIZIA Immobilien AG or its associated companies.

CORRESPONDING AUTHOR

Dr. Marcelo Cajias
Fuggerstraße 26, 86150 Augsburg DE
marcelo.cajias@patrizia.ag
+49 176 80293 102

UNDERSTANDING REAL ESTATE INVESTMENTS THROUGH BIG DATA GOGGLES

A GRANULAR APPROACH ON INITIAL YIELDS

ABSTRACT:

Initial yields are used by institutional investors and investment managers to assess the pricing conditions of real estate markets. In contrast to commercial real estate, initial yields in the residential sector are hard to quantify, especially due to the lack of comparables. In the era of digitalisation and big data residential assets are mostly brought to the market via digital multiple listing systems. The paper develops semiparametric hedonic models for extracting the implicit information to calculate residential net initial yields for both a buy-to-hold and rental investment strategy based on more than 3 million observations. The results are robust and confirm that the pricing conditions of residential markets are captured by the hedonic approach, enhancing the transparency in real estate markets.

UNDERSTANDING REAL ESTATE INVESTMENTS THROUGH BIG DATA GOGGLES

A GRANULAR APPROACH ON INITIAL YIELDS

1. INTRODUCTION

Net initial yields area a common indicator used by institutional investors and investment managers to assess the pricing conditions of real estate markets. Net initial yields (NIY) measure the net rental income of a specific property relative to the gross purchase price plus transactions costs. They are a very popular indicator in real estate investment markets worldwide, partly because they express the initial pricing level of real estate, but especially because they are easy to calculate and interpret: the lower the NIY the more attractive is the market and higher prices will result. Nowadays, the assessment of markets via NIYs works pretty well when targeting commercial real estate markets as these are provided by several data provider and international brokers such as RCA, JLL or CBRE among others. But if something is for sure in the real estate industry is that residential NIYs are more than scarce.

The reasons why the residential market lacks standardized NIYs lie in the nature of the asset class. Firstly, residential markets lack of a central data collection system that gathers both transaction prices and contract rents, as neither brokers, nor institutional players, nor private owners of residential assets are obliged to disclose this information. Secondly, the users are to a great extent unsophisticated private persons interested in either renting or buying a property; an issue that complicates the calculation of a reliable NIYs as investment comparables are scarce and the holding period of residential assets is relatively long. Finally, in many countries information on residential transactions are held by governmental property valuation committees – mostly regionally organized – and not publicly or easily accessible for investors or researchers.

This situation raises the question whether an alternative measure for assessing the pricing conditions of residential markets is needed or whether we need to look for a novel approach that approximates the price-rent-relationship. This is where the paper aims at focussing on. In the era of digitalisation and big data residential assets are mostly brought to the market via digital multiple listing systems (MLS) in order to reach a broader range of potential users and increase their liquidity in contrast to traditional channels such as newspapers. MLS have evolved continuously over time and have contributed substantially to a harmonization of the hedonic characteristics of residential assets. That means that any user aiming at bringing its asset into the market for selling or letting via a MLS has to comply with minimum requirements in terms of hedonic characteristics which leads to a harmonization of data basis. Thus, this paper aims at applying a semiparametric hedonic approach for approximating residential NIYs by using more than 3 million big data observations from MLS.

The theoretical foundations for NIYs are simple: the price (P) of any (residential) asset equals the present value of the rental income (R) minus expenses at a specific discount rate (i). Following the Gordon growth model in perpetuum the present value of a residential asset can be easily simplified as:

$$P = \frac{R}{i} \tag{1}$$

Although the Gordon growth approximation is anything but accurate in estimating the future performance of capital expenditures of repositioning or conversion strategies of a specific asset, it provides a initial signal about the relative asset performance. In real estate finance NIYs would be estimated in simple terms as:

$$i = \frac{R}{P} \tag{2}$$

In contrast to the theoretical approach, the empirical approach of estimating NIYs is much more complicated and this time not because location, location and location plays an essential

role, but because real estate assets can't be sold or rented at the same time. Thus, in order to fulfil minimum statistical requirements, the NIY calculation needs to include sufficient data in the numerator and in the denominator of equation (2). Otherwise, the NIYs would tend to be skewed and biased and would mimic a wrong market pricing. In this context, the German residential market offers a unique experimental scenario with the second largest private rental market in Europe after Switzerland. At the same time, the size and quality of the data employed for the empirical analysis is expected to reduce the bias and skewness in the NIY approximation.

In general, research focussing on real estate initial yields focusses on explaining their behaviour and incorporating them in an applied context, rather than focussing on the formation of the price-rent-ratio. For example, Hutchison, et al. (2011) focusses on gathering best practices regarding the appropriate benchmark risk free rate investors use when estimating real estate yield gaps in the UK. In contrast, Jones, Dunse, Cutsforth (2015) go a step further and analyse the relationship between property yields and government bond yields in Australia, the US and the UK and conclude that the structural break during the GFC has led to strong variations in the yield gap. Instead, we take several steps back and focus on the approach for estimating NIY.

The investigation of the formation of the price-rent-ratio has not received substantial attention in real estate research over the last decade and remains elusive and partly hard to quantify. This might be less ascribed to the theoretical inconsistency in the estimation of initial yields, but especially to the lack of adequate data capturing rents and prices consistently and simultaneously. As one of the first papers in this area, Linneman and Voith, 1991 describe the methodological challenges when estimating NIYs and confirm that a simple OLS-approach works well whenever the sample is not biased towards different asset types (e.g. houses and dwellings) and the model controls for households' characteristics such as income. Therefore,

we focus in the paper on estimating NIYs for only dwellings including households' characteristics in the parameterization of the hedonic regressions.

The current paper builds upon the theoretical approach and the empirical results of Lisi, 2015 and Hill, Iqbal and Syed, 2016. Although Lisi argues that the estimation of NIY (in his case capitalization rates) is achievable without any information from the rental side, his approach in novel since he derives NIYs reweighting the value of each hedonic characteristic from transaction and implicit prices by univariate R^2s. In other words, NIYs are measured as the aggregated distance between the true and the implicit willingness to pay for each single hedonic attribute. Lisi argues that transaction prices reflect the overall value appreciation of the asset and that rental prices are irrelevant when estimating NIYs. Although their results show a concordance with NIYs provided by brokers, our approach makes usage of the information held in rents allowing deeper insights on the data generating process of the price-rent-relationship. In contrast, our approach follows the empirical strategy of Hill, Iqbal and Syed with some modifications in the model parameterization. While they focus on the estimation of the price-rent-relationship in Australia based on a similar data basis with 730,000 observations, they put a lot of emphasis on measuring when prices and rents reach intertemporal disequilibrium. Instead of looking at the estimation of NIYs, their empirical approach provides strong evidence that the extraction of implicit prices and rents from hedonic models – either via single imputation, time dummies, quantile or quality adjusted methods – is valid and able to capture market pricing conditions accurately.

Although the estimated residential NIYs in this paper track initial asking prices and initial rents rather than contract prices and contract rents, they serve as an important indicator of the current pricing level of residential real estate across both cross-sections, time and space. In contrast to other European countries, the size of the rental market in Germany is large and most of the market activity takes place via ML services as a traditional marketing channel for

both landlords and tenants. In contrast to registry or mortgage approval databases, the advantage of MLS databases – such as the empirica database in our case – relies on the fast access to real data when estimating hedonic models, see Cajias (2018). The deviation to contract rents and prices is therefore not expected to lead to an error bias, especially after controlling for full hedonic characteristics, see Schimizu, Nishimur and Watanabe (2012), Lyons (2013) for examples in Japan and Ireland respectively.

The empirical results illustrate on the one hand that residential prices and rents across 161 German residential markets have increased continuously over the last years after controlling for hedonic, spatial and temporal variables. Consequently, NIYs have reached record low levels, whereas the premium to risk-free government bonds has risen dramatically. On the other hand, the results confirm the reliability of the employed approach as the economic spatial hierarchy of German residential markets is fulfilled, e.g. the attractiveness and economic relevance of Munich is expected to lead to lower NIYs in comparison to Bonn, the former capital city of Germany.

The next section explains the methodology followed by the data description. Afterwards, the results are presented from three different perspectives covering the cross-sectional variation, the development over time and potential regional clusters. Finally, the paper summarizes the main findings, pitfalls and possible extensions.

2. ECONOMETRIC APPROACH FOR ESTIMATING NET INITIAL YIELDS

The markets to be modelled comprise of dwellings for rent and sell rather than single family houses, student housing or multifamily assets. Let $\tilde{P}_{ijmt}$ and $\tilde{R}_{ijmt}$ represent the log asking prices and log asking rents of dwellings i, in time t, in ZIP code j and market m. Each dwelling has continuous and binary hedonic characteristics in the matrix X^{P}_{ijmt} and X^{R}_{ijmt}. ZIP-specific covariates are included in $\tilde{Z}_j$, time-fixed-effects are controlled via $\mu_t^{\tilde{P}}$ and $\mu_t^{\tilde{R}}$ and ir_t controls for the financial conditions. The models include quarterly dummies to control for

the unobserved serial error component and they capture the quarter of the entry of any dwelling offered in the MLS starting in 2013-Q2 until 2018-Q1, i.e. $t-1=20$ dummy variables. The hedonic equations are calculated for each m separately and include non-linear variables via $f(\dot{x}_{ijmt}^{\tilde{P}})$ and $f(\dot{x}_{ijmt}^{\tilde{R}})$ parameterized as:

$$\begin{Bmatrix} \tilde{P}_{ijmt} = X_{ijmt}^{\tilde{P}}\beta^{\tilde{P}} + \tilde{Z}_j\delta^{\tilde{P}} + \mu_t^{\tilde{P}}\theta_t^{\tilde{P}} + ir_t\gamma^{\tilde{P}} + f(\dot{x}_{ijmt}^{\tilde{P}}) + \varepsilon_{ijmt} \\ \tilde{R}_{ijmt} = X_{ijmt}^{\tilde{R}}\beta^{\tilde{R}} + \tilde{Z}_j\delta^{\tilde{R}} + \mu_t^{\tilde{R}}\theta_t^{\tilde{R}} + r_t\gamma^{\tilde{R}} + f(\dot{x}_{ijmt}^{\tilde{R}}) + \epsilon_{ijmt} \end{Bmatrix} \tag{3}$$

The equations in (2) are estimated as a Generalized Additive Model (GAM) with an intercept based on Hastie and Tibshirani (1990), whereas selected metric covariates are included additionally as smooth functions (e.g.: cubic, cyclic cubic, penalized or thin plate splines) to increase the explanatory power based on the results of: Cajias and Ertl 2018, Geniaux and Napoléone, 2008; Wood, 2006. ε_{ijt} and ϵ_{ijt} are iid errors and $\beta^{\tilde{P}}$, $\beta^{\tilde{R}}$, $\delta^{\tilde{P}}$, $\delta^{\tilde{R}}$, $\theta_t^{\tilde{P}}$, $\theta_t^{\tilde{R}}$, $\gamma^{\tilde{P}}$ and $\gamma^{\tilde{R}}$ are the coefficients to be estimated. The GAM approach corresponds to the expansion of the OLS estimator by a penalized smooth term λ minimizing the sum of squared errors SSE as:

$$SSE = \sum_i (y_i - \sum_k \beta_k X_{ik})^2 - \lambda \int (f''(x))^2 dx \tag{4}$$

The second term in (4) corresponds to the non-linear function of covariates with at least one second derivative, while the parameter λ determines the amount of smoothness that has to be introduced additionally to the OLS estimator, see Wood, 2006. Based on pooled cross-sectional observations the estimated hedonic equations adjust $\tilde{P}_{ijmt}$ and $\tilde{R}_{ijmt}$ for hedonic characteristics, temporal and spatial effects and estimate implicit prices and rents $\widehat{\tilde{P}_{ijmt}}$ and $\widehat{\tilde{R}_{ijmt}}$. The regression models are estimated in parallel with an elastic computing system in the cloud due to the high amount of data and huge memory requirements.

In order to estimate the NIY we need to track a dwelling that has been sold out and rented at every point in time in each market. Since this is unlikely to be the case, we need to extract

implicit prices and rents from (3) to estimate a NIY for an artificial identical dwelling. In other words, we need to estimate the willingness to pay for the same asset from both a buy-to-hold and rental investment strategy. Let the set of artificial hedonic characteristics to be defined as $\bar{X}_t$ independently from j and m for both investment strategies. That means that equation (3) estimates the implicit asking price and implicit asking rent for the same dwelling in each market across time. One may ask whether the data bases are comparable for estimating such an artificial implicit dwelling. This issue is covered in the next section, but given the criteria harmonization in the German MLSs over the last years data size and sample comparability are less crucial. Especially, missing characteristics and household bias are not a problem in comparison to the studies of Hill, Iqbal and Syed, 2016 and Linneman and Voith, 1991.

The Net Initial Yield (NIY) calculation is as follows. Let the estimated NIY be denominated as $\hat{\imath}_{mt}$ in time t and market m. In order to incorporate as much as real investment information as possible we assume a transaction costs of $TC = 4\ \%$ in the case of a buy-to-hold strategy, e.g. broker fee, technical, legal and due diligence, etc. Additionally, we include property transaction taxes which are set up by each German federal states as Tax_{mt}. Since we track net asking rents, i.e. without heating, we do not incorporate further costs in the numerator. The NIY for market m at quarter t is estimated as:

$$\hat{\imath}_{mt} = 100 * \frac{12 * e^{\left(X_{ijmt}^{\tilde{P}} \hat{\beta}^{\tilde{P}} + \tilde{Z}_j \hat{\delta}^{\tilde{P}} + \mu_t^{\tilde{P}} \hat{\theta}_t^{\tilde{P}} + ir_t \hat{\gamma}^{\tilde{P}} + \hat{f}\left(\dot{x}_{ijmt}^{\tilde{P}}\right)\right)}}{(1+Tax) * (1+TC) * e^{\left(X_{ijmt}^{\tilde{R}} \hat{\beta}^{\tilde{R}} + \tilde{Z}_j \hat{\delta}^{\tilde{R}} + \mu_t^{\tilde{R}} \hat{\theta}_t^{\tilde{R}} + r_t \hat{\gamma}^{\tilde{R}} + \hat{f}(x_{ijmt}^{\tilde{R}})\right)}} \tag{5}$$

, or simplified to

$$\hat{\imath}_{mt} = 100 * \frac{12 * e^{\widehat{\tilde{R}_{mt}|\bar{X}_t}}}{(1+Tax) * (1+TC) * e^{\widehat{\tilde{P}_{mt}|\bar{X}_t}}} \tag{6}$$

For example, given an implicit monthly rent of 1,600 €/p.m. and an implicit price of 600,000 € in Munich, where the property tax is 3.50 %, the NIY in t corresponds to:

$$\hat{\imath}_{Munich,t} = 100 * \frac{12*1{,}600}{(1+0.035)*(1+0.04)*600{,}000} = 2.97\ \% \tag{7}$$

3. DATA COLLECTION AND DESCRIPTIVE STATISTICS

We gather data from six sources. Real estate prices and rents come from Empirica systems (www.empirica-systeme.de) which collects georeferenced real estate data from the 100 most important German Multiple Listing Systems (MLS) such as ImmoScout, ImmoNet or ImmoWelt for whole Germany. Based on an algorithm empirica identify doubles and harmonizes the big data. Furthermore, we obtain the number of households and the purchasing power per household on a ZIP basis from the GfK for 2017 and gather the effective interest rate for residential loans with a maturity higher than 10 years from Reuters Eikon on a monthly basis. We extracted the geo-shapefiles of the German territory from Eurostat in order to calculate two spatial gravity variables: the distance to the centroid of the city and to the ZIP centroid. After merging the data and calculating the gravity variables with R, we proceed with the data selection. This is a very important step to avoid skewed and biased NIYs. In order to benchmark our estimated NIY with real data we obtain the average net initial yields from CBRE, however only for existing dwellings and for only the top 7 German residential markets. Finally, we gather demographic data on the NUTS3 level from Oxford Economics. More specifically, we gather the population and the working population.

--- Insert Table 1: Data mining and sample description ---

The initial sample comprises of more than 3.9 million observations across 403 NUTS3 areas. NUTS3 regions correspond to the "Nomenclature of territorial units for statistics", which is a hierarchical system for dividing up the economic territory in Europe. While the NUTS1 consists on major socio-economic regions, the NUTS3 regions cover small regions similar to counties or administrative districts. (www.ec.europa.eu/eurostat/web/nuts/overview). In order to ensure a sufficient and adequate data basis we decided to select NUTS3 regions with at least 100 observations per quarter in the price and rental sample respectively. The vector of

selected NUTS3 in the rental sample was longer in comparison to the price sample and we chose the common NUTS3 areas fulfilling the requirement of 100 observations per quarter. This step led to a reduction of data of around 22 % leading to a final data set of 3,055,343 observations from 2013 Q1 until 2018 Q1. The rental sample accounts for 68 % of the final data, whereas prices of dwellings for the remaining 32 %.

Figure 1 shows the distribution of both samples – the ownership market and the rental market – across time, markets and space. The upper left graph shows that the number of adverts in both samples is highly correlated across time and that market activity takes place in both markets uniformly. In other words, the data generating process of both samples is not skewed over time as the relative shares of adverts follow the same path. The upper right graph provides the cumulative share of data across markets in both samples, i.e. a cumulative concentration analysis. The 45° line would mean that the number of observations is equally distributed across markets, but the sample shows a slight bias toward the rental markets due to a higher amount of data. The maps at the bottom show the spatial distribution of the samples across the German markets. Each dot represents one observation and provides an overview of the markets covered in the paper that fulfil the requirement of 100 observations per quarter. Overall the data basis has slightly more observation in the rental market, but the distribution in contrast to the price data is homogenously distributed across space, time and markets, reducing the biasedness and skewness in the estimation of the NIY.

--- Insert Figure 1: Sample description over time, markets and space ---

Each of the $i = 3{,}055{,}343$ dwellings in the $m = 161$ residential markets starting in 2013 Q1 until 2018 Q1 – i.e. $t = 1, \dots, 21$ – has the following hedonic characteristics: living area, age, latitude, longitude, number of rooms, with built-in kitchen, with bathtub, with parking slot, with balcony, with terrace, with elevator, as-good-as-new and refurbished. The gravity variables included measure the distance to the ZIP and NUTS3 centroid in kilometres and the

GAM regressions in (3) include the purchasing power per household and the number of households in log for each of the $j = 1, \ldots, 4015$ ZIP codes ($\tilde{Z}_j$). In order to control for financial conditions the regressions include the effective interest rate for housing loans with a maturity of more than 10 years at the month of entry of each advertise gathered from the Bundesbank via Reuters Eikon. The integration process of the six different sources is shown graphically in Figure 2.

--- Insert Figure 2: Integration process of big data research approach ---

--- Insert Table 2: Mean of selected variables by purchasing power groups ---

--- Insert Table 3: Variable description and artificial dwelling specification ---

Given the 3.06 million observations the provision of detailed descriptive statistics and correlations is complicated and meaningless. The appendix shows the detailed descriptive statistics. Instead, we provide a short overview of selected variables across purchasing power quantiles, which is an appropriate proxy for spatial and economic developments. Prices and rents increase for higher purchasing power quantiles, which corresponds to higher expenditure levels for housing by households. In contrast, the average age of the dwellings decreases the higher is the purchasing power which points to lower prices and rents for older dwellings. The mean living area of dwellings in the ownership sample has a low variation across high and low purchasing power categories. Most interestingly however is the average living area of dwellings in the rental market as it ranges from 64 m² to 82 m², meaning that a larger share of smaller dwellings is advertised in the rental market. Finally, the descriptive statistics show that higher prices and rents are achieved outside the city centre as the gravity variable and the purchasing power quantiles correlate positively on average. Table 3 provides broad descriptive statistics of the covariates as well as the specification of the artificial vectors $\bar{X}_t$

which are used for the estimation of the implicit prices and rents $\widehat{\tilde{P}_{\iota jmt}}$ and $\widehat{\tilde{R}_{\iota jmt}}$. The next section presents the econometric results of the NIYs.

4. ECONOMETRIC RESULTS

In this section, we present the empirical estimated net initial yields - $\hat{\iota}_{mt}$ - based on the willingness to pay for an artificial identical dwelling in a buy-to-hold and rental investment strategy under consideration of transaction fees and taxes. The approach considers more than 3 million observations from 2013Q1 until 2018Q1 over 161 NUTS3 German residential markets. Overall, we estimate $m * t = 3{,}381$ net initial yields.

4.1. Cross-sectional distribution of Net Initial Yields

The average of the NIYs over the entire sample period is 3.51 %. The lowest NIYs were to find in Munich, Berlin, Frankfurt, Stuttgart, Augsburg and Ebersberg; markets that correspond to the most attractive in terms of economic activity, innovations and labour conditions, but also locations with strong demand for housing. In contrast, the less attractive markets based on the analysis were Olstholstein, Meißen, Aurich, Rostock and Kleve; markets with substantial risk profiles in terms of future labour supply and market activity. Figure 3 shows the behaviour of our estimated NIY in contrast to the share of working population in each market and the growth in the working population. It is to expect that markets with a higher proportion of workers relative to the overall population will be able to growth steadily in the long-run and consequently demand more housing and provide stable rental growth. Thus, when looking at the left panel in Figure 3 the relationship between NIY and the share of working population is negative pointing to a higher investment attractiveness in “young” cities. The panel also shows the downward shift in the NIY between 2013Q1 and 2018Q1 confirming that prices and rents within the cities do respond to labour composition. Finally, the panel in the right shows the growth in the working population and the NIY in 2018Q1. To

put it in simple words, cities with a continuous growth in their working population experience the lowest NIYs and consequently high investment attractiveness.

--- Insert Figure 3: Net initial yields in contrast to demographic indicators ---

--- Insert Figure 4: Mean and standard deviation of net initial yields by market ---

Figure 4 shows the mean and the standard deviation of the NIYs for each market in ascending order across the entire estimation period. Two main finding are to highlight. First, the estimated NIYs vary within a range of ca. 200 basis points (BP) starting at 2.5 % in a uniform distribution. This means that the employed data and the econometric approach enable investors for looking at buy-to-hold and rental investment strategies accurately. In other words, the smooth ascending distribution of the NIYs across the markets does reflect the relative performance of markets accurately without abrupt movements. Finally, the NIYs are useful from an investment perspective as they indeed represent the current financial performance without an overestimation. To put it simple, the yield ranges – measured as the standard deviation – of the most and less attractive markets are comparable reflecting an equally distributed risk-return investment universe across markets, e.g. an investment strategy of 4 % will be less likely in Munich than in less attractive cities. Overall, the distribution of the estimated NIYs shows a concordance with main demographic indicators as well as a real convergence in terms of market attractiveness and risk.

4.2. Temporal distribution of Net Initial Yields

--- Insert Figure 5: Distribution of net initial yields over time ---

While the cross-sectional perspective has confirmed that the estimated NIYs obey an urban economic hierarchy the intertemporal analysis in figures 5, 6 and 7 show their development over time. The boxplots presented in Figure 5 show on the one hand that the NIY have decreased continuously over time across the German residential markets. The yield

compression took place at a rate of ca. 3 BP per quarter or ca. 10 BP per year. By the end of the observation period in 2018 Q1 the average initial performance of the real estate investments fell by 67 BP to 3.17 % on average. More interestingly is the fact that the yield compression has taken place across every market. Thus, regardless of the market size and location prices and rents have increased continuously leading to a uniform decline of NIY to historic low levels. Based on the empirical results an investor willing to deploy its capital with a target NIY of at least 4.0 % (without leverage) in 2013 would have had the opportunity to screen several investment strategies across different markets. However, the same approach in 2018 would lead to a reconsideration of the investment strategy as only a few markets could be targeted for such a return profile.

--- Insert Figure 6: Distribution of the bond yield gap over time ---

--- Insert Figure 7: Slope shift in implicit price and rents: 2013Q1 and 2018Q1 ---

So far, the results show that the employed data and empirical approach do well in mimicking the performance of residential assets in Germany. But, how do the estimated NIY behave relatively to other investments assets? Figure 6 shows the relative performance of residential investments in contrast to the German 10-year government bond yields, i.e. the bond yield gap in BP. While the compression in the NIYs has decreased steadily over the last six years, it has been accompanied at the same time by a strong reduction in the yields for 10-year government bonds. Yet, the bond yield gap has risen up to 400 BP and in some cases by more than 450 BP, reflecting the attractiveness residential assets as an own investment asset class. Although the gap has fallen over the last seven quarters residential assets continue to be strongly demanded by institutional investors, aiming at exactly benefit from this yield gap. Finally, Figure 7 shows the mean implicit prices and rents for the same artificial dwelling across each market in 2103Q1 and 2018Q1 used in the calculation of the NIYs. The respective regression

slope has increased meaning that prices have increased stronger than rents, increasing the value in the denominator of (5) stronger that the value in the numerator.

4.3. Spatial distribution of Net Initial yields

From a spatial perspective, the NIY compression has taken place across all markets over the observation period, but predominantly in the urban areas surrounding the top 7 German markets Berlin, Munich, Dusseldorf, Cologne, Stuttgart, Frankfurt and Hamburg. The lowest NIYs are to find in these cities and the surrounding regions are increasingly experiencing a stronger demand for residential assets given the low new supply levels.

--- Insert Figure 8: Spatial distribution of net initial yields 2013Q1, 2018Q1 and change ---

Figure 8 shows the NIYs in 2013Q1 and 2018Q1 and the corresponding yield compression in BP. While the majority of the markets in 2013Q1 show NIYs higher than 3.5 % in grey the map in the centre shows the development of the NIY into levels far below 3.0%. Yet, there is on the one hand a substantial movement in the NIYs on a regional perspective, but the yield compression in the right panel shows that almost every market suffered a yield compression regardless of the location and the closeness to the top 7 markets. Although the price-rent-relationship convergences across real estate markets from a spatial perspective and signals stability and low downside risks, it is necessary to notice that possible market corrections arising from macroeconomic divergences will likely affect NIYs in secondary and tertiary markets stronger relative to the main urban areas.

5. SENSITIVITY AND ROBUSTNESS ANALYSIS

Figure 9 puts the estimated NIYs in contrast to the NIYs provided by CBRE for the top 7 German markets and only for existing dwellings. The comparison confirms on the one hand that the empirical big data approach presented in this paper and CBRE's approach based on investment comparables do follow the same temporal path of rising prices and rents. In other words, CBREs NIY's approach based on contract and appraisal rents and prices of mostly

institutional deals mimics a similar price momentum in comparison to the proposed big data approach. Obviously, the differences are on the nature of the data – i.e. institutional portfolios vs. private persons – but the bottom line of both approaches regarding the market pricing is the same. This confirms that the pricing of the German housing market is transparent and perhaps less likely to present elevated levels of information asymmetry since different investment approaches lead to a common market understanding. On the other hand, the gap between the mean NIY for the top 7 markets from both the empirical and CBRE's approach is constant over time. The gap results principally from the different investment approach, i.e. an artificial newly-built dwelling and an existing dwelling in case of the CBRE data.

Finally, figure 10 shows the sensitivity of the econometric approach when varying the artificial hedonic vector to mimic the NIY for both a newly-built dwelling and an existing dwelling as defined in the last column of Table 3. From a theoretical point of view, households would be willing to pay more for a higher housing utility leading to high prices and rents and consequently to low yields. Therefore, the NIY of existing dwellings should be higher than for newly-built dwellings as the willingness to pay decreases. Figure 10 shows as expected a rightward shift in the NIY distribution pointing to lower prices and rents and consequently higher yields for the hedonic vector of existing dwellings in comparison to newly-built dwellings.

Since the robustness check with real market data from CBRE and the sensitivity results based on an investment strategy in newly-built and existing dwellings, the proposed methodology behind the estimated NIYs is robust and reflects actual market developments from both the ownership and rental market.

--- Insert Figure 9: Comparison of estimated vs. CBRE net initial yields ---

--- Insert Figure 10: Slope shift in implicit price and rents: 2013Q1 and 2018Q1 ---

6. RECOMMENDATIONS AND CONCLUSION

In search for investable products, institutional investors increasingly set their focus in the residential asset class in order to diversify across different levels of income and capital returns as well as different risk profiles. Especially in Europe, the residential sector has opened over the last decade and is becoming more and more mainstream. When assessing the pricing condition of real estate markets, institutional inventors and investment managers usually look at the level and development of net initial yields, which are at the same time traditionally provided by international brokers, like JLL or CBRE, mostly for the commercial sector. The investment case in residential markets is certainly different as net initial yields are scarce not only due to the long holding period and the lack of a centralized data base, but especially due to the relatively low number of investment comparables. This is the gap the paper aims at closing at.

In the era of digitalization residential assets are predominantly brought to the market via digital multiple listing systems allowing landlords increase asset liquidity and reach wider range of potential users. The paper makes use of more than 3.06 million big data observations and proposes a semiparametric regression approach for estimating net initial yields adjusting for spatial, time and hedonic characteristics, but also controlling for property taxes and transaction costs. The 3,381 estimated net initial yields in the 161 German markets showed a spatial pattern with the biggest and most attractive cities showing the lowest yields as well as a self-adjusting in the markets surrounding the top cities. The development of net initial yields over time reflects that prices have increased stronger than rents over the last six years leading to rock bottom yields for residential assets and a significant premium in comparison to government bond yields. Finally, the approach responds to the spatial hierarchy of markets in Germany, meaning that the level of the estimated yields is accurate and from an investment perspective achievable.

Transparency is an essential barrier when assessing the pricing conditions of markets and deriving investment decisions. Although international brokers do provide detailed investment comparables on – mostly commercial – real estate markets the residential sector remains a puzzle when it comes to investment yields. The paper sheds light on this problem from a conceptual, methodological and empirical perspective and confirms that residential investment yields are deducible by making usage of advanced hedonic models and big data.

7. REFERENCES

Anglin, P., & Gençay, R. (1996), Semiparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, 11(6), 633–648. doi:10.1002/(SICI)1099-1255(199611)11:6<633::AID-JAE414>3.0.CO;2-T

Bao, H., & Wan, A. (2004), On the Use of Spline Smoothing in Estimating Hedonic Housing Price Models: Empirical Evidence Using Hong Kong Data. *Real Estate Economics*, 32(3), 487–507. doi:10.1111/j.1080-8620.2004.00100.x

Cajias, M. & Freudenreich, P. (2018), Exploring the determinants of liquidity with big data – market heterogeneity in German markets, *Journal of Property Investment & Finance*, 36 (1), pp.3-18, https://doi.org/10.1108/JPIF-01-2017-0006

Cajias, M. & Ertl, S. (2018), Spatial effects and non-linearity in hedonic modeling: Will large data sets change our assumptions?, *Journal of Property Investment & Finance*, Vol. 36(1), pp.32-49, https://doi.org/10.1108/JPIF-10-2016-0080

Cajias, M. (2018), Is there room for another hedonic model?: The advantages of the GAMLSS approach in real estate research, *Journal of European Real Estate Research*, Vol. 11(2), pp.204-245, https://doi.org/10.1108/JERER-07-2017-0025

Davidson, R. & MacKinnon, J.G. (2003), Econometric Theory and Methods, second edition, New York, Oxford University Press.

Diebold, F.X. (2004), Elements of forecasting, Thomson South-Western.

Eurostat (2013), Handbook of residential property price indices, Publication office of the European Union, ISBN: 978-92-2-127359-2.

Eurostat (2013), Technical manual on owner-occupied housing and house price indices, Publication office of the European Union, http://ec.europa.eu/eurostat/documents/7590317/0/Technical-Manual-OOH-HPI-2017/.

Hutchison, N., Fraser, P., Adair, A. & Srivatsa, R. (2011), The risk free rate of return in UK property pricing, *Journal of European Real Estate Research*, 4(3), 165-184.

Geniaux, G., & Napoléone, C. (2008), Semi-Parametric Tools for Spatial Hedonic Models: An Introduction to Mixed Geographically Weighted Regression and Geoadditive Models. In A. Baranzini, J. Ramirez, C. Schaerer, & P. Thalmann (Eds.), Hedonic Methods in Housing Markets (pp. 101–126). New York, NY: Springer New York.

Hastie, T. and Tibshirani, R. (1990), Generalized Additive Models, Monographs on statistics and applied probability, 1. edition, Chapman and Hall/CRC.

Hill, R. & Syed, I. (2016) "Hedonic price–rent ratios, user cost, and departures from equilibrium in the housing market", *Regional Science and Urban Economics*, 56(C), 60-72.

Jones, C., Dunse, N. & Cutsforth, K. (2015), "The changing relationship between government bond yields and capitalisation rates – Evidence from the UK, USA and Australia", *Journal of European Real Estate Research*, 8(2), 153-171.

Linneman, P. & Voith, R. (1991) "Housing price functions and ownership capitalization rates", *Journal of Urban Economics*, 30, 100-111.

Lisi, G. (2015), Use of hedonic prices to estimate capitalization rate. *International real estate review*, 18(3), 303-316.

Lyons, R.C. (2013), Price signals in illiquid markets – The case of residential properties in Ireland, 2006-2012, Trinity Economics Papers, TEP No. 0613.

Mason, C. & Quigley, J.M. (1996), Non-parametric Hedonic Housing prices, *Housing Studies*, 11(3), pp. 373–385.

McCord, M., Davis, P., Haran, M., McIlhatton, D. & McCord, J. (2014), Understanding rental prices in the UK: a comparative application of spatial modelling approaches, *International Journal of Housing Markets and Analysis*, 7(1), pp. 98–128.

Pace, K. (1998), Appraisal Using Generalized Additive Models. *Journal of Real Estate Research*, 15(1/2), 77–99.

R Development Core Team. R. (2016), A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, ISBN 3-900051-07-0.

Schimizu, C., Nishimura, K.G. & Watanabe, T. (2012), House prices from magazines, realtors and the land registry, Property markets and financial stability, *Bank for international Settlements BIS Papers*, No. 64, pp. 29-36.

Widłak, M., Waszczuk, J. and Olszewski, K. (2015), Spatial and hedonic analysis of house price dynamics in Warsaw, *National Bank of Poland*, Working Papers No. 197.

Wood, S. (2006), Generalized additive models: An introduction with R. Texts in statistical science, Chapman & Hall/CRC.

Table 1: Data mining and sample description

	Submarket		All data
	Rental market	Ownership market	
All rental market data	2,698,550 (68%)	1,234,656 (32%)	3,933,206
% data reduction due to small markets	22.8%	21.2%	21.9%
Final estimation sample	2,082,179 (68%)	973,164 (32%)	3,055,343

Notes: Data reduction due to markets with less than 100 observations per quarter. This corresponds to small areas lacking sufficient data for yield estimation.

Table 2: Mean of selected variables by purchasing power groups

Purchasing power by household	Purchasing power quintiles									
	ZIPs with low purchasing power							ZIPs with high purchasing power		
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Ownership market										
Asking prices	226,205	227,445	235,075	210,786	217,577	221,995	226,446	233,914	264,961	348,017
Distance to centre	4.9	7.6	6.5	7.2	7.6	8.6	8.4	9.6	8.6	8.5
Living area	82.0	81.8	83.3	82.1	83.8	84.9	84.4	85.2	87.4	89.4
Dwelling's age	60.3	52.5	48.2	39.7	36.8	34.0	31.8	29.6	29.2	27.2
Rental market										
Asking rents	462.3	472.0	529.8	541.4	573.4	577.4	627.2	642.4	698.6	830.7
Distance to centre	5.0	5.1	6.0	5.8	6.3	7.1	7.7	8.6	8.3	8.4
Living area	64.7	65.4	67.8	68.8	70.9	72.3	75.1	77.0	80.0	82.1
Dwelling's age	69.9	65.9	60.2	57.3	52.1	47.8	44.6	40.4	36.6	35.2

Notes: Values measure the mean of the selected variables over time by sample. Purchasing power quintiles range from 26,000 €/per household/p.a. to 103,000 €/per household/p.a.

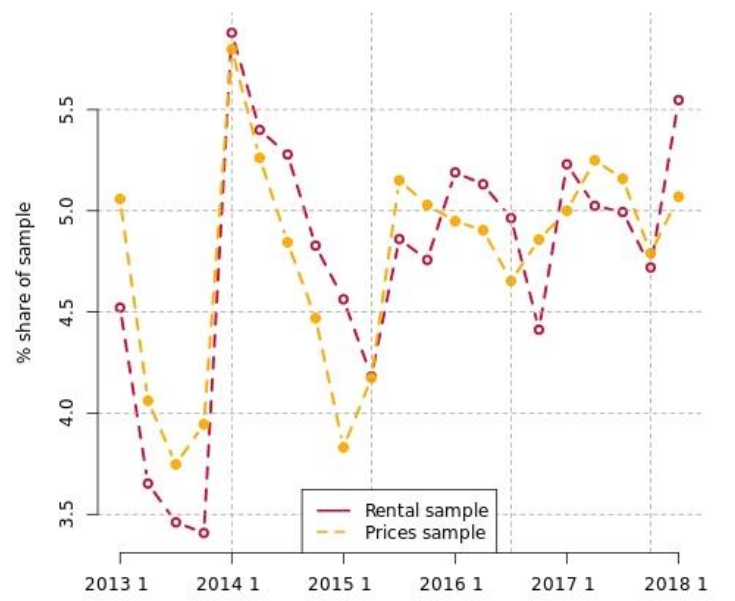
Table 3: Variable description and artificial dwelling specification

Sample Variables	Unit	Source	Ownership sample		Rental sample		Artificial hedonic vectors	
			Mean	Sd	Mean	Sd	Newly built dwellings	Existing dwellings
Asking rents and prices	€/p.m. and €/asset	Empirica	595.05	382.60	241,194	229,670	/	/
Housing loan interest rate	%	Eikon	2.026	0.469	2.038	0.480	/	/
Log purchasing power	log €/household/p.a.	GfK	10.654	0.209	10.713	0.201	/	/
Log households	log number of households		9.227	0.491	9.141	0.593	/	/
Centroid ZIP	Km.	Eurostat and R	1.086	0.947	1.061	0.902	/	/
Centroid city centre	Km.		6.819	5.038	7.738	6.213	/	/
Log living area	log m²	Empirica	4.209	0.385	4.346	0.424	75	120
Dwelling age	Years		51.006	34.117	38.941	33.805	5	20
Number of rooms	Integer		2.621	0.957	2.934	1.134	3	4
With bathtub	Binary		0.518	0.500	0.473	0.503	1	1
With built-in kitchen	Binary		0.409	0.492	0.401	0.491	1	1
With parking slot	Binary		0.447	0.497	0.691	0.463	1	0
With balcony	Binary		0.153	0.360	0.257	0.445	1	0
With terrace	Binary		0.589	0.492	0.643	0.480	1	0
With elevator	Binary		0.261	0.439	0.399	0.490	1	1
As-good-as-new	Binary		0.128	0.335	0.240	0.430	1	0
Refurbished	Binary		0.179	0.418	0.082	0.400	0	0

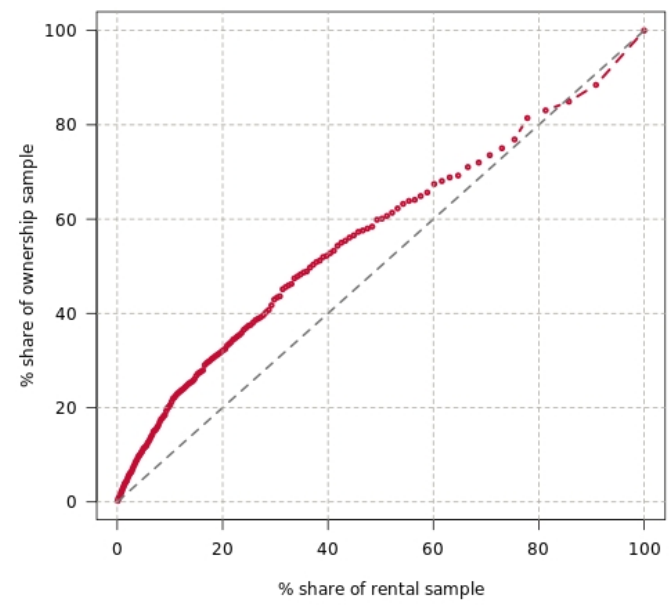
Notes: Final data set of 3,055,343 observations from 2013 Q1 until 2018 Q1.

Figure 1: Sample description over time, markets and space

Sample distribution over time

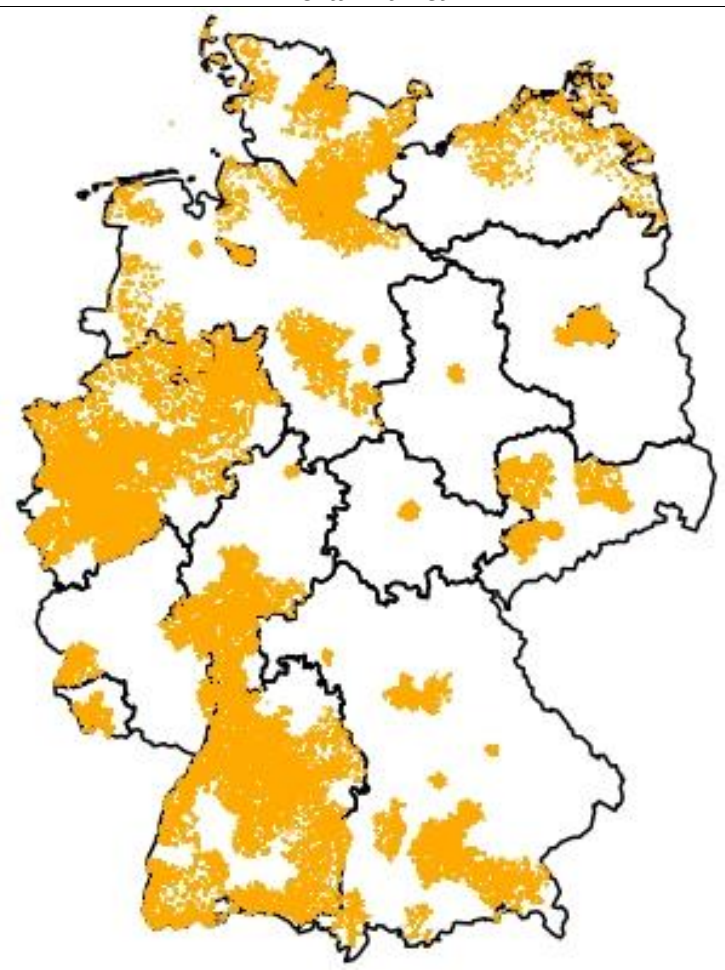


Cumulative concentration by market – 45° line

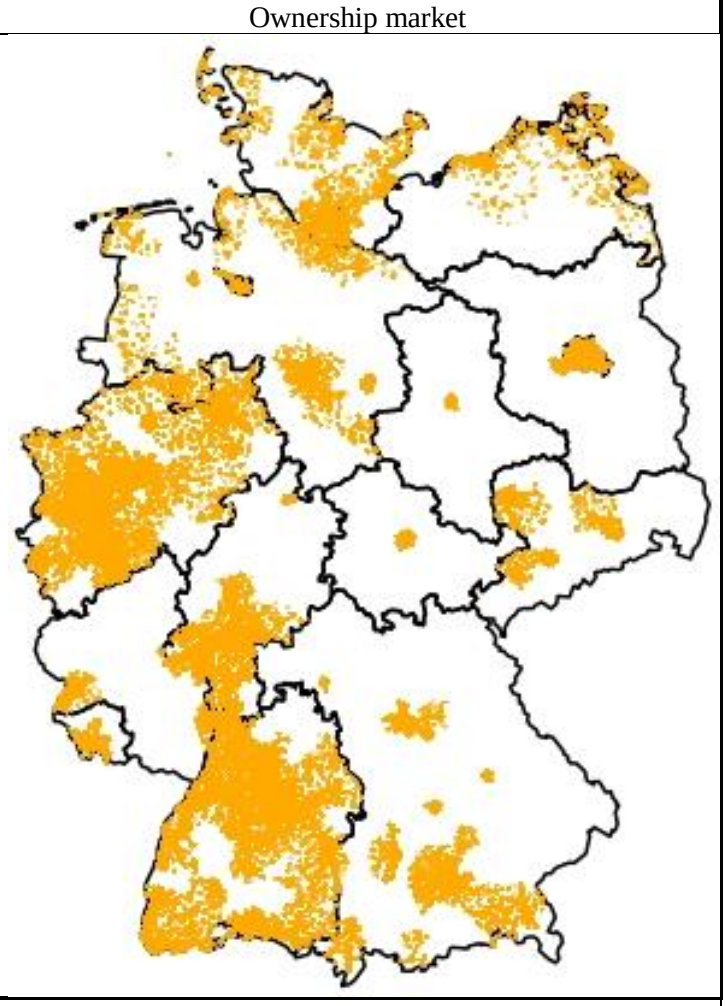


Spatial concentration of sample

Rental market

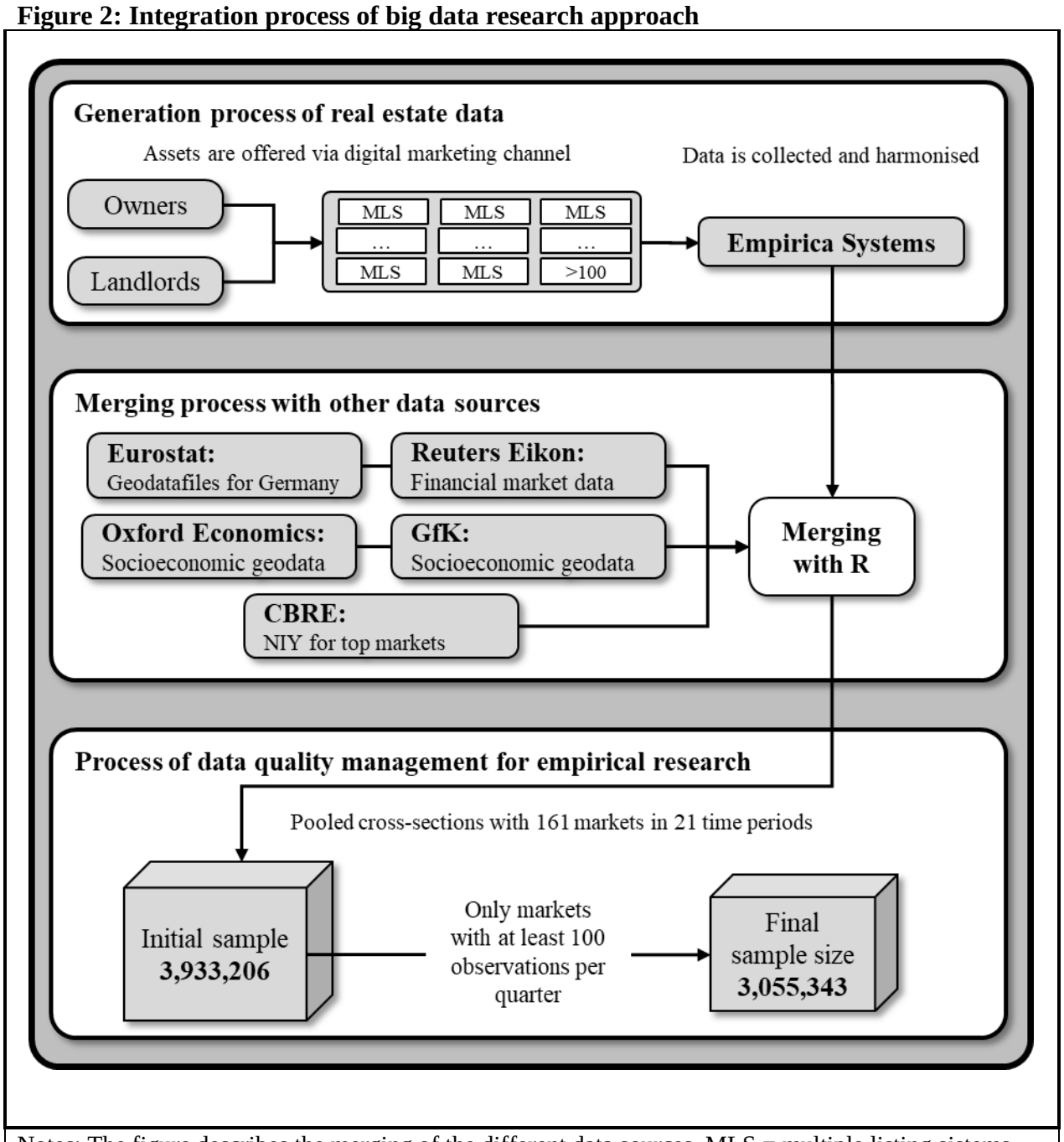


Ownership market



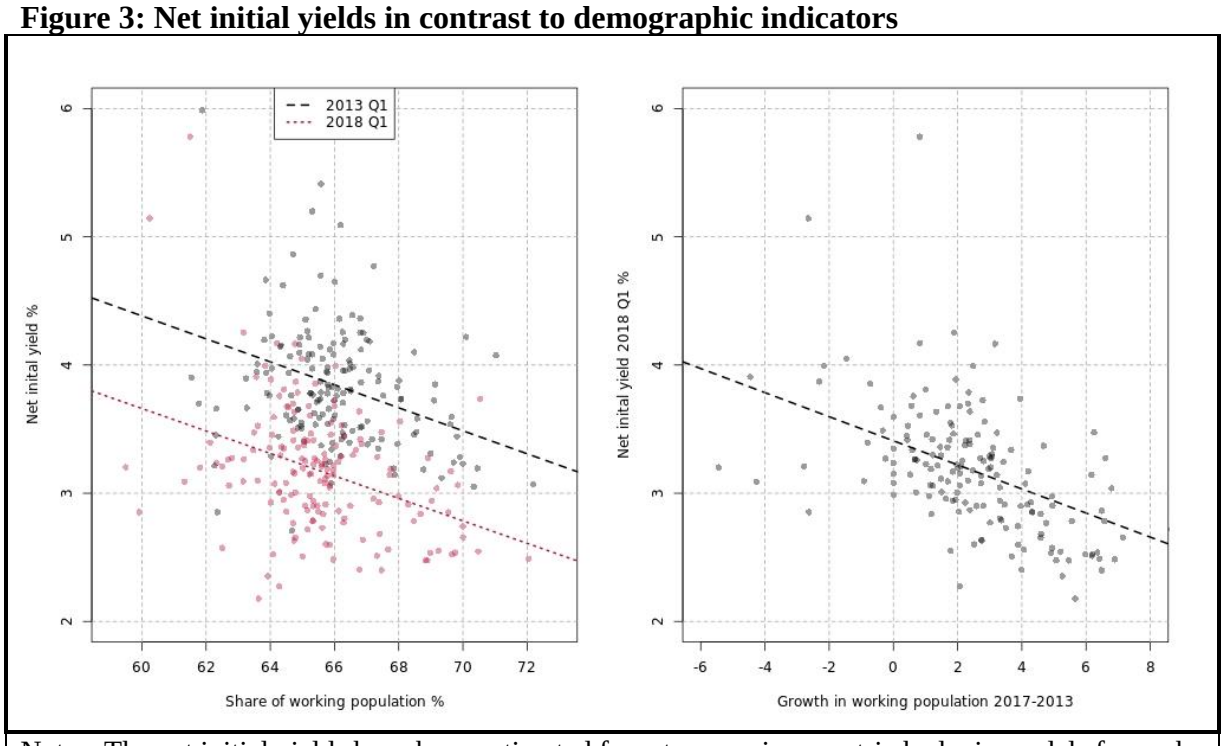
Notes: The upper left plot shows the relative distribution of the price and rental sample over time. The upper right plot shows the cumulative share of each sample by markets, whereas the 45° line corresponds to a uniform distribution. Each of the dots in the maps show the adverts in the price and rental sample based on 3,055,343 observations from 2013 Q1 until 2018 Q1. Markets with at least 100 observations per quarter have been chosen.

Figure 2: Integration process of big data research approach



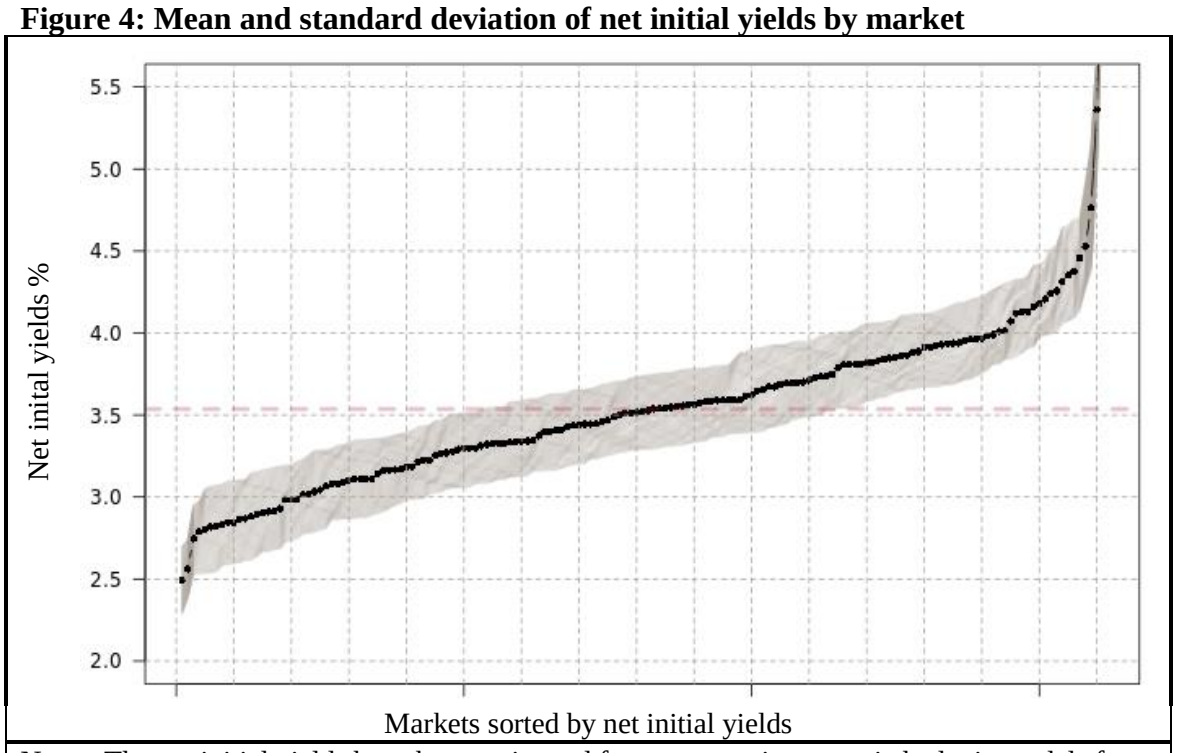
Notes: The figure describes the merging of the different data sources. MLS = multiple listing sistems, NIY= net initial yields.

Figure 3: Net initial yields in contrast to demographic indicators



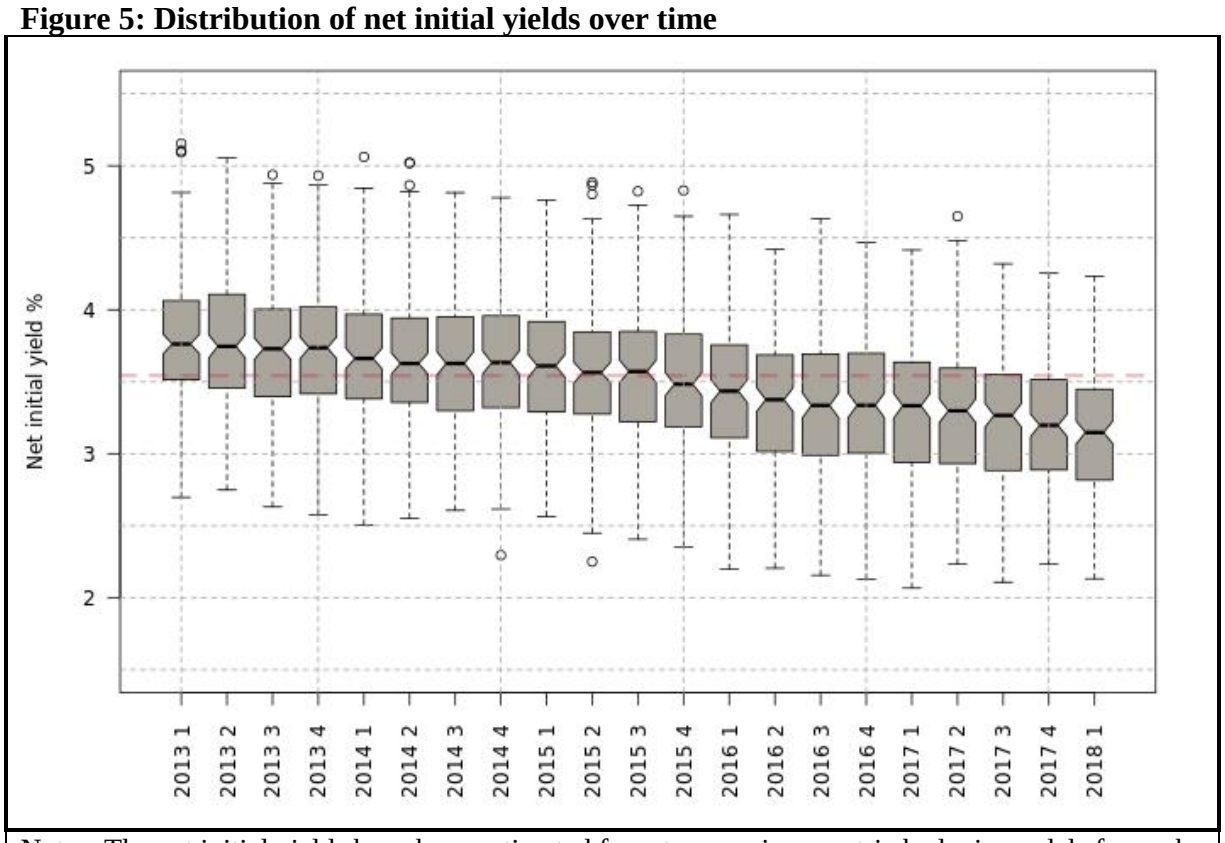
Notes: The net initial yields have been estimated from two semiparametric hedonic models for each market independently and adjusted by transaction costs for a buy-to-hold and a rental investment strategy. The demographic data corresponds to the definition of Eurostat on a NUTS3 level.

Figure 4: Mean and standard deviation of net initial yields by market



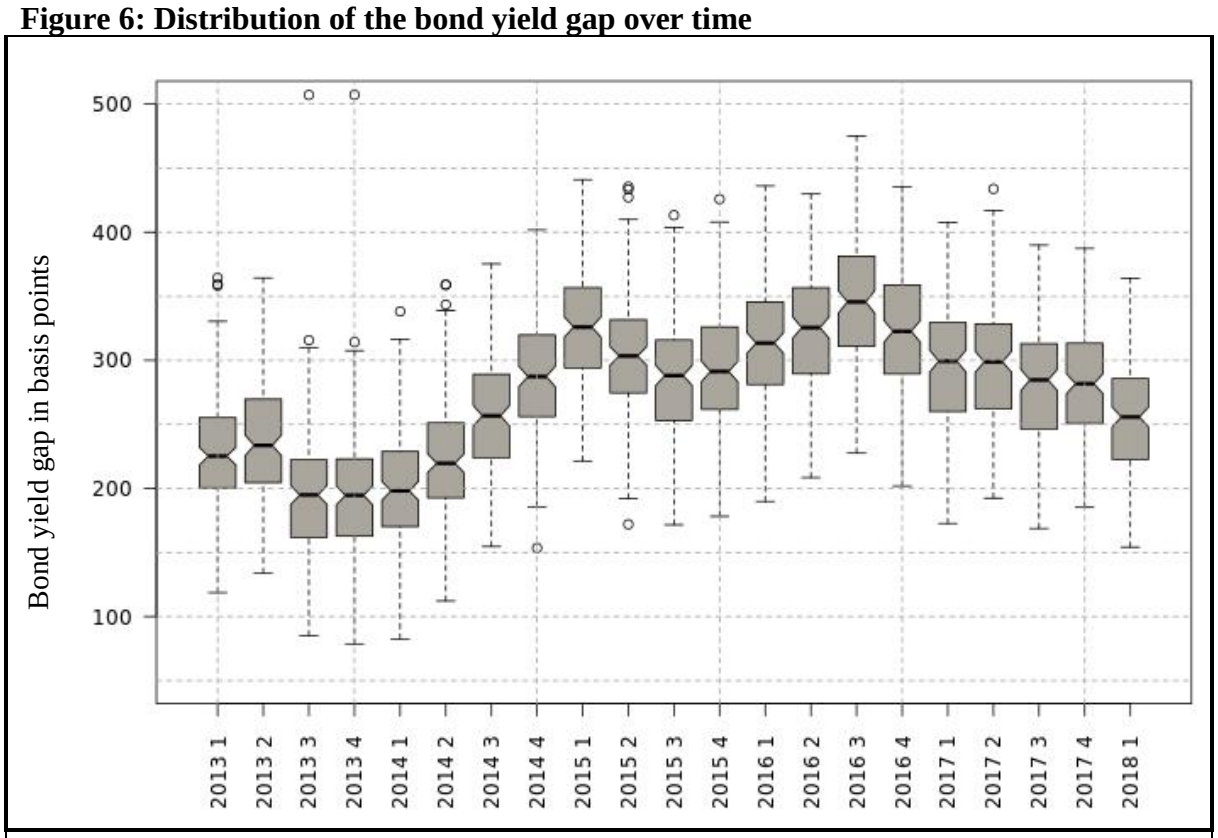
Notes: The net initial yields have been estimated from two semiparametric hedonic models for each market independently and adjusted by transaction costs for a buy-to-hold and a rental investment strategy. The range corresponds to the standard deviation across 2013 and 2018 on a quarterly basis.

Figure 5: Distribution of net initial yields over time



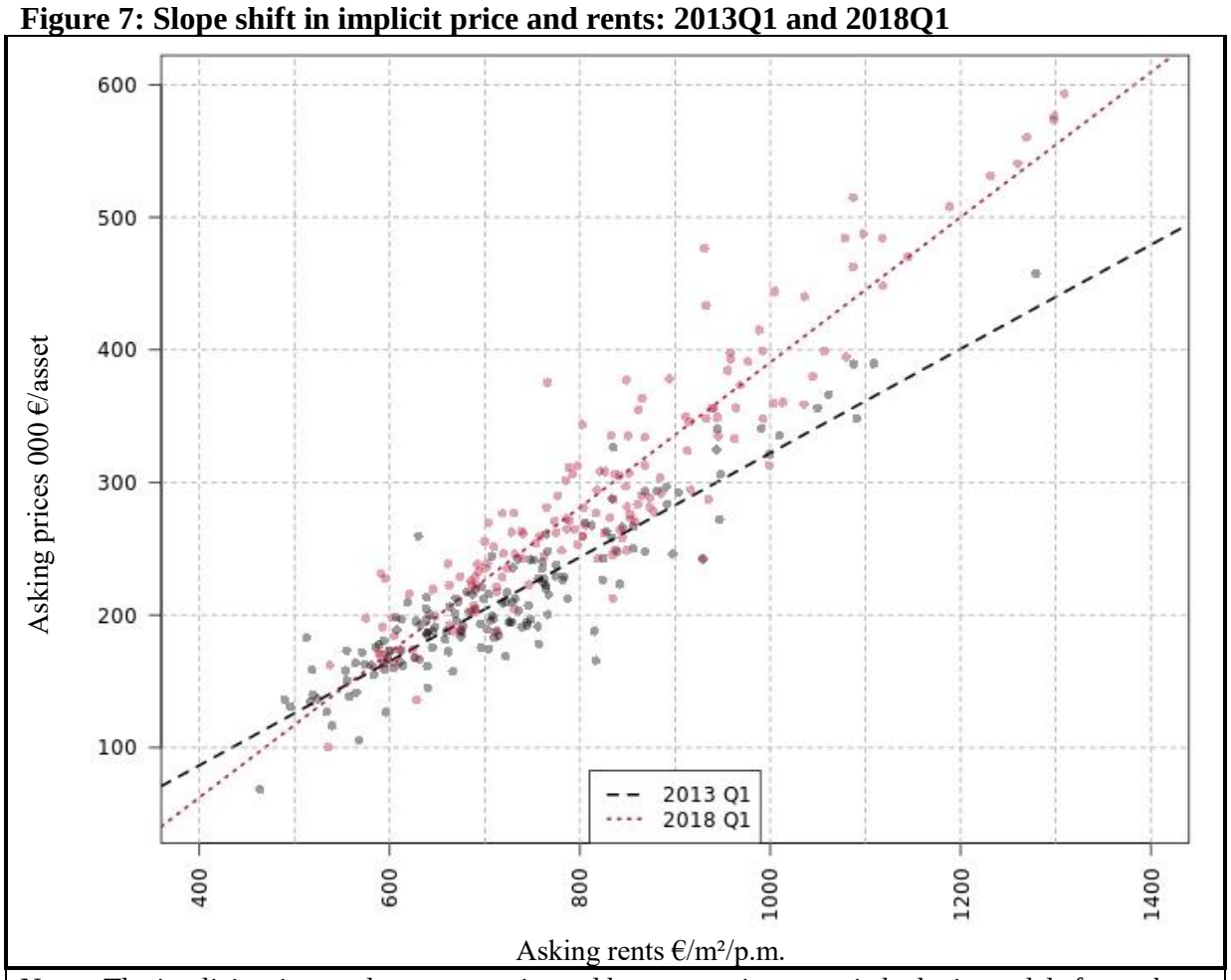
Notes: The net initial yields have been estimated from two semiparametric hedonic models for each market independently and adjusted by transaction costs for a buy-to-hold and a rental investment strategy. The line corresponds to the overall average of the net initial yields.

Figure 6: Distribution of the bond yield gap over time



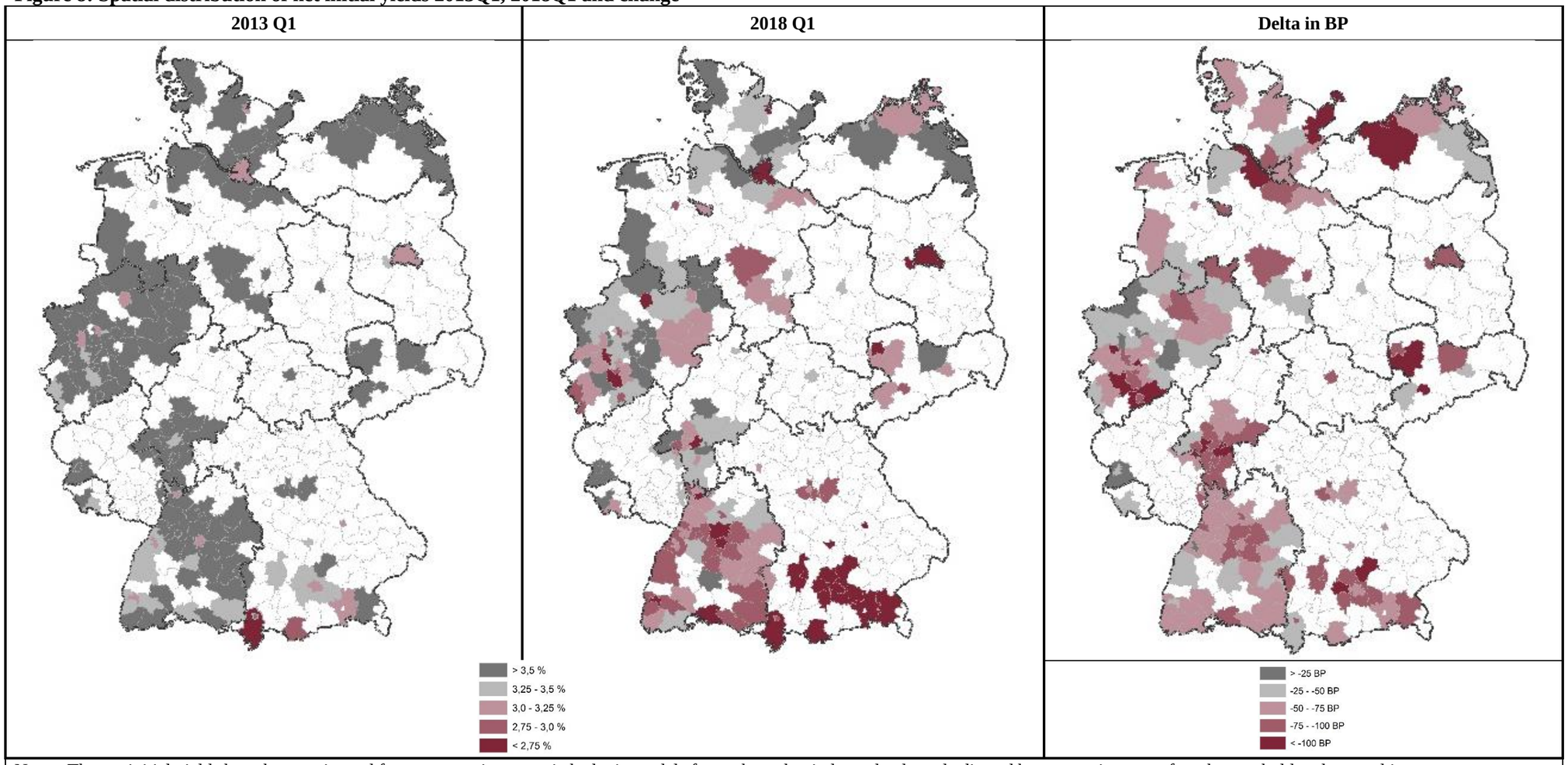
Notes: The bond-yield-gap is calculated by subtracting the net initial yields from the German ten-year government bond yield. The net initial yields have been estimated from two semiparametric hedonic models for each market independently and adjusted by transaction costs for a buy-to-hold and a rental investment strategy.

Figure 7: Slope shift in implicit price and rents: 2013Q1 and 2018Q1



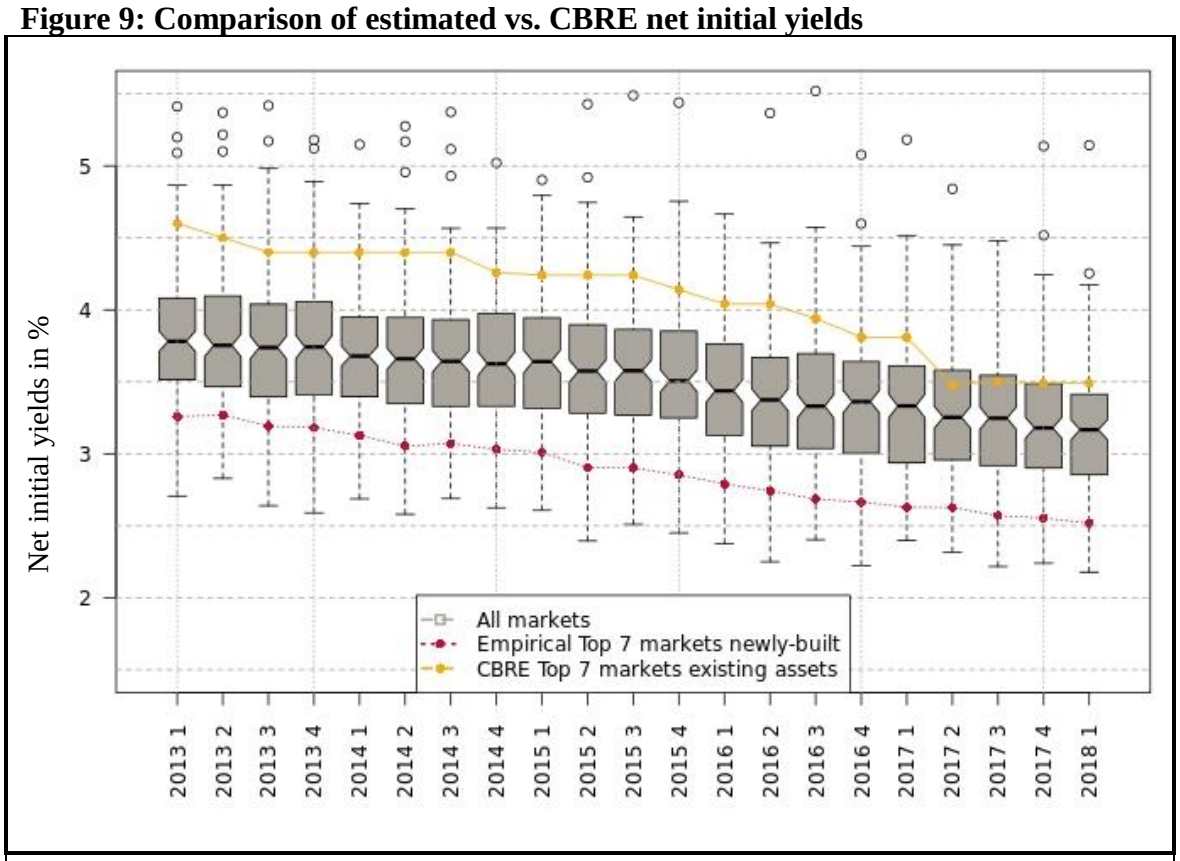
Notes: The implicit prices and rents are estimated by two semiparametric hedonic models for each market independently for an identical artificial dwelling. The rise in the slope represents a stronger increase in implicit prices relative to rents.

Figure 8: Spatial distribution of net initial yields 2013Q1, 2018Q1 and change



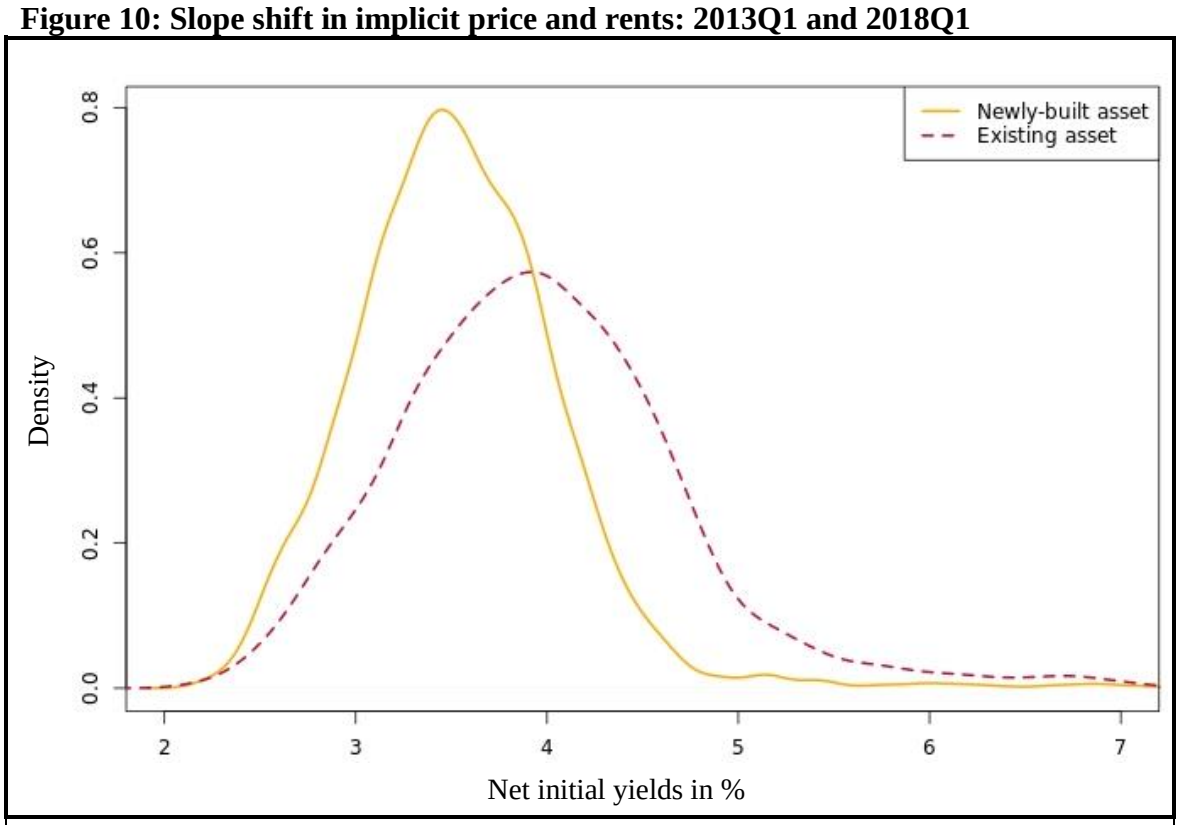
Notes: The net initial yields have been estimated from two semiparametric hedonic models for each market independently and adjusted by transaction costs for a buy-to-hold and a rental investment strategy. The vector of hedonic characteristics for the estimation of the net initial yields is for both samples – prices and rents – identical. The change in the net initial yields is measured in basis points on NUTS3 areas.

Figure 9: Comparison of estimated vs. CBRE net initial yields



Notes: Net initial yields from CBRE measured by the average for existing dwellings on the basis of investment comparables rather than adverts. Empirical estimated net initial yields represent newly-built dwellings and follow the same path as commercial yields.

Figure 10: Slope shift in implicit price and rents: 2013Q1 and 2018Q1



Notes: The density function shows the concentration of the net initial yields based on two different specifications of the artificial hedonic vector. The first vector corresponds to a newly-built dwelling and the second to an existing asset. The net initial yields for the latter are higher representing lower attractiveness.

Appendix 1.1: Sample size, descriptive statistics and explanatory power by market

NUTS3 Code	Rental Sample							Price Sample						
	N	Asking rents €/p.m.		Average			Adj. R² of GAM regression	N	Asking prices €		Average			Adj. R² of GAM regression
		Mean	Sd	Living area m²	Age	Distance to city Km.			Mean	Sd	Living area am²	Age	Distance to city Km.	
DE111	19.032	914	501	34,1	34,8	2,4	86,10%	14.225	316.770	256.938	37,9	35,8	2,3	85,52%
DE112	10.888	733	307	30,4	20,6	3,7	89,33%	11.820	225.378	120.809	31,2	20,9	3,7	90,26%
DE113	11.717	712	288	29,2	25,9	2,6	89,29%	11.595	218.409	121.739	32,6	25,0	2,9	89,16%
DE114	4.209	567	217	28,0	26,4	3,0	86,96%	5.467	171.806	92.731	32,7	26,0	3,2	87,41%
DE115	11.317	701	281	27,5	25,2	3,3	87,85%	14.605	222.207	121.386	29,1	24,7	3,5	88,40%
DE116	6.943	675	268	30,2	23,2	2,8	87,11%	10.780	208.168	126.933	30,4	23,0	2,9	88,47%
DE117	2.786	660	312	29,7	26,4	1,1	87,81%	3.383	200.431	131.919	30,6	26,3	1,1	89,44%
DE118	4.859	619	241	29,5	22,4	4,5	86,93%	6.625	180.972	92.573	29,8	21,7	4,7	87,99%
DE11A	2.464	557	206	27,6	27,7	3,9	85,37%	3.066	153.405	83.995	29,1	21,1	4,4	88,73%
DE11D	3.480	568	216	31,0	27,0	5,5	83,96%	4.692	172.214	102.153	29,3	23,4	5,7	87,10%
DE121	2.283	786	409	35,2	38,4	2,0	86,39%	2.719	333.715	259.077	45,1	38,6	2,0	84,64%
DE122	9.206	703	308	31,9	35,1	1,8	87,38%	4.267	228.148	146.532	39,9	30,6	1,7	86,43%
DE123	7.689	624	236	30,6	23,9	4,0	87,90%	7.123	199.289	111.388	34,6	22,2	4,1	89,13%
DE124	3.057	599	238	31,4	29,1	2,7	88,02%	4.338	180.278	91.495	37,1	28,6	2,6	83,58%
DE125	6.004	823	522	39,1	37,0	1,3	90,31%	2.302	271.236	260.383	44,0	32,6	1,4	84,59%
DE126	13.239	637	348	32,3	32,8	1,8	88,97%	7.201	216.921	172.244	39,8	32,8	1,8	84,63%
DE128	14.734	622	255	31,5	23,1	3,5	86,31%	10.225	191.071	124.946	35,6	21,1	3,7	86,05%
DE129	5.163	521	228	31,3	26,4	0,9	89,35%	4.159	169.095	115.644	37,2	29,2	1,0	88,79%
DE12A	2.992	509	212	29,6	26,1	3,4	86,22%	4.512	121.905	92.993	34,3	21,9	3,5	87,92%
DE12B	2.702	566	202	30,1	26,0	3,2	85,57%	2.984	173.184	97.714	36,2	25,1	3,2	85,71%
DE131	5.286	783	416	36,1	30,7	1,9	90,80%	5.782	310.410	213.611	40,6	28,4	1,8	89,87%
DE132	4.454	734	290	31,9	24,8	4,7	88,78%	5.383	244.253	138.245	35,1	22,3	4,7	86,15%
DE134	5.621	606	237	29,7	29,6	5,2	84,57%	6.360	186.448	96.404	32,3	27,1	5,2	85,25%
DE136	3.728	524	212	30,9	26,7	4,1	86,13%	5.118	142.751	95.893	34,1	24,0	4,5	88,23%
DE138	5.428	710	354	32,2	28,2	5,0	87,96%	6.802	252.778	189.174	33,9	26,4	4,9	86,85%
DE139	3.644	794	340	33,7	28,8	3,8	86,81%	4.994	232.890	139.960	34,9	26,0	3,9	89,84%
DE13A	2.618	653	257	30,7	26,1	5,8	86,56%	4.555	174.305	120.033	37,2	22,5	5,8	89,89%
DE141	4.541	652	251	29,5	27,1	2,4	87,19%	5.317	203.137	114.465	32,6	25,1	2,4	87,81%
DE142	4.963	679	306	31,0	25,8	2,8	89,73%	4.233	219.071	133.875	33,3	24,2	2,8	88,25%
DE143	2.828	483	195	28,6	25,8	3,2	86,56%	2.215	139.303	91.399	34,4	23,9	3,0	86,50%
DE144	5.288	718	310	30,0	33,7	1,6	84,72%	2.903	270.707	159.709	33,8	29,8	1,7	85,49%
DE145	3.175	597	220	28,3	26,6	4,8	83,89%	2.650	184.047	87.465	30,4	22,7	4,5	86,09%
DE146	3.723	601	198	29,4	24,1	6,4	83,46%	2.668	173.564	91.293	31,1	23,2	6,5	88,12%
DE147	4.986	805	382	32,3	24,2	5,6	87,10%	6.529	279.591	196.300	36,3	22,3	5,4	86,90%
DE148	5.036	619	244	30,0	27,1	3,6	85,31%	5.695	194.340	110.659	30,6	24,1	3,5	87,28%
DE211	3.972	775	249	24,9	23,4	1,2	84,35%	3.964	267.742	120.192	26,7	22,8	1,2	85,85%
DE212	50.663	1.215	720	37,0	33,7	2,1	86,86%	44.217	540.142	421.868	41,7	29,7	2,1	83,92%
DE217	3.407	804	309	27,2	22,8	2,8	89,33%	2.344	326.614	158.473	28,1	19,3	2,9	88,03%
DE218	3.964	823	302	28,1	19,6	3,9	88,81%	2.380	354.225	191.537	33,3	19,2	4,1	87,87%
DE21B	5.084	706	254	27,9	21,2	4,7	87,73%	2.226	271.554	144.495	33,4	19,0	4,8	85,91%
DE21C	4.971	830	288	27,4	20,9	4,0	85,61%	4.495	309.796	154.305	28,7	20,0	4,1	89,22%
DE21D	2.273	704	334	33,2	25,7	3,8	89,66%	2.298	336.944	249.476	40,7	21,3	3,6	89,36%
DE21H	8.735	960	384	28,9	19,2	4,3	88,72%	6.487	379.712	221.854	31,4	17,6	4,5	84,63%
DE21K	4.606	723	296	33,0	27,2	5,8	88,52%	3.970	263.237	154.802	36,5	23,3	6,1	87,75%
DE21L	3.993	1.043	572	38,4	24,2	3,0	89,70%	2.229	505.867	402.007	48,1	23,7	3,1	86,57%
DE21M	2.306	621	247	28,8	28,6	5,4	86,25%	3.185	193.552	122.023	33,4	23,3	5,9	83,99%
DE232	6.515	682	321	32,1	29,4	1,2	89,76%	4.247	262.375	167.308	36,3	27,8	1,1	89,53%
DE252	5.857	676	303	29,3	25,6	1,0	89,17%	2.762	259.864	170.009	37,4	25,4	0,9	90,52%
DE253	6.534	645	317	33,2	41,8	1,0	88,60%	6.297	211.704	136.435	38,5	37,0	1,0	86,87%
DE254	27.396	624	303	29,2	32,3	1,7	87,36%	17.379	215.431	136.399	33,6	34,7	1,7	83,85%
DE258	2.924	593	212	26,0	24,5	1,7	86,00%	2.110	207.190	110.139	30,4	25,4	1,5	87,70%
DE259	3.665	571	229	28,2	25,0	3,6	88,82%	3.233	212.984	121.536	31,9	25,4	3,1	87,52%
DE263	4.949	639	291	27,4	26,4	1,3	89,01%	2.154	196.149	121.640	31,0	23,8	1,4	86,22%
DE271	10.797	605	267	28,3	33,3	1,4	89,53%	9.590	211.367	137.694	33,9	30,4	1,4	87,24%
DE273	2.239	560	243	28,2	30,4	0,8	88,90%	2.194	188.090	119.640	35,2	27,9	0,7	85,39%
DE276	3.673	582	215	27,6	21,5	5,2	87,25%	5.001	195.863	93.021	28,1	18,7	5,0	90,07%
DE279	3.692	610	247	29,3	25,1	2,8	88,85%	3.978	191.591	139.944	32,2	23,5	2,6	89,26%
DE27E	2.182	566	266	32,3	26,5	4,3	86,84%	4.050	174.846	109.330	32,4	20,6	4,5	85,14%
DE300	190.294	696	470	33,9	39,3	4,0	87,59%	112.287	320.533	302.990	48,2	44,3	3,7	86,28%
DE404	10.664	684	358	31,8	45,2	2,7	91,57%	4.706	320.937	240.110	43,8	49,0	2,5	87,50%
DE501	21.725	548	306	27,2	27,7	4,2	88,19%	8.909	162.947	126.127	34,1	26,3	3,7	83,14%
DE600	106.961	797	483	30,2	34,6	3,8	86,77%	34.433	386.802	344.575	42,6	36,8	3,8	85,91%
DE711	5.959	707	330	31,5	29,8	2,0	87,15%	2.394	262.110	167.632	39,8	30,7	2,1	83,78%
DE712	43.543	1.017	637	37,2	39,6	1,9	88,97%	15.162	436.931	381.443	46,2	36,2	2,0	85,91%
DE713	5.782	691	333	31,1	36,1	0,9	89,25%	3.338	227.819	143.210	35,6	30,6	0,9	87,96%
DE714	16.714	788	455	35,2	39,7	1,6	90,21%	6.550	318.723	251.711	44,1	38,7	1,7	88,24%
DE715	5.402	624	259	31,0	25,7	3,1	84,16%	3.932	187.257	111.165	35,6	23,0	3,2	84,06%
DE716	7.286	647	234	30,4	23,8	3,1	85,94%	3.528	202.212	107.206	35,3	24,1	3,2	83,24%
DE717	6.972	635	248	28,1	23,1	3,3	86,52%	3.752	186.988	101.510	33,4	21,8	3,6	88,08%
DE718	9.571	841	430	34,0	28,7	2,1	90,15%	4.877	287.357	200.724	39,3	24,8	2,1	88,43%
DE719	10.534	570	216	28,2	25,7	4,1	86,10%	6.149	174.308	105.179	32,3	24,6	4,1	83,06%
DE71A	9.660	801	385	32,7	23,2	2,2	91,05%	5.177	253.163	170.677	34,5	22,5	2,0	88,17%
DE71C	12.656	686	267	28,6	22,7	2,9	87,61%	7.940	201.505	116.821	32,0	20,5	3,1	86,77%
DE71D	6.159	640	300	31,6	28,0	4,5	89,08%	3.156	213.141	146.722	38,5	26,1	4,6	85,86%
DE71E	7.789	642	270	30,4	31,0	3,3	88,13%	3.612	214.481	132.237	35,9	27,4	3,1	84,91%
DE721	5.853	574	247	31,9	30,7	2,7	87,54%	2.448	179.104	103.731	36,1	25,5	3,0	86,96%
DE731	10.335	507	256	28,6	29,9	1,4	86,90%	2.984	167.616	123.780	37,2	33,2	1,5	89,11%
DE803	13.991	390	218	21,3	25,8	1,2	92,68%	2.516	237.819	193.154	36,2	40,2	1,3	85,15%
DE80K	4.304	401	179	23,1	38,9	5,3	85,83%	2.243	194.434	157.543	31,0	32,6	7,5	72,89%
DE80L	6.621	395	175	22,2	36,7	14,1	83,98%	4.732	223.117	179.407	29,5	28,6	17,0	76,25%
DE80N	6.126	400	199	22,7	34,5	9,6	86,79%	3.050	186.450	159.769	33,4	34,9	7,1	81,45%

Appendix 1.2: Sample size by market and descriptive statistics

NUTS3 Code	Rental Sample							Price Sample						
	N	Asking rents €/p.m.		Average			Adj. R² of GAM regression	N	Asking rents €		Average			Adj. R² of GAM regression
		Mean	Sd	Living area m²	Age	Distance to city Km.			Mean	Sd	Living area am²	Age	Distance to city Km.	
DE911	10.352	521	285	29,1	29,7	1,8	87,96%	4.248	190.364	132.693	39,7	35,7	1,9	85,54%
DE916	4.470	342	146	25,8	25,9	4,1	86,94%	3.232	56.777	62.586	32,5	22,5	4,2	80,17%
DE925	7.493	395	169	25,3	27,0	4,1	88,94%	2.354	116.639	82.345	34,4	29,8	4,4	81,36%
DE929	38.494	545	286	27,7	28,2	5,3	85,73%	17.535	173.465	131.647	36,1	29,0	5,7	82,12%
DE932	2.929	428	167	22,8	29,9	9,5	82,88%	2.235	164.115	121.035	32,7	23,0	8,6	74,18%
DE933	8.346	621	216	25,8	24,9	5,3	87,54%	2.845	180.449	89.287	29,4	21,0	5,0	82,64%
DE935	5.625	614	250	27,0	32,6	3,1	88,68%	2.122	198.378	133.326	33,5	24,2	3,1	85,21%
DE939	7.655	550	206	24,1	28,4	8,5	87,91%	2.109	165.689	101.577	30,1	24,8	8,5	77,22%
DE943	8.942	547	221	25,7	29,5	1,3	89,16%	3.847	193.307	116.156	31,5	24,2	1,2	89,79%
DE944	7.216	508	233	25,9	27,7	1,4	86,33%	2.315	166.040	115.080	33,5	27,4	1,5	88,06%
DE947	3.084	454	158	24,8	28,8	4,9	77,32%	3.193	223.282	200.189	30,2	25,5	7,6	87,80%
DE949	4.421	479	150	23,0	21,4	5,0	83,76%	2.201	144.515	66.854	31,2	18,9	5,3	82,88%
DE94E	4.873	462	172	27,4	25,1	6,7	87,04%	2.288	138.942	79.226	33,0	22,7	6,5	85,33%
DEA11	48.197	764	485	33,8	29,8	2,6	89,51%	14.361	346.917	343.120	49,2	30,9	2,7	88,74%
DEA12	30.458	362	141	20,0	23,5	2,0	81,03%	6.215	101.368	80.561	32,4	27,5	2,0	74,50%
DEA13	42.861	437	224	23,5	24,0	1,9	86,94%	9.300	160.261	149.287	37,5	29,5	2,0	86,96%
DEA14	12.487	463	224	26,2	28,6	1,4	87,47%	4.177	142.209	137.062	35,7	29,0	1,5	83,52%
DEA15	16.822	432	186	23,7	25,6	1,5	87,40%	4.537	125.480	94.192	33,2	25,7	1,6	84,70%
DEA16	9.978	460	244	25,4	25,1	1,3	89,63%	2.645	176.737	150.023	38,5	27,9	1,3	87,75%
DEA17	7.026	389	155	22,2	26,3	1,5	83,99%	2.939	124.533	87.605	30,3	30,2	1,6	84,76%
DEA19	7.106	473	193	24,8	31,7	1,2	85,69%	2.483	151.849	88.277	32,7	32,1	1,2	84,40%
DEA1A	23.247	411	185	25,3	30,1	1,7	87,32%	7.754	120.909	98.438	35,7	33,3	1,6	81,84%
DEA1B	6.439	488	161	25,3	27,0	4,6	79,87%	2.132	146.013	79.221	33,7	27,8	5,0	72,73%
DEA1C	21.819	542	230	23,8	24,1	2,6	85,62%	9.436	168.549	111.827	31,4	20,8	2,8	86,05%
DEA1D	17.552	591	271	27,1	24,0	2,3	87,42%	7.114	184.642	132.396	36,6	21,3	2,3	86,58%
DEA1E	7.525	495	199	25,4	27,9	2,8	83,96%	3.059	151.225	88.242	33,0	25,3	2,8	84,25%
DEA1F	12.977	467	175	22,3	24,8	3,3	82,74%	5.552	143.570	82.854	27,7	23,8	3,2	83,90%
DEA22	22.990	659	316	30,6	31,3	1,6	87,97%	6.060	232.342	162.284	39,7	25,6	1,6	84,38%
DEA23	49.697	723	371	30,1	29,6	2,8	86,96%	18.110	267.574	222.680	41,2	28,6	2,9	81,76%
DEA24	6.730	529	203	23,4	25,5	1,1	86,74%	2.544	172.957	102.229	26,4	21,8	1,1	82,26%
DEA26	6.342	438	153	25,0	24,5	4,6	81,57%	2.061	127.524	91.076	36,8	27,1	4,9	82,84%
DEA27	14.970	585	230	25,5	22,7	3,9	85,43%	6.323	167.534	113.117	34,8	22,0	3,9	79,52%
DEA2A	6.638	438	160	24,7	27,0	3,9	84,32%	2.047	120.107	72.347	34,3	25,0	3,8	78,37%
DEA2B	11.380	599	246	27,2	23,5	2,6	88,62%	4.275	185.473	114.378	33,9	20,8	2,4	79,76%
DEA2C	20.554	574	225	27,2	25,6	3,8	84,74%	6.641	196.823	129.451	36,9	24,4	3,7	81,01%
DEA2D	19.723	540	274	28,7	31,0	3,4	85,33%	5.556	171.360	124.823	37,6	30,0	3,8	83,56%
DEA32	17.952	336	123	19,7	22,5	1,6	81,73%	3.041	92.127	72.907	30,8	27,1	1,6	78,07%
DEA33	12.661	678	347	31,3	27,4	2,1	88,40%	5.901	236.509	176.693	35,9	25,7	2,1	83,03%
DEA34	5.257	488	178	24,6	24,2	7,1	83,79%	2.689	153.907	74.151	30,8	23,9	7,1	79,27%
DEA36	26.099	393	147	21,0	24,0	3,1	82,08%	7.420	123.426	74.447	29,0	25,4	3,3	78,58%
DEA37	8.276	467	160	23,5	25,9	4,7	84,31%	3.871	148.619	72.034	29,5	24,0	4,4	82,59%
DEA38	4.433	459	171	24,9	32,5	2,9	83,73%	2.363	135.869	78.243	29,4	25,1	3,3	81,12%
DEA41	14.246	473	218	25,0	28,5	2,4	83,81%	4.922	147.748	100.068	31,0	28,5	2,3	88,35%
DEA42	7.628	500	188	24,5	24,8	5,1	87,75%	3.598	155.900	79.291	27,2	20,4	4,6	86,77%
DEA43	6.230	419	155	24,5	27,7	2,8	83,36%	2.547	128.906	78.349	30,0	28,2	2,5	82,52%
DEA45	8.395	430	173	25,5	27,6	3,2	84,37%	4.347	133.732	91.589	32,5	26,7	2,7	81,57%
DEA46	8.275	422	148	24,6	27,5	2,7	81,44%	2.656	133.833	89.489	34,4	28,6	2,7	83,51%
DEA47	7.130	485	200	27,5	22,9	4,2	87,03%	3.850	170.203	91.449	31,1	21,1	3,8	86,53%
DEA51	18.453	411	193	23,1	24,3	1,9	85,60%	3.976	141.185	111.450	32,0	28,4	1,9	86,90%
DEA52	30.904	423	210	23,0	25,2	2,4	83,40%	7.448	150.812	108.798	34,8	29,6	2,4	85,29%
DEA55	9.489	352	123	19,7	24,9	1,4	83,50%	2.067	102.543	63.362	29,7	30,6	1,3	75,45%
DEA56	12.110	427	193	24,9	29,1	2,2	87,07%	4.150	143.941	92.709	34,9	27,7	2,2	86,07%
DEA57	3.748	398	150	26,2	27,8	5,3	85,71%	2.684	87.139	70.894	34,5	21,3	4,6	80,27%
DEA58	11.487	377	153	22,5	26,1	4,4	85,84%	3.685	106.535	61.383	31,8	24,9	4,2	76,13%
DEA5B	7.601	464	180	26,4	30,6	4,2	86,39%	2.861	147.297	88.256	33,5	27,2	4,5	82,66%
DEA5C	15.626	397	156	21,2	25,1	3,5	83,95%	4.043	132.715	78.073	28,6	24,2	3,6	83,93%
DEB21	5.732	613	284	32,8	33,6	1,4	89,22%	4.404	247.181	137.180	38,6	27,2	1,3	91,32%
DEB25	3.317	614	204	26,7	26,9	4,4	81,94%	2.075	226.624	105.879	33,0	23,7	4,7	80,04%
DEB34	4.511	573	314	32,6	30,0	0,7	89,52%	4.374	149.927	113.469	33,8	28,3	0,8	86,26%
DEB35	11.042	715	371	31,2	29,3	1,3	89,96%	4.274	233.118	149.310	34,7	25,0	1,4	88,40%
DEB3I	2.546	605	205	28,9	20,7	1,9	87,73%	2.721	181.533	86.405	34,5	21,4	1,8	88,17%
DEB3J	6.176	617	266	31,3	29,6	3,5	87,26%	3.206	205.134	133.164	37,2	24,8	3,2	86,29%
DEC01	8.258	506	220	31,1	29,7	2,7	86,64%	5.340	162.298	125.106	38,3	26,6	2,8	84,74%
DEC04	2.079	577	241	30,5	30,7	3,0	83,97%	2.201	146.844	80.328	35,2	26,2	3,1	82,13%
DED21	72.148	452	254	26,4	39,3	2,1	90,92%	15.815	206.225	144.573	40,2	46,6	2,2	88,99%
DED2E	12.517	364	180	23,5	40,9	3,5	89,98%	2.102	127.861	119.501	36,0	40,4	3,0	84,14%
DED41	33.362	317	132	21,1	35,1	1,4	87,19%	3.920	93.823	96.850	35,6	43,8	1,8	84,05%
DED45	19.407	310	126	22,0	37,8	3,6	85,12%	2.202	56.607	59.215	31,2	41,2	3,9	69,01%
DED51	92.069	426	263	27,8	40,1	1,9	91,23%	18.157	193.462	180.690	43,9	49,0	1,8	87,10%
DED52	11.051	341	175	22,7	36,2	3,6	89,05%	2.453	118.394	133.158	38,1	35,0	4,2	84,23%
DEE03	23.005	347	148	20,7	35,2	1,5	90,05%	2.261	105.142	88.013	38,4	45,7	1,2	83,10%
DEF02	18.799	448	218	22,7	30,5	2,1	85,47%	4.014	181.066	146.931	33,8	35,3	2,3	84,42%
DEF03	11.562	450	226	24,5	31,5	2,8	87,12%	2.874	180.309	144.270	33,9	31,5	4,6	85,16%
DEF07	2.802	471	236	23,9	30,8	10,1	85,99%	4.046	327.318	245.036	34,0	26,2	15,1	83,06%
DEF08	5.276	481	229	25,0	27,6	9,6	86,14%	8.241	187.510	149.556	32,9	23,1	9,8	76,01%
DEF09	14.108	576	232	24,1	26,7	3,8	87,10%	5.629	177.729	106.844	29,3	23,6	5,5	88,67%
DEF0B	6.820	448	185	24,4	26,9	6,0	84,56%	2.350	154.543	117.693	34,8	27,7	7,2	83,53%
DEF0D	11.789	578	219	22,7	21,7	2,6	87,08%	3.826	174.225	99.753	29,0	21,0	2,8	82,87%
DEF0F	7.952	634	247	25,8	26,1	3,4	87,75%	3.700	208.836	112.922	29,7	22,4	3,3	85,73%
DEG01	13.052	451	229	25,8	39,3	1,8	91,32%	3.058	169.346	117.973	37,7	45,6	2,1	86,61%