Price Uncertainty and Returns to Housing *

Francisco Amaral †

please click here for the latest version

January 9, 2024

Using a novel micro-level dataset that encompasses the universe of housing transactions in major German cities and spans the past four decades, I document significant differences in the predictability of sales prices across individual houses. On average, residential properties with greater price uncertainty have lower transaction prices, generate higher net rental yields, and yield larger total returns. These properties are traded in smaller and more illiquid markets, implying that price uncertainty and, consequently, return premia arise in markets with less efficient buyer-seller matching. I rationalize these findings within a bargaining model where a risk-averse investor faces greater resale price uncertainty due to higher matching frictions.

Keywords: housing markets, liquidity, asset returns
JEL codes: E21, G11, G52, N90, R21, R31

*This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866. For extremely helpful feedback I thank Filipe Correia, Mintra Dwarkasing, Stephanie Ettmeier, Ricardo Gabriel, Jonathan Halket, Christian Hilber, Chi Kim, Matthijs Korevaar, Martin Kornejew, Dmitry Kuvshinov, Timothy Meyer, Cathrin Mohr, Marco Pagano, Monika Piazzesi, Giovanni Puopolo, Francisco Queirós, Ricardo Reis, Björn Richter, Farzad Saidi, Christophe Spaenjers, Martin Schmalz, Moritz Schularick and Kaspar Zimmermann. I also thank conference participants at the "Of Shifts and Shocks: Macrohistory Workshop" workshop in Bonn, the 2023 ECHOPPE housing conference in Toulouse, the 2023 Portuguese Economic Journal conference in Braga, the 2023 International AREUEA Conference in Cambridge, the ReCapNet 2023 conference in Mannheim and the 2023 UEA North-American Meeting in Toronto as well as seminar participants at the finance seminar of the University of Groningen, the finance seminar at the Amsterdam Business School and the macro-finance and finance seminars of the University of Bonn. All remaining errors are my own.

†ECONtribute & MacroFinance Lab, University of Bonn, amaralfranc@gmail.com.

1 Introduction

Buying a house is the most important financial decision that households make in their lifetime. Understanding the factors driving the willingness to pay for a house is, therefore, of great importance. While several papers have examined the role of location, credit conditions, income and other factors (e.g. Duranton and Puga, 2015; Van Nieuwerburgh and Weill, 2010; Mian and Sufi, 2009), in this paper, I focus on how uncertainty about the market value of a house affects its transaction price and return.

This uncertainty, defined as the expected variance of the distribution from which the price of a house might be drawn at a point in time, differs substantially across individual houses (Jiang and Zhang, 2022; Kotova and Zhang, 2021). There are different factors that can contribute to price uncertainty in housing markets. The heterogeneity and illiquidity of houses introduce uncertainty about the outcome of the bargaining process between buyers and sellers (Goetzmann et al., 2021; Sagi, 2021). Additionally, houses are generally held for extended periods of time, rendering them susceptible to various shocks that can impact their fundamental value (Han, 2013; Sinai and Souleles, 2005), making them harder to value ex-ante.

However, we still know little about the extent to which price uncertainty influences the trading decisions of buyers and sellers and, consequently, impacts housing prices and returns. This is a challenging task, as it requires highly granular data on housing markets. Using a newly-collected transaction-level dataset covering the universe of apartment transactions over the last 40 years in four of the largest German cities[1], I am able to shed light on this issue. I find evidence that this uncertainty is priced in housing markets: apartments with higher price uncertainty trade, on average, at lower prices. The magnitude of the effect is large, I estimate that apartments with high price uncertainty trade at a price that is, on average, 5% lower than comparable apartments with lower price uncertainty, a result in the same order of magnitude as existing estimates of foreclosure discounts (Conklin et al., 2023).

Nevertheless, I find that these apartments can still be rented out at standard rates, resulting in higher rental yields, as measured by the ratio of net rental income to transaction price. I then measure total returns as the sum of rental yields and capital gains at the apartment level. I find that apartments with greater price uncertainty tend to yield higher total returns, due to higher rental yields. Again, these differences are economically significant. The data suggests an average annual return premium of 50 basis points for apartments exhibiting greater price uncertainty. This is approximately 10% of the average yearly total housing return in Germany over the past four decades.[2]

I rationalize these findings through the lens of a bargaining model. The model

[1] Berlin, Hamburg, Cologne and Duesseldorf.

[2] I take the estimates for Germany from Amaral, Dohmen, Kohl, et al. (2021).

features a risk-averse housing investor who acquires a property for renting and future sale. Consistent with my empirical results, the model predicts that properties with greater re-sale value uncertainty will transact at lower prices and have higher rental yields. Assuming that matching frictions drive the price uncertainty, the model also predicts that total returns are higher due to increased rental yields. Confirming the assumptions of the model, I find that apartments with greater price uncertainty are traded in smaller and less liquid markets, which suggests that matching frictions in the housing market underlie the uncertainty surrounding transaction prices.

The primary data source I use in this paper is a transaction-level dataset introduced in Amaral, Dohmen, Schularick, et al. (2023), which contains detailed information on the universe of residential real estate transactions in large German cities over the last half-century. The dataset provides comprehensive information on property characteristics as well as transaction types. This feature enables me to control for differences in observable property characteristics and effectively identify transactions of the same properties over time. The data set also includes information on the realised rental income after costs, which allows me to build net rental yields at the property level.

My focus is on the most prevalent housing type in large German cities: apartments. The market for apartments in large German cities provides an ideal setting to examine the relationship between price uncertainty and housing returns. In contrast to U.S. cities (Glaeser and Gyourko, 2007), there is minimal segmentation between the home-ownership and rental markets for apartments in large German cities. Owner-occupied and rental apartments exhibit only marginal differences in terms of their location and characteristics.

Following Jiang and Zhang (2022), I measure price uncertainty at the apartment transaction level as the predicted variance of the pricing error from a hedonic housing price model. I provide evidence that the error in the hedonic model is not driven by omitted variable bias or is simply noise. I do this by showing that the errors are spatially independent and that their magnitude is highly persistent over time within apartments.

Then, I introduce the measure of price uncertainty into a hedonic model of house prices and show that higher uncertainty significantly predicts lower transaction prices for all cities in my sample. Importantly, I also show that properties with higher price uncertainty do not have lower demand, as proxied by online search behaviour, thus reinforcing that uncertainty is being priced in. These effects are economically very relevant: transitioning from the lowest to the highest quintile of price uncertainty predicts a decrease in the final transaction price by approximately 5% to 7% for properties transacted in the same neighborhood and year-quarter, while controlling for property characteristics.

However, the effect is notably weaker for rents. Rental rates are similar across properties with different price uncertainty. The rationale behind this outcome is that such apartments face greater illiquidity in the sales market compared to the rental market. This aligns with the fact that in large German cities, the rental market is larger and more liquid than the sales market. I then find that properties with higher price uncertainty exhibit, on average, higher rental yields. Transitioning from the lowest to the highest quintile of the price uncertainty distribution predicts an increase of between 35 and 60 basis points in rental yields for transactions during the same year-quarter in the same neighborhood, while controlling for property characteristics.

By identifying repeated sales of the same apartments over time, I am able to construct property-level capital gains, which, when combined with rental yields, provide measures of property-specific total returns. I then find that apartments with higher price uncertainty experience, on average, the same rate of price appreciation as other apartments. In other words, the data shows that price uncertainty is uncorrelated with the level of capital gains in housing markets. Finally, I show that properties with higher price uncertainty have, on average, higher total returns, driven by rental yields.

To make sure my results are not driven by time-varying market conditions, I employ portfolio sorting methods and hedonic regressions to construct a time series of prices and returns for properties with high and low price uncertainty. I show that portfolios with high price uncertainty yield higher total returns, and this return premium is not driven by heterogeneous exposure to systemic risk as measured by the returns on city market portfolio. Additionally, I conduct a battery of robustness tests to ensure that my results are not influenced by measurement error and to exclude alternative mechanisms. Importantly, I do not find that properties with higher price uncertainty have lower demand based on online search behaviour.

I map these empirical findings to a bargaining model featuring a risk-averse investor, who faces uncertainty regarding the future rental income and resale value of the house. Intuitively, the investor's risk aversion explains why properties with greater price uncertainty exhibit lower transaction prices and higher rental yields. More interestingly, the model reveals that the impact of price uncertainty on capital gains depends on the source of that uncertainty. Under the assumption that price uncertainty arises from matching frictions, the model predicts that increased price uncertainty does not result in higher capital gains, as supported by the data.

In line with the mechanism of my model, I find that apartments with higher price uncertainty are traded in smaller markets. More specifically, there is a lower number of similar properties on the market, making it more challenging to price these apartments.[3] Additionally, I show that properties with higher levels of price uncertainty are less liquid. On average, they have a longer expected time on the market and a lower probability of sale. Furthermore, the final transaction price is, on average, significantly lower when compared to the original asking price.

[3] This concept builds on the idea of house atypicality (Haurin, 1988).

Related literature This paper contributes to the literature on price dispersion in housing markets (Giacoletti, 2021; Sagi, 2021; Piazzesi and Schneider, 2016) by showing how the interaction between rental and sales markets affects transaction prices and returns. Jiang and Zhang (2022) show that price uncertainty impacts the quality of housing as collateral and, consequently, negatively affects the credit conditions offered by banks. I complement this work by showing that price uncertainty also influences transaction prices and returns.

This paper also speaks to the literature on the risk factors driving returns to housing (e.g., Demers and Eisfeldt, 2022; Amaral, Dohmen, Kohl, et al., 2021; Han, 2013; Cannon et al., 2006). While this literature has primarily focused on identifying systemic housing risk factors, this paper provides evidence that property-specific idiosyncratic risk factors are priced in housing markets.

My research also connects to the literature on decentralized asset markets. Duffie et al. (2007) develop a search-and-bargaining model for financial assets traded in decentralized markets to understand how trading frictions affect asset prices. Gavazza (2011) constructs a model of the commercial aircraft market to illustrate how market thickness affects liquidity and prices. The author shows that airplanes traded in thinner markets typically trade at lower prices, mirroring the effect I find for housing assets. The interplay between liquidity in the rental and sales market aligns with the theoretical framework presented in Pagano (1989), who examines how trading frictions can influence the relationship between market size and asset liquidity across various markets and thus determine the distribution of trading activity.

The rest of the paper is structured as follows. Firstly, I present the data and provide evidence on market liquidity in the German housing market. Secondly, I describe the measurement of price uncertainty and present the empirical framework. Thirdly, I show that properties with higher price uncertainty are sold at a discount and have higher returns. I present the data and provide evidence on market liquidity in the German housing market. Fourthly, I derive a theoretical framework and characterize the optimal bid of a risk-averse investor when facing uncertainty about future cash-flows. Fifthly, I provide empirical evidence linking price uncertainty to market size and liquidity at the property level. The last section concludes.

2 Data

In the empirical analysis of this paper, I combine three distinct datasets. The first dataset comprises comprehensive transaction-level information on the universe of real estate transactions in major German cities dating back to the 1980s. I employ this dataset to estimate price uncertainty and construct two market liquidity measures for four large German cities. The second dataset, which I developed from scratch, contains net rental values in German cities based on property size, age, and location, spanning from the 1980s onwards. I use this data set to provide rental income information for those observations for which it is missing in the main transaction data set. The third dataset encompasses real estate advertisements in Germany since 2010. I merge this dataset with the transaction-level data to obtain additional liquidity measures at the transaction level. I will now provide a more detailed description of each dataset.

Transaction-level data set - This data set, introduced in Amaral, Dohmen, Schularick, et al. (2023), consists of transaction-level data encompassing all residential real estate transactions in 20 major German cities dating back to the 1960s. The underlying data is sourced from the local real estate expert committees, known as "Gutachterausschüsse," who receive comprehensive information about each real estate transaction from notaries. This valuable information encompasses the transaction price, date, as well as various property characteristics such as size, location, and building year. Additionally, it includes details about the type of transaction, including whether it was conducted at arm's length or not. In many instances, this data is further enriched by gathering additional information directly from buyers and sellers regarding the property, such as whether the property has a garage or not. The scope of this novel data set is, to the best of my knowledge unique, in that existing transaction-level data sets only contain representative sales information for a shorter period of time.[4]

For the main empirical analysis, I only use data on sales of apartments. The reason for this is twofold. Firstly, the housing stock in large German cities is mostly composed of apartments and therefore there are considerably more apartment sales than of other types of housing. This contrasts with most cities in the U.S., where the predominant type of housing is single-family. Secondly, apartments are more homogeneous types of housing than single-family or multi-family housing. This increases the statistic precision of the hedonic analysis, which I will carry on later.

Before conducting the empirical analysis, I ensure the integrity of the transaction data by meticulously filtering out non-arm's length sales. This encompasses a range of transactions, such as property sales between relatives, leaseholds ("Erbbau"), package sales involving multiple properties sold together, sales of social housing, transactions involving official government institutions at the local or federal level, foreclosures and any sales flagged by the "Gutachterausschüsse" as not aligning with genuine market prices. Additionally, I exclude all transactions that have missing information regarding the date, sales price, size, location, or building year. To enhance the sample quality, I implement supplementary cleaning procedures. Specifically, I eliminate "house flips" and cases where the reported sale price appears anomalous, as well as duplicates. This

[4]For example, in the case of the U.S. existing data sets, such as Corelogic, only have a reoresentative sample since the 2000s.

is accomplished by removing all transactions of a property if it undergoes two sales within a year or if its annualized appreciation or depreciation exceeds 40% for any given pair of sales. This approach adheres to the standard methodology established in the literature (e.g. Giacoletti, 2021). Please note that in all specifications, the sales price measure I utilize is net of additional costs that do not directly pertain to the value of the structure and land of the respective property. In other words, the sales price is adjusted for inventory costs (e.g., if the kitchen is included in the apartment sale) or additional infrastructure expenses (e.g., when the owner of the apartment is entitled to use a parking spot or garage). The valuation of these additional costs is specified in the contract and is also reported by the "Gutachterausschüsse". Table 7 presents the summary statistics of the data by city after the cleaning procedures.

Rental values data set - To complement the rental income information provided by the "Gutachterausschüsse", I collected net rental value data from an independent source. I then merge it with the transaction-level data. The rent data is obtained from the so-called "Mietspiegel", which provides rent per square meter estimates for apartments in German cities based on factors such as size, age, and location of the apartments. The rent estimates are net of utilities and maintenance costs. These rent estimates are then matched with the transaction data, conditional on the size, age, and location of the property. The specific details regarding the data source and the matching process are provided in Appendix J. In the empirical section, I will present the results for both the full sample, where transactions were matched with rents based on characteristics, and for the subsample in which both transaction prices and rental income are observed for the same property at the same point in time.

Real estate advertisement data - To be able to measure asset liquidity at the transaction property level, I combined transaction-level data with advertisement data. The advertisement data was sourced from Value AG, a German real estate company that has consistently been collecting online real estate advertisements and integrating them with data from local real estate agents, resulting in a comprehensive and extensive data set that covers the period from 2012. This data set from Value AG encompasses crucial information on property characteristics obtained from the advertisements. Leveraging this information, I employed a nearest neighbor algorithm to match the transaction data with the advertisement data.

2.1 Liquidity in German Housing Markets

Several studies examining the structure of decentralized asset markets have provided evidence that larger markets enhance the efficiency of matching between buyers and sellers. This effect results in reduced price dispersion and decreased uncertainty surrounding the value of traded assets (Sagi, 2021; Gavazza, 2011). In this section, I present various pieces of evidence regarding the size and liquidity of the rental and sales markets for apartments in major German cities. Overall, the empirical evidence indicates that the apartment rental market is significantly larger, thicker, and more liquid than the apartment sales market.

Using data from the largest real estate online platform in Germany for the period between 2010 and 2018.[5] For this analysis, I exclude from the original dataset all ads with missing information about price, rent, or size. Additionally, I also remove ads flagged as potential duplicates. This issue may arise when an ad is deliberately removed and then re-uploaded shortly afterward to increase its visibility.

In Table 1, I present various indicators for the rental and sales markets of apartments in the four cities in the sample. The second column shows the homeownership rate in 2010 by city.[6] On average, only one-fifth of the population actually resides in owner-occupied housing, while the rest rents. Similar to other developed countries, homeownership rates in German cities are substantially below the national average, which has remained around 45% over the last decade (Kohl, 2017).

The third and fourth columns display the average number of sales and rental ads for apartments per year by city.[7] On average, each year there are four times more rental ads than sales ads, confirming that the rental market is not only larger in terms of its inventory but also in terms of the number of properties available.

However, what truly matters is the number of potential buyers per ad, i.e., the market thickness. To approximate the market thickness, I use data on the number of times the seller was contacted by potential buyers through the website for a specific ad. For clarification, this metric is not equivalent to hits per ad. To contact the seller, a potential buyer (or renter) must click on the ad and then select the 'Contact Seller' option. The average number of contacts for sales and rental ads is displayed in columns 5 and 6. On average, rental ads attract four times as many customers showing explicit interest compared to sales ads, indicating that the rental market is considerably thicker.

Finally, in columns 7 and 8, I present the average number of days that sales and rental ads remain on the website. Not surprisingly, we observe that sales ads stay, on average, twice as long on the website, suggesting that the time on the market is substantially shorter for apartment rentals than for apartment sales.[8] Overall, there is considerable evidence that the rental market for apartments is larger and more liquid than the sales

[5] The data is originally from `www.immoscout.de` and was provided by RWI and Immobilienscout24 (2021). The data is available from 2007 onwards, but due to issues in the identification of duplicate ads, it only becomes representative from 2010.

[6] The data is obtained from Eurostat.

[7] I determine the year based on the initial date of the advertisement.

[8] Furthermore, rental ads potentially have a much higher chance of actually resulting in a rental contract than sales ads have of resulting in a sales contract.

market in large German cities.

Table 1: *Summary statistics for apartment sales and rentals by city, 2010-2018*

City	Homeownership rate (in %)	# ads per year		Contact clicks per ad		Duration of ads (in days)	
		Sales	Rentals	Sales	Rentals	Sales	Rentals
Berlin	13.7	25767	109804	4	32	26	18
Cologne	26.7	7086	27517	12	50	40	20
Duesseldorf	21.6	6567	28544	10	28	35	21
Hamburg	21.4	9883	27657	8	33	36	17

Note: The homeowneship rate refers to 2010 and the data source is Eurostat. The rest of the data refers to the period 2010-2018 and is based on own calculations with data from www.immoscout.de, which provided to me by RWI and Immobilienscout24 (2021). "Duration of ads" measures the days between the day the ad was posted and the day the ad was removed. "Contact clicks per ad" refers to the average amount of times that the seller was contacted by potential buyers via the website about the ad.

3 Measurement and Empirical Framework

Following the real estate literature (e.g Kotova and Zhang, 2021), I measure idiosyncratic price deviations at the apartment transaction level as the difference between the transaction price and the expected market value, which is determined using a hedonic regression estimated on apartment repeat sales.[9]

For each city separately, I regress the natural logarithm of the transaction price for property i in year-quarter tq on a time-invariant apartment fixed effect, y_i, year-month fixed effects, η_{tm}, year-quarter-neighborhood fixed effects, $\kappa_{n,tq}$, and a second-order polynomial function of apartment characteristics (age and size) interacted with year fixed effects, $f_c(x_i, ty)$:

$$ln(p_{i,tq}) = y_i + \eta_{tm} + \kappa_{n,tq} + f_c(x_i, ty) + u_{i,tq}, \quad (1)$$

where $u_{i,tq}$ is a mean-zero error term with variance σ^2. Specification (1) combines elements of repeat-sales and hedonic models of housing prices. The apartment fixed effect term, y_i, absorbs all features of an apartment, observed and unobserved, which are time-invariant, such as a balcony facing the sea or the floor number. The η_{tm} and $\kappa_{n,tq}$ terms absorb parallel shifts in housing prices in a city and in neighborhoods over time, for example due to gentrification.[10] The $f_c(x_i, ty)$ term allows apartments with different observable characteristics x_i to appreciate at different rates: for example, the $f_c(x_i, ty)$ term allows larger apartments to appreciate faster than smaller apartments, or newer apartments to appreciate faster than older apartments. I use an additive functional form

[9] A very similar approach to estimate the market value is employed in Kotova and Zhang (2021); Buchak et al. (2020).

[10] More precisely, I use the definition of "Stadtbezirke" to divide the cities in different neighborhoods.

for $f_c(x_i, ty)$:[11]

$$f_c(x_i, ty) = g_c^{sqmt}(sqmt, ty) + g_c^{yrbuilt}(yrbuilt, ty) \quad (2)$$

The functions $g_c^{yrbuilt}$ and g_c^{sqmt} are interacted second-order polynomials in their constituent components. The squared terms of the polynomial function accommodate the possibility that the effect of size and age on transaction prices may vary along the distribution. For instance, larger apartments might appreciate at a different rate than smaller apartments, and this effect may not follow a monotonic pattern. Please note that for Hamburg, information identifying the exact apartments was not available. As such, I use building fixed effects instead of apartment fixed effects to measure price deviations for Hamburg. For more details please refer to Appendix C.3.

The residuals, $u_{i,tq}$ from equation (1) quantify the discrepancy between the transaction price and the expected market value of the apartments. Consequently, the squared residuals serve as a measure of price dispersion at the apartment transaction level. Table 2 displays the summary statistics for the apartment repeat sales for all cities in the sample.[12] In terms of the standard deviation of the residuals, Cologne has by far the lowest level, 10.1%, followed by Duesseldorf with 14.9%, Hamburg with 17.7% and Berlin with 18.2%.[13] Using the same method Kotova and Zhang (2021) estimate the standard deviation of residuals for single-family houses in California to be in the range between 11.1% and 13.5% depending on the city.

3.1 Stylised facts about price dispersion

While the concept of price dispersion is clear in theory, the infrequent transactions of properties, coupled with the significant heterogeneity among houses, complicates the empirical task of estimating price dispersion. Therefore, in this section, I first present additional evidence regarding the distribution of estimated price dispersion across space and over time to validate the estimates. Secondly, I will discuss several potential biases that could arise in equation (1) and demonstrate that the main results of the paper are not influenced by these biases. For the sake of brevity, the results in this section will be referenced in the text but will be presented in Tables and Figures in appendices D and I.

Distribution of dispersion across space and over time By definition, the market value of a property should reflect the value of common property characteristics in the market. In other words, the residuals in Equation (1) should capture the cross-sectional variation

[11] In principle, it would be better to estimate a fully interacted polynomial in all house characteristics. However, as argued by Kotova and Zhang (2021), that is not computationally feasible.

[12] Table 7 in the Appendix presents the summary statistics for all apartment sales, i.e. not just the repeat sales.

[13] Figure 9 in the Appendix plots the distribution of residuals, u_{it}, with the mean and standard deviation for the cities in my sample separately.

Table 2: *Summary statistics for apartment repeat sales by city*

	Berlin					
	N	Mean	SD	P25	Median	P75
Price (thousand €)	67195	186	162.3	83.9	135	229
Size (m^2)	67195	74	28.8	53.5	67.3	89.2
Construction year	67195	1932	38.3	1903	1912	1961
Residuals, $u_{i,tq}$ (%)	67195	0	18.3	-10.9	0	11.3
Rental yield (%)	67195	3.5	1.7	2.3	3.2	4.3
	Hamburg					
	N	Mean	SD	P25	Median	P75
Price (thousand €)	49506	306	263.6	130	234	394
Size (m^2)	49506	76	30.8	54	70	91
Construction year	49506	1974	41	1953	1978	2012
Residuals, $u_{i,tq}$ (%)	49506	0	17.8	-8.4	0	9.5
Rental yield (%)	49506	4.2	2	2.9	3.8	5
	Cologne					
	N	Mean	SD	P25	Median	P75
Price (thousand €)	49963	140	103.6	75	112.5	170
Size (m^2)	49963	69	24.7	52.4	67	84
Construction year	49963	1968	23	1959	1971	1983
Residuals, $u_{i,tq}$ (%)	49963	0	13.6	-7.8	0	7.9
Rental yield (%)	49963	5.7	2.4	4.1	5.4	6.8
	Duesseldorf					
	N	Mean	SD	P25	Median	P75
Price (thousand €)	25238	156	136.2	76.7	117.4	185
Size (m^2)	25238	74	28.7	54	70	90
Construction year	25238	1961	24.7	1953	1962	1976
Residuals, $u_{i,tq}$ (%)	25238	0	14.8	-8.4	0	8.7
Rental yield (%)	25238	5	2.2	3.7	4.6	5.8

Note: Table reports summary statistics for all apartment resales for Berlin (1986-2022), Hamburg (2002-2022), Cologne (1989-2022) and Duesseldorf (1984-2022). Note that before 1992 the data for Berlin refers only to West-Berlin. Prices are in nominal terms. Please note that in the case of Hamburg, the total number of sales does not refer to repeat-sales, as data on the number of the apartments is missing in the original data. Please refer to Appendix C.3 for more information.

in the idiosyncratic component of housing prices. This aligns with the bargaining model in Section 2, where, in the first period, the investor encounters uncertainty surrounding the idiosyncratic component of prices in the final period. To test whether the residuals are idiosyncratic, I now examine their spatial distribution. If Specification (1) is correctly defined, then we should expect the residuals to be spatially independent. To test for this, I estimate spatial correlation in the residuals, $u_{i,tq}$ using Moran's I. A positive Moran's I indicates that apartments with positive residuals are surrounded by other apartments with positive residuals. In Figure 10, I plot Moran's I for the residuals and the transaction prices. In contrast to the transaction prices, the results suggest that the residuals are spatially independent, as I cannot reject the null hypothesis of no spatial autocorrelation. In Appendix D, I provide a more detailed explanation of how Moran's I is estimated and present the results of this analysis.

In the bargaining model in Section 2, I assume that the investor knows the variance of the sales price in the final period. In other words, the investor is aware of the price dispersion of a given house. To justify this assumption, it is necessary for the estimated price dispersion, $u^2_{i,tq}$, to be predictable over time. Specifically, I test for all pairs of transactions in the data set whether the variance of the residuals at the point of sale predicts the variance of the residuals at the point of re-sale:

$$u^2_{i2} = \beta_1 u^2_{i1} + \beta_2 hp_i + \kappa_{nt} + \lambda_m + \epsilon_{it}, \tag{3}$$

where u_{i2} and u_{i1} are the idiosyncratic price residuals at the points of re-sale and sale respectively of property i. hp_i measures the holding period in months for property i, while δ_m are monthly fixed effects and κ_t are neighborhood fixed effects. The results can be found in Table 8, which shows that properties sold and re-sold in the same neighborhood and in the same month show considerable persistence in their idiosyncratic price dispersion. An increase in one standard deviation of the sales' price dispersion predicts an increase in 0.66 standard deviations in the resale price dispersion. One concern is that these results are being driven by the buyers, if a specific buyer is bad at pricing a house at the moment of sale, then probably as well at the moment of re-sale. This could potentially explain the high level of persistence in the variance. To address this concern, I show that the persistence in variance is also strongly positive and statistically significant when testing the relation between first and third sale. The results can be found in Table 9 in Appendix D. Additionally, the cross-sectional correlation at the point of sale and re-sale of idiosyncratic shocks is 0.66, which is higher that than most risk factors used in the stock pricing literature (Bali et al., 2016). The results can also be found in Appendix D.

Biases The baseline regression model in (1) may yield biased results due to several factors. Therefore, I highlight potential issues that may arise and explain how I address them in the robustness analysis presented in Section I of the appendix.

Firstly, in regression (1), I incorporate apartment fixed effects. However, this approach may pose challenges since most properties are sold only a few times within the sample period. This could potentially lead to an "incidental parameters problem" (IPP) problem, whereby the estimate $\hat{\sigma}^2$ would be inconsistent. Moreover, it is crucial to consider whether properties sold more than once are representative of the entire population of transacted properties. If these repeat sales do not accurately reflect the broader sample, the generalizability of my results could be compromised. To address these concerns, I run regression (1) while excluding the apartment fixed effects. The findings of this analysis are presented in Section I of the Appendix, where I show that the main results remain consistent and robust.

Secondly, in my baseline regression analysis, I do not explicitly account for the influence of varying holding periods on the sales prices. It has been demonstrated by Giacoletti (2021) that longer holding periods are correlated with greater idiosyncratic shocks. To ensure that this factor is not driving my results, I conduct an additional regression analysis that incorporates holding period fixed effects. The findings of this analysis demonstrate that the results remain robust and unaffected by the inclusion of holding period fixed effects.

Thirdly, it is important to consider that sales prices may be influenced in a systematic manner by the characteristics of both buyers and sellers. If certain types of households, such as affluent ones, tend to concentrate in specific areas, then regression (1) already accounts for this by incorporating location fixed effects. However, it is also possible that businesses or large investment funds have the ability to negotiate more favorable prices compared to individual households. To address this potential influence, I perform an additional regression analysis that includes buyer and seller fixed effects, along with their interaction term. These controls account for whether the buyers or sellers are private companies or households. The robustness analysis presented in Section I demonstrates that the results remain unaffected even after incorporating these additional controls.

3.2 Empirical Framework and Identification

The theoretical framework outlined in Section 2 guides the empirical tests concerning the effects of price dispersion on transaction prices and returns. While in the model, the investor adjusts the optimal bid based on the expectation of price dispersion, the price dispersion measured in the previous section reflects realized price dispersion. In other words, the residuals, $u_{i,tq}$, from Equation (1) are only observed ex-post and thus represent a biased measure of investors' expectations. Therefore, I approximate the information set available to a potential investor about a specific property before purchasing it. To achieve this, I employ the method introduced in Jiang and Zhang (2022). Using the observable characteristics of the properties and the transaction values of similar properties that were sold in the same period, I obtain a prediction of idiosyncratic price dispersion at the property level. More specifically, I estimate the following regression:

$$u^2_{i,tq} = g_c(x_i, tq) + \epsilon_{it} \tag{4}$$

$$\hat{\sigma}^2_{i,tq} = \hat{g}_c(x_i, tq), \tag{5}$$

where u^2 are the squared residuals estimated from equation 1 and $g_c(x_i, tq)$ is a smooth function of observable property characteristics interacted with quarter fixed effects. The characteristics are size, age and location and g is an additive function that takes the form:

$$g_c(x_i, tq) = g_c^{loc}(tq, \kappa) + g_c^{sqmt}(tq, sqmt) + g_c^{yrbuilt}(tq, yrbuilt), \quad (6)$$

where κ are neighborhood fixed effects and g^{sqmt} and $g^{yrbuilt}$ are second-order polynomials that interact time quarter fixed effects with size and year of construction respectively. I then use the predicted values, $\hat{g}_c(x_i, t)$, as an estimate of the property transaction level predicted price dispersion.

Cross-sectional variation The objective of the empirical analysis in this paper is to investigate whether expected price dispersion can predict prices and returns in the housing market, taking into account property characteristics. Consequently, the challenge in this context lies in the potential correlation between predicted dispersion and property characteristics that can impact transaction prices. To address this challenge, the analysis will involve comparing contemporaneous transaction prices of properties that are similar in size, age, and location but differ in terms of their predicted dispersion. To clarify, this section of the paper focuses on exploring the cross-sectional variation in the data. This approach differs from most asset pricing settings, which concentrate on more liquid asset classes. In the context of housing markets, analyzing the time variation of different properties is often impractical because each property is typically sold only every few years, and properties vary significantly in their holding periods. Since the measure of predicted dispersion will be derived from estimated coefficients, the empirical results in this section will rely on two-stage least-squares (2SLS) regressions, in which :

$$\text{Stage 1:} \quad u_{i,tq}^2 = g_c(X_i, tq) + B_X X_i + \eta_{tm} + \kappa_{n,tq} + e_{i,tq} \quad (7)$$

$$\text{Stage 2:} \quad y_{i,tq} = \gamma \hat{u}_{i,tq}^2 + B_X X_i + \eta_{tm} + \kappa_{n,tq} + \epsilon_{i,tq}, \quad (8)$$

where X_i is a vector of property characteristics that include size and age, κ_{nt} are year-quarter fixed effects and μ_{dt} are year-neighborhood fixed effects. The dependent variable $y_{i,tq}$ can refer to the transaction price, the net rent at the time of the transaction or the rental yield.[14]

[14]Please note that all regression output results presented in this paper will display standard errors that have been adjusted for the use of an estimated regressor in the second stage, achieved through the utilization of a sandwich variance estimator.

Alternative identification To further reinforce identification, I will also instrument the price dispersion using measures of market thickness at the property level that do not directly depend on property characteristics. As demonstrated in various theoretical and empirical papers, thicker markets tend to exhibit less price dispersion (e.g. Sagi, 2021; Gavazza, 2011). Extending this concept to housing markets, I create two measures of market liquidity at the property-transaction level to predict price dispersion. The proposed instruments are based on the premise that each property may potentially have its own market. Given the nature of the data I am working with, these measures will primarily capture sellers' market liquidity. However, we anticipate that general equilibrium factors will influence both sellers' and buyers' market liquidity, resulting in a high degree of correlation between these two measures across different properties and over time.

Following Jiang and Zhang (2022), I build an instrument based on the distance of the properties' i characteristics to the mean characteristics of the properties' sold in the same city and within the same year. This measure captures the degree of thinness in the local property market for property i. For instance, there will be less suply and demand for an old and large apartment in a city predominantly composed of new and small apartments. The instrument is then built as:

$$Z_{it}^{m} = (X_{it}^{m} - \bar{X}_{ct}^{m})^2, \ \forall m \in \{size, age, location\}, \tag{9}$$

where *size* measures the living area of the apartment in square meters, *age* is the building year of the apartment and *location* is the geographical location of the apartment given by its latitude and longitude. The instrument for location measures the distance between properties' i latitude and longitude and the average latitude and longitude of all the properties being sold within the city in year t. Given that distances for size, age and location are all measured in different units, the distances are all standardized to have mean 0 and standard deviation of 1 for each year t.

In addition, I construct a market thickness measure based directly on the relative frequency of specific combinations of apartment characteristics. The aim is to capture how often a particular combination of characteristics appears on the market at a given point in time. Typical combinations of characteristics will appear more frequently on the market, indicating a higher supply and demand for those specific characteristics. To achieve this, I divide the distribution of size, age, and location into eight equally sized bins for the entire sample. Then, for each year t, I calculate the relative frequency of each bin by dividing the total number of transacted properties with those specific characteristics by the total number of transactions in that year:

$$Z^m_{it} = \frac{obs_{it}}{obs_t}, \; \forall m \in \{size, age, location\} \tag{10}$$

I perform two-stage least squares regressions in which observed idiosyncratic price dispersion is instrumented by the measures of market liquidity for the different property characteristics, Z^m_i. By introducing the instruments separately for each characteristic, I am enabling each characteristic's illiquidity to have a distinct impact on the price dispersion.

Time-series variation Housing differs from more liquid asset classes in that houses are transacted very infrequently. Consequently, the transaction price of a house is not observed every period. To analyze the time-series variation in the measure of predicted dispersion, I employ a portfolio sorting analysis, where I sort transactions into specific portfolios each period based on the level of predicted dispersion in the apartment transactions. I then utilize fixed-effects panel regression methods to estimate the impact of predicted dispersion on expected returns.

I first sort all transactions into equal-sized portfolios based on their predicted dispersion $\hat{\sigma}_{it}$ every quarter. Given the size of the sample I first construct six equally sized portfolios. For each one of the p portfolios I estimate total quarterly housing returns as the sum of capital gains and rental returns:

$$\text{Total housing return}_{p,tq} = \underbrace{\frac{P_{p,tq} - P_{p,tq-1}}{P_{p,tq-1}}}_{\textbf{Capital Gain}} + \underbrace{\frac{R_{p,tq}(1 - c_{tq})}{P_{p,tq-1}}}_{\textbf{Net Rent Return}} \tag{11}$$

where $P_{p,tq}$ is the hedonic transaction price in portfolio p in quarter tq, $R_{p,tq}$ is the hedonic rental payment and c_{tq} are utility, maintenance and vacancy costs as a share of the rent. To estimate the value of the hedonic price and the hedonic rental return, I employ rolling window time-dummy hedonic methods, which ensure that fluctuations in the return series are not driven by changes in the sample of transactions sold over time. Based on the transactions assigned to each portfolios, I first build rolling-window time-dummy hedonic housing price indices for each portfolio.[15] Based on these hedonic indices, I build quarterly capital gains series. Using the individual rental yields data constructed based on the *Mietspiegel* data, I then build rental yield rolling-window time-dummy time-dummy hedonic indices for each portfolio, which I benchmark to the mean portfolio rental yield in last period to have a time-series of rental yields for each portfolio. For more details on the construction of the return series please refer to Appendix E. All returns are then transformed into log points, to be more robust to outliers (Bali et al., 2016).

[15] For an overview of hedonic pricing methods in the context of housing markets, please refer to Hill (2013).

To assess the impact of predicted price dispersion on housing returns, I conduct the following fixed-effects panel regression:

$$y_{p,tq} = \beta_0 + \gamma \hat{u}_{p,tq} + B_X X_{p,tq} + \eta_{tq} + \epsilon_{p,tq}, \tag{12}$$

where the dependent variable $y_{p,tq}$ is one of the outcomes of interest (total returns, excess returns, capital gains, rental yields) for portfolio p in year-quarter tq. $\hat{u}_{p,tq}$ is the average predicted dispersion in portfolio p and $X_{p,tq}$ is a vector of the average characteristics of the transactions that compose portfolio p and η_{tq} are time fixed-effects.

4 Empirical results

In this section, I test the model predictions outlined in Section 2. Using the transaction level data, I exploit within neighborhood-year variation to assess the relationship between predicted price dispersion and transaction prices, as well as rents. I show a significant negative effect on transaction prices, which is notably less pronounced in the case of rents. Subsequently, I proceed to evaluate the impact of predicted price dispersion on rental yields, capital gains, and total returns. I identify a clear positive relationship with rental yields, no discernible pattern concerning capital gains, and as a result, a strong positive correlation with total returns.

Shifting the focus to across-portfolio variation, I demonstrate that portfolios with higher levels of predicted dispersion significantly outperform others in terms of housing returns, reinforcing the findings at the property transaction level. Moreover, the analysis reveals that the premium associated with investing in higher predicted dispersion portfolios varies over time, increasing during market downturns.

4.1 Transaction level data

Predicted dispersion, transaction prices and rents To better understand the effects of price uncertainty on net rental values and transaction prices, I conduct 2SLS regressions as in (7) for each city separately, where the outcome variables are transaction prices and rents net of utilities and maintenance costs.[16] To ensure comparability between the results of prices and net rents, I initially standardize the variables to have a mean of zero and a standard deviation of one. Figure 1 illustrates a bin scatter plot based on the regression results, showing both transaction prices and rents by city. A discernible pattern emerges across all cities in the sample. While transaction prices are significantly lower for higher levels of predicted price dispersion, rents largely remain constant across

[16] I employ the net rents value to ensure that our results are not influenced by variations in utility costs.

the distribution of price dispersion.

The tables in Appendix F present the 2SLS regression output underlying the binned scatter plots. They confirm that the coefficient of predicted price dispersion on transaction prices is two to three times larger than the one on net rents. Additionally, the coefficient on net rents is for some cities statistically insignificant, indicating that apartments with higher predicted price dispersion are not rented out at lower levels of rent. The effect of predicted price dispersion on transaction prices is not only highly statistically significant, with all cities displaying coefficients significant at the 1% level, but it is also economically relevant. Controlling for property size, age, neighborhood, and the year quarter of sale, the data suggests that moving from the first to the fifth quintile of predicted price dispersion distribution results in a decrease in sales prices ranging from 7,000€ (in Cologne) to 10,000€ (in Berlin). In other words, apartments sold in the same neighborhood and in the same year-quarter with similar characteristics, on average, will display differences in prices that amount between 4% and 7% of the average sales price in their respective cities.[17]

As a robustness test, I replicate the same analysis as described above but limit the dataset to include only those observations for which both price and rent information is available. This targeted focus on a subset of observations serves to alleviate potential biases that may arise from the data matching process discussed in Section 3 of this paper. The results are presented in Appendix F.2. The regressions conducted on this subset of observations corroborate the findings obtained from the main sample.

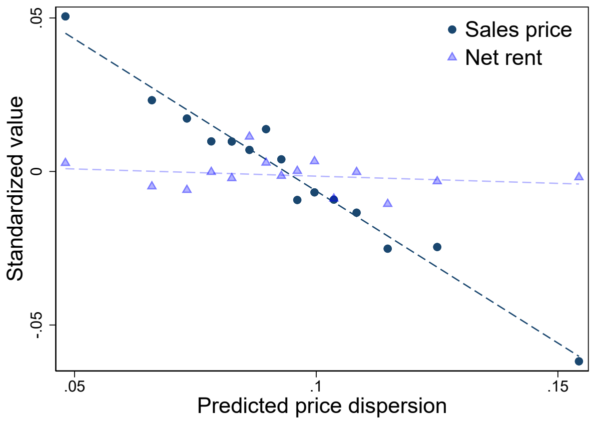
Furthermore, I show that price uncertainty also affects the second moment of the distribution of transaction prices. Properties characterized by higher price uncertainty exhibit prices with a greater standard deviation. The detailed results can be found in the appendix G. This highlights that the observed discount in transaction prices, which I am quantifying, is due to uncertainty rather than solely from lower demand for this category of properties. While lower demand may account for the lower transaction prices of these properties, it does not explain the increased volatility in transaction prices.

Predicted dispersion and returns to housing As illustrated in Figure 1, the effect of price dispersion on transaction prices is notably stronger than its effect on rental values. To understand whether this translates into a positive effect of price uncertainty on rental yields, I perform 2SLS regressions as in Equation (7), using rental yields at the point of sale as the outcome variable. Based on the regression results, I create binned scatterplots for each city in my sample, which are displayed in Figure 2. The regression output for each city can be found in the tables in Appendix H. In all cities, there exists

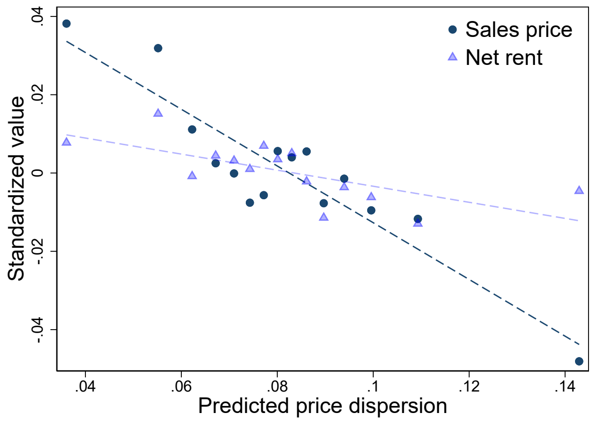
[17] These results are obtained by running regression (7) for each city separately and including a categorical variable for the quintiles of the idiosyncratic price uncertainty distribution.

Figure 1: *Predicted price dispersion, sales prices and net rents*

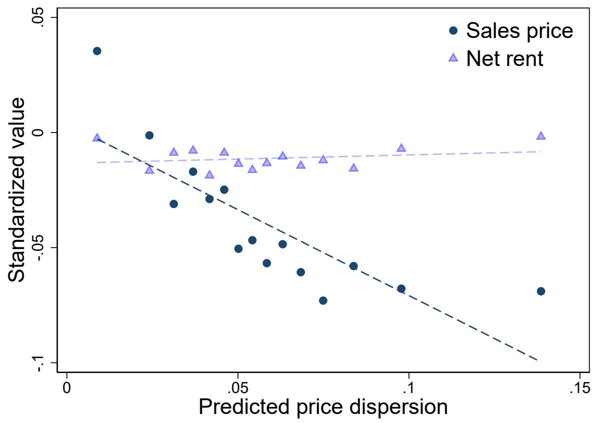
(a) *Berlin*



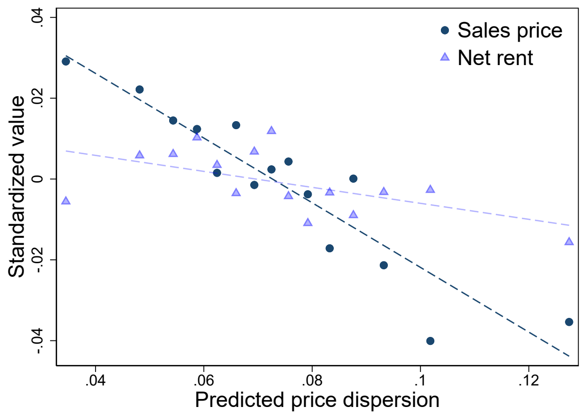
(b) *Hamburg*



(c) *Cologne*



(d) *Duesseldorf*



Note: All panels display bin scatters with 15 bins. Each bin represents the average value of sales prices and net rents residualized based on regression (7)*. Both sales prices and rents are standardized to haven mean zero and standard deviation of one.*

a clear positive and significant relationship between predicted dispersion and rental yields. The effects are not only statistically highly significant but also economically significant. When comparing sales of apartments within the same neighborhood in the same year quarter and controlling for size and building age, moving from the lowest to the highest quintile of predicted dispersion predicts, on average, an increase of between 20 (Dusseldorf) to 34 (Hamburg) basis points in the rental yield. This constitutes a substantial effect, as it represents between 4% (Dusseldorf) to 8.5% (Hamburg) of the average rental yield in the respective cities over the time period covered in the sample. This result also holds for the subsample, in which both prices and rents are observed at the point of sale as shown in Appendix H.1.

Next, I analyze the relationship between price uncertainty and capital gains, as well as total returns. The objective is to determine whether the predicted dispersion at the point of sale can serve as a predictor of future capital gains or total returns. Using the detailed information on the precise location of the apartments within the building, I match transactions of the same apartment over time, allowing me to construct apartment-level capital gains.[18] By utilizing the rental yield values at the point of the first sale, I subsequently calculate the average yearly total return for each pair of transactions involving the same apartment. Unlike rental yields, capital gains and total returns are observed only over the holding period and not for each transaction. Therefore, the unit of observation is now pairs of apartment sales and re-sales. Since the holding period is not observed at the time of the first sale, I cannot use the 2SLS framework as in Equation (7). Instead, I employ a two-step estimator in which I additionally control for the length of the holding period in the second stage. To be specific, I include a categorical variable that divides the holding period into 10 equally sized categories. First, I run the same first-stage regression at the apartment transaction level as outlined in Equation (7). Then, in the second stage, I regress the outcome variable for each pair of sales (j) on the predicted dispersion of the first sale, while controlling for property characteristics, time and neighborhood fixed effects, and the length of the holding period ($hp_{i,j}$) between the sale and re-sale.

$$\text{Stage 1:} \quad u^2_{i,tq} = g_c(X_i, tq) + B_X X_i + \eta_{tm} + \kappa_{n,ty} + e_{i,tq} \tag{13}$$

$$\text{Stage 2:} \quad y_{i,j} = \gamma \hat{u}^2_{i,tq} + B_X X_i + \eta_{tm} + \kappa_{n,ty} + \beta hp_{i,j} + \epsilon_{i,tq}, \tag{14}$$

where X_i is a vector of property characteristics that include size and age, η_{tq} are year-quarter fixed effects and $\kappa_{n,ty}$ are year-neighborhood fixed effects. The dependent variable y_{it} can refer to the capital gains and total returns.

I present the results in the form of binned scatterplots in Figure 2. Predicted price dispersion at the point of the first sale does not appear to be a reliable predictor of capital gains. Across all cities, no robust relationship is evident, and the coefficient on predicted dispersion is consistently statistically insignificant. This finding aligns with existing evidence regarding idiosyncratic risk in housing markets, which indicates that idiosyncratic price risk predominantly materializes at the points of sale and re-sale, thus not being attributable to changes in the house's fundamentals over time (Giacoletti, 2021; Sagi, 2021). In other words, we would not anticipate real estate with high predicted dispersion to appreciate at a different rate than real estate with lower price dispersion.

As for total returns, the pattern differs. In this case, I observe a robust and statistically significant positive effect of predicted price dispersion on total returns across all cities. Controlling for property characteristics, time, and neighborhood fixed effects, properties with higher levels of price uncertainty at the point of the first sale outperform the rest of

[18] Utilizing the information on the length of the holding period, I proceed to annualize the capital gains.

the market in terms of future total returns. It is important to note that by incorporating property characteristics, time, and location fixed effects, the results demonstrate that these excess returns cannot be attributed to varying exposures to the market portfolio. The regression output for each city can be found in the tables in Appendix H. Similarly, for total returns, the effects are not only statistically highly significant but also economically substantial. When comparing sales of apartments within the same neighborhood in the same year quarter and controlling for size and building age, moving from the lowest to the highest quintile of predicted dispersion predicts, on average, an increase ranging from 40 (Berlin) to 57 basis points (Cologne) in future total returns. This represents a significant impact, accounting for approximately 4% (Berlin) to 6% (Cologne) of the average total return in the respective cities over the time period covered in the sample. Please note that in the case of Hamburg, the number of observations is very limited due to the absence of key information necessary for the identification of repeat-sales.[19] Given this limited dataset, it is not surprising that the effects are not statistically significant. Nevertheless, they consistently exhibit the expected direction.

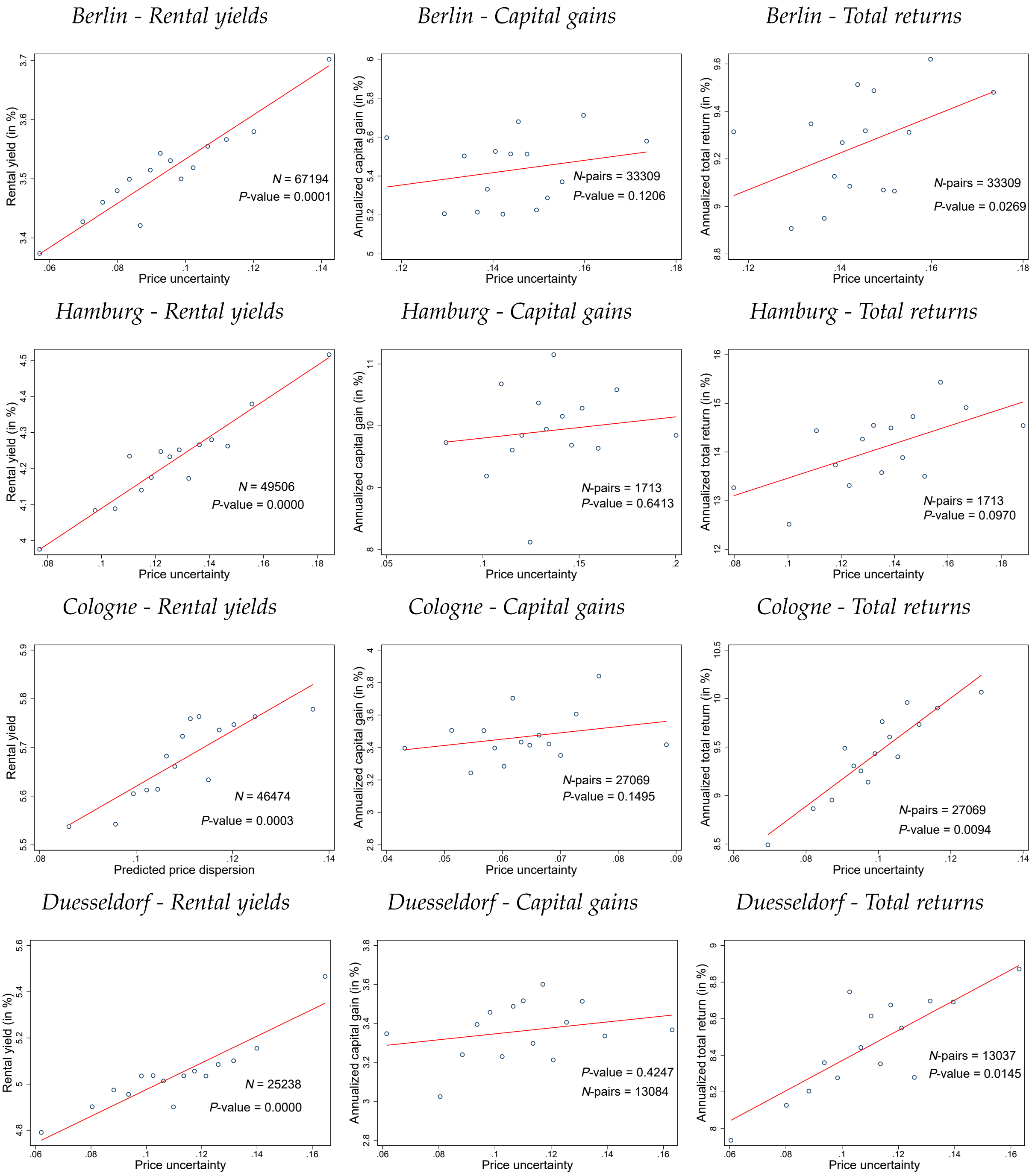
Predicted dispersion and rental yields for multi-family housing The model presented in Section 2 characterizes the optimal bid for a housing investor. Typically, large real estate investors hold multi-family houses in their portfolios rather than individual apartments. Furthermore, assuming risk-averse investors who consider resale risk at the time of purchase is even more appropriate for characterizing investment decisions in the multi-family housing market. This is because, in this market, average transaction prices are very high, representing a significant portion of investors' total portfolios. The dataset, constructed by Amaral, Dohmen, Schularick, et al. (2023), also includes information on transactions involving multi-family housing. In this section, I replicate the analysis conducted in the previous sections, but this time utilizing data on multi-family house transactions. To account for specific characteristics of the multi-family housing market, I incorporate additional control variables when measuring price dispersion, such as the building's lot size or the percentage of commercial use of the property.[20] I then employ the same 2SLS approach as in (7) to investigate the impact of predicted dispersion on rental yields. For multi-family housing, data on rental income after accounting for maintenance and utilities costs at the point of sale is available for a significant portion of the transactions for the cities of Berlin and Hamburg.[21] Therefore, in this analysis, I only consider observations for which I simultaneously have data on transaction prices

[19]For more details, please refer to Appendix C.3.

[20]Please note that I exclude all buildings in which commercial properties occupy more than 20% of the usable area of the building.

[21]Unfortunately, in the case of Cologne most of the transactions of multi-family housing have missing information about the size of the houses. For Duesseldorf, only information on the gross rental yields (before excluding maintenance costs) is available, as such, not allowing for a clear comparison across transactions.

Figure 2: *Biscatter of housing returns on predicted price dispersion by city*



Note: The first column displays a binscatter of rental yields on predicted price dispersion based on the 2SLS regression output of (7). Here the unit of observation are transactions. The second and third columns displays a binscatter of capital gains and total returns on the predicted price dispersion at the point of the first sale based on the two-step regression estimator (13). Here the units of observation are pairs of sale and re-sale of the same apartment.

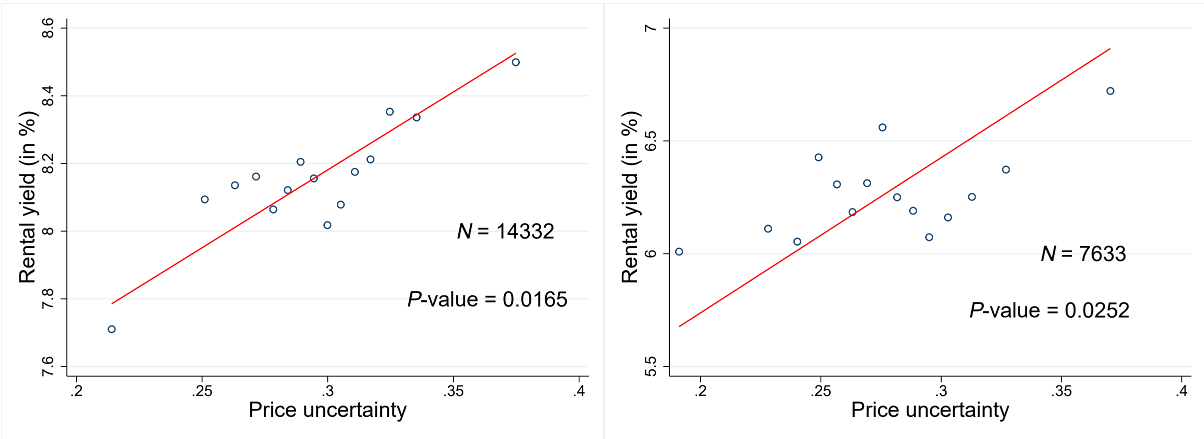
and rental incomes.

The results are presented in Figure 3. Similar to the findings for apartments, a positive and robust relationship between predicted dispersion and rental yields at the transaction-house level is evident. After accounting for property characteristics, time and location fixed effects, it becomes clear that multi-family houses with higher predicted averages tend to yield higher rental returns on average. Once again, the results exhibit not only statistical significance but also considerable economic relevance. Transitioning from the lowest quintile to the top quintile of predicted price dispersion in a given year and neighborhood, while controlling for property characteristics, predicts a rental yield increase of 65 basis points in Berlin and 40 basis points in Hamburg. These increases represent 7% and 6% of the average rental yields observed over the sample period in Berlin and Hamburg, respectively. For detailed regression output tables, please refer to Appendix H.2.

To further test the robustness of these findings, I conducted the same analysis exclusively on multi-family houses without any commercial properties. Commercial properties are often more challenging to value. Importantly, the results remain consistent, indicating that even for 100% residential multi-family housing, rental yields increase with predicted dispersion.

Figure 3: *Rental yields & predicted dispersion for multi-family housing by city*

(a) *Berlin* **(b)** *Hamburg*



Note: The binned scatters are based on regression (7) with ratio of net rental income to transaction price of multi-family houses at the point of sale as the outcome variable. Panel a) displays the results for Berlin for the period between 1970 and 2022 and Panel b) displays the results for Hamburg for the period between 1991 and 2022. The regression output tables can be found in Appendix H.2.

4.2 Portfolio sorting analysis

Properties are traded very infrequently, which means that a time series for the value of a specific property is not observed, and the variation that I can analyze at the transaction level is cross-sectional. Using the portfolio price and return time-series constructed in Section 4.2, I can, however, also analyze the time-series variation. Since these portfolios were built based on hedonic methods that control for property characteristics, the differences in performance across the portfolios arise solely due to their differences in value uncertainty. From this perspective, the hedonic portfolio sorting analysis can be interpreted as a multi-sort portfolio analysis, where researchers aim to control for specific asset characteristics to isolate the effects of the risk factor (Bali et al., 2016). In this section, I first demonstrate that portfolios containing properties with higher levels of uncertainty outperform the rest of the market. Decomposing the total returns, I then illustrate that the return differences across portfolios stem from variations in rental returns and not from capital gains. This corroborates the results from the transaction-level analysis. Finally, I also establish that exposure to the market portfolio remains constant across the portfolios, and higher value uncertainty portfolios exhibit higher alphas.

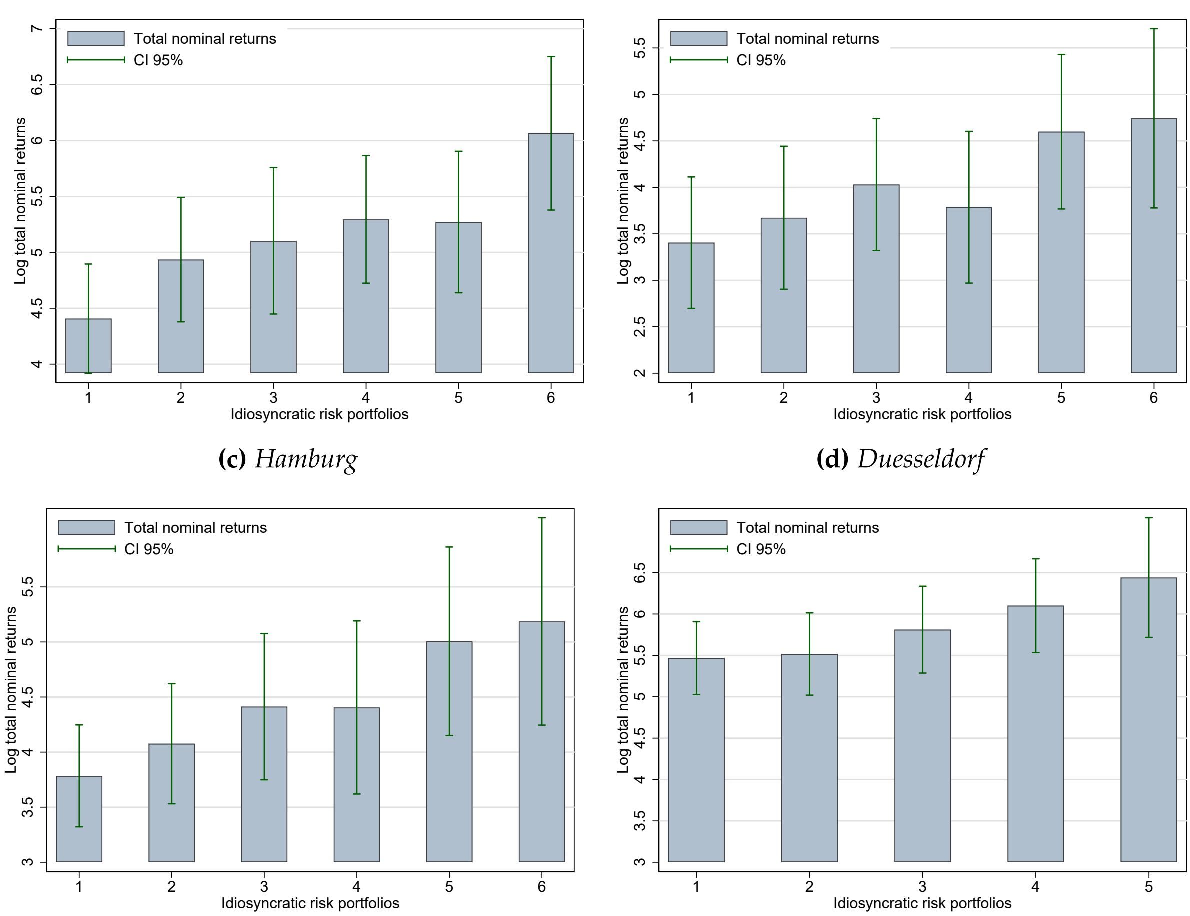
In Figure 4, I plot the total nominal returns by portfolio for each city over the entire sample period, accompanied by the 95% confidence intervals. The portfolios are sorted based on the level of predicted dispersion associated with the transaction of the properties within them. Portfolio 1 consists of transactions of properties with the lowest levels of predicted dispersion, whereas portfolio 6 comprises transactions of properties with the highest levels of predicted dispersion. All three cities exhibit a consistent pattern: total returns consistently rise with increasing value uncertainty, demonstrating a nearly monotonic relationship.

To assess the magnitude of the return differences depicted in Figure 4, I conduct hypothesis tests to determine if these differences are statistically significant. Following the best practices in the asset-pricing literature, I test the differences in log excess returns. I build the excess returns by subtracting the returns on short-term German government bonds. Additionally, I perform a decomposition of the return differences into separate components, namely capital gains and rental return differences. The results can be found in Table 3, where I provide the average log excess return difference between portfolio 6 and portfolio 1 as well as the difference between portfolio 6 and the average of the rest of the portfolios for the all cities separately.

For all cities, investing in portfolio 6 provides a statistically significantly higher return than investing in portfolio 1. For example, for Cologne investing in the portfolio with the highest value uncertainty provides a premium of 149 basis points per quarter over investing the portfolio with the lowest value uncertainty. Not only is this premium statistically significant, but it also economically very large. The most documented investment risk strategies in the stock market literature, such as the small-minus-big, high-minus-low or momentum strategies yield a return premium in the range of 100 to 150 basis points per quarter (e.g. Ehsani and Linnainmaa, 2022; Fama and French, 2015). The differences in returns between portfolio 6 and an average of the rest of the portfolios is also statistically significant, although not as large as the difference to portfolio 1.

The second and the third column also show the differences in capital gains and rental returns. It becomes evident that most of the return differences come from the differences in rental returns and not from the differences in capital gains. This result is consistent

Figure 4: *Log total nominal returns & predicted dispersion by city*



(a) *Cologne* **(b)** *Berlin* **(c)** *Hamburg* **(d)** *Duesseldorf*

Note: The Figure shows the total nominal returns on six equally-sized portfolios built based on predicted disperion quantiles. The returns to the portfolios are constructed using hedonic regressions controlling for property characteristics. For more details please refer to section 3.2

with the predictions of the model and the transaction-level results, that also indicated that rental returns and not capital gains increase with value uncertainty.

Exposure to systematic risk Although the differences in total returns are quite large, this does not necessarily mean that varying levels of value uncertainty are driving the return differences. It could be the case that the different portfolios have different exposures to systematic risk in the market. In order to test this hypothesis, I run the following regression:

$$ln(r_{pt}) = \alpha_p + \beta_p ln(r_{mt}) + \epsilon_{pt}, \quad (15)$$

where $ln(r_{pt})$ is the log total excess return for portfolio p, which is constructed by subtracting the risk-free return to the nominal total return of each portfolio, and $ln(r_{mt})$ is the total excess return for the city of Cologne. In Figure 5 I plot the both the α coefficient for each portfolio p, as well as the β coefficient. While the show the exact

Table 3: *Portfolio return differences in log points by city*

Portfolio	Excess Returns	Capital Gains	Rent Returns	N
Cologne				
P6 vs P1	1.55*** (0.29)	-0.00 (0.27)	1.39*** (0.07)	264
P6 vs rest	1.28*** (0.27)	-0.03 (0.25)	1.15** (0.56)	792
Berlin				
P6 vs P1	1.06** (0.44)	0.45 (0.44)	0.61*** (0.14)	214
P6 vs rest	0.84** (0.35)	0.36 (0.34)	0.47 (0.43)	642
Hamburg				
P6 vs P1	1.47*** (0.47)	0.32 (0.40)	1.14*** (0.20)	166
P6 vs rest	0.84*** (0.32)	0.33*** (0.03)	0.51* (0.29)	498
Duesseldorf				
P5 vs P1	0.70*** (0.26)	0.04 (0.22)	0.64*** (0.15)	306
P5 vs rest	0.45** (0.22)	0.08 (0.17)	0.37 (0.79)	765

Note: Differences are measured as coefficients in a random effects panel regression of the dependent variable (log capital gain, log rental yield and log total housing return respectively) on a P6 dummy and year fixed effects. Driscoll-Kraay standard errors (in parenthesis). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

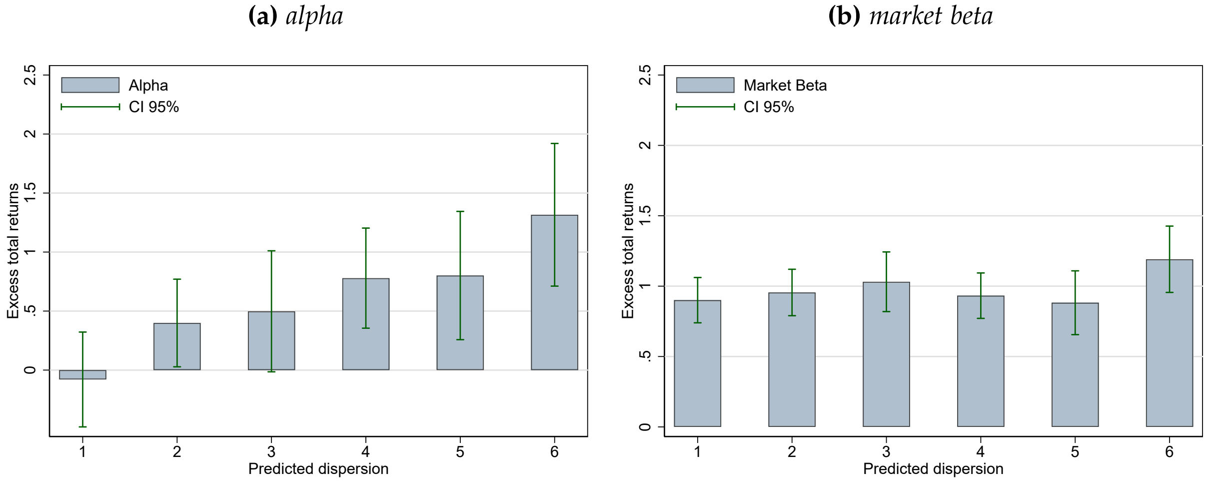
same pattern as the total nominal returns, the do not show significant differences across the portfolios. This indicates that the observed differences in total nominal returns in **??** are not being driven by differential exposure to the market portfolio returns. To the extent that the market portfolio returns represent systematic, non-diversifiable, risk, this indicates that the main results are not driven by different exposures to systematic risk.

4.3 Robustness analysis

In this section, I present the robustness analysis conducted to ensure that alternative factors are not influencing the primary results of the paper. Here, I specifically emphasize demonstrating the robustness of the findings that confirm the three main predictions of the model about properties with higher value uncertainty: i) they trade, on average, for lower prices, ii) they yield, on average, higher rental yields and iii) they do not realize, on average, higher capital gains but have higher total returns.

Observed rental income and price at transaction level The main results, both at the transaction-level and in the portfolio-sorting analysis, are based on a matched sample of transactions where I align transaction prices with net rental income from the *Mietspiegel*,

Figure 5: *Total returns controlling for systematic risk exposure, Cologne 1989-2022*



Note: The six equally-sized portfolios are built on the predicted dispersion. Standard errors are adjusted for time-series autocorrelation using Newey-West with 6 quarter lags.

as explained in the data section. This approach introduces two potential sources of bias. Firstly, it's not certain whether all the properties I match are actually being rented, raising questions about the accuracy of predicting the rental income these properties would generate if they were indeed in the rental market. Secondly, the matching process relies on the same set of characteristics used to predict price dispersion. If I do not find an effect when regressing rental values on predicted price dispersion, it could potentially invalidate the rest of my results.

To address these concerns, I replicate all my analyses using only the sample of properties for which I simultaneously observe rental income and transaction prices. I conduct this analysis for both apartments and multi-family housing. This approach addresses both concerns, as the presence of information on rental income guarantees that the property is indeed being rented out and provides the exact level of rental income, eliminating the need for estimation. As I demonstrate in Appendices F.2 and H.1, all results hold when using the samples for which both rental income and prices are observed, with the results being particularly robust in the case of multi-family housing.

All sales As mentioned in Section 3, the inclusion of apartment fixed effects in my baseline regression could pose two problems. Firstly, considering that properties are rarely transacted during the sample period, the limited number of observations per fixed effect has the potential to introduce bias to the coefficients and consequently impact the estimated residuals in my baseline regression (1). Secondly, properties that transact more than once might not be representative of the universe of property transactions and have special characteristics that might bias the results (Haan and Diewert, 2011). In order to assess whether these issues might be influencing my results, I conducted a new analysis where I excluded the apartment fixed effects from the baseline regression and using the exact same sample, I run the following regression:

$$ln(p_{i,tq}) = \eta_{tm} + \kappa_{n,tq} + f_c(x_i, ty) + u_{i,tq}, \tag{16}$$

which is equal to the regression (1), but excludes the property fixed effects. Compared to my baseline results, the regression without apartment fixed-effects yields significantly greater dispersion in the residuals. However, the residuals from the specification without apartment fixed-effects exhibit a strong positive correlation with the residuals from the specification with apartment fixed-effects. Therefore, it is unsurprising that I am able to replicate the main results. I present the detailed results in Appendix I.1. Summary statistics for all sales by city can be found in Table 7 in the Appendix.

Housing renovations and price dispersion The model specification in equation (1) does not take into account the potential impact of renovations on the value of apartments. Without explicitly controlling for renovations, it is possible that the residuals are picking up this effect. Therefore, the model may not be capturing a measure of idiosyncratic price deviation, but rather a measure of the enhancing value of renovations. Additionally, apartments continuously depreciate, which counteracts the effects of renovation. To determine whether these two effects are a significant source of measurement error in my analysis, I adapt (1) by including building-time fixed effects. Since the largest renovation works are typically done simultaneously for all apartments within a building, this approach should already control for the most significant renovation works. To estimate the idiosyncratic price deviations, I use the following regression:

$$ln(P_{i,tq}) = b_{i,ty} + \eta_{tm} + \kappa_{n,tq} + fc(x_i, ty) + u_{i,tq}, \tag{17}$$

where $b_{i,ty}$ is a building-year fixed effect that captures building specific characteristics that also change over time. Given the large number of buildings for which there are several apartment transactions every year, I am able to estimate the coefficients precisely. As I show in Appendix I.3 controlling for building renovations does not change my results.

Adjusting for heterogeneity in holding periods Equation (1) does not explicitly take into account the relation between the variance of the residuals and holding period. Giacoletti (2021) shows that the variance of the residuals increases slightly with holding period. Additionally, the properties, which are sold more often will have smaller residuals by construction, since the apartment fixed-effects in equation (1) will be better estimated. To take these issues into account, I add a second step to the estimation of the transaction level price dispersion, in which I explicitly regress the squared residuals from equation (1) on a smooth function of the holding period, hp_i, interacted with the number of sales, $sales_i$:

$$u_{i,tq}^2 = g_c(sales_i, hp_i) + e_{i,tq}^2. \quad (18)$$

Then I take $e_{i,t}^2$ as my new measure of the price dispersion. I then follow the same steps as described in Section 3 of the paper and analyse the relation between prices and returns and the new measure of predicted price dispersion that explicitly takes into account the relation between the squared residuals from regression (1) and the length of holding period as well as the number of times the property was transacted. As shown in Appendix I.4, all the main results hold.

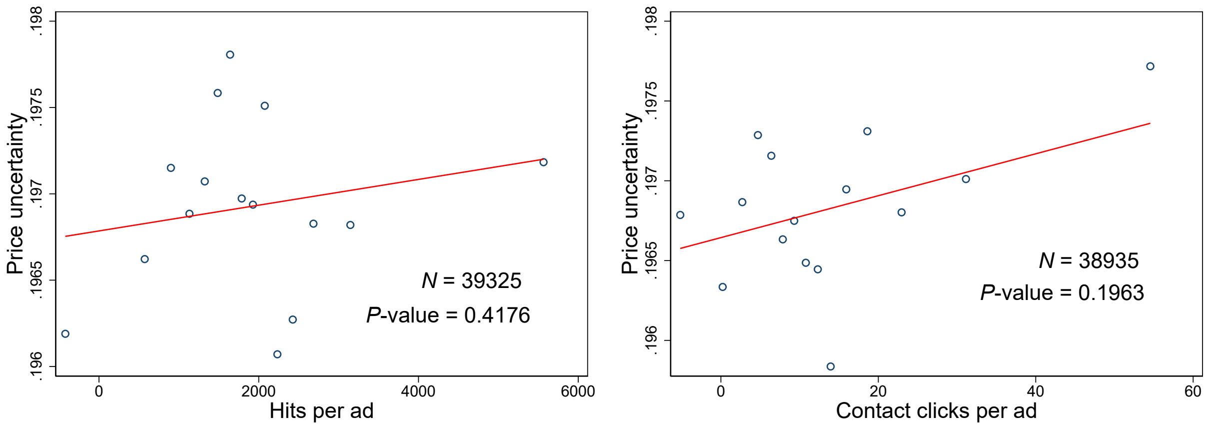
Instrumenting price uncertainty In my baseline analysis, I regress realized variance of the pricing errors on property characteristics to generate a prediction of price uncertainty. This process inherently creates a correlation between the measure of ex-ante price uncertainty and the property characteristics. This correlation could potentially raise concerns about the main results of the paper, as I might be capturing a mechanical effect related to a preference for specific types of characteristics in transaction prices.

To address this issue, I employ direct measures of sellers' market thickness to predict price uncertainty. The first measure is based on the Euclidean distance between property characteristics and the mean characteristic in the market. The second measure relies on the relative frequency of specific characteristics in the market. Detailed information on the construction of these measures can be found in Section 4.2. As demonstrated in Appendix I.2, using these measures of market thickness to predict price dispersion does not alter the main results and, in some cases, even reinforces them. These results are also confirmed in the following section, where I show a strong correlation between measures of market size and liquidity and value uncertainty at the transaction property level.

Bargaining power and transaction prices One of the predictions of my bargaining model in section 5 is that a higher bargaining power of the buyer could explain my empirical results. Intuitively, if properties with higher price uncertainty have thinner buyer markets, then we would expect buyers to have higher bargaining power, consequently driving down the prices of these properties. To test this hypothesis, I approximate the number of potential buyers per house by using click data from online advertisements. Specifically, I construct two measures of buyer market thickness using information on the number of clicks per ad and the number of times the seller is contacted per ad. While clicking on an ad does not necessarily indicate an intent to buy and is thus a noisy proxy for potential demand, contacting the seller of the property involves writing a text and clearly demonstrates an interest in acquiring the property. I then regress price uncertainty on these two proxies of buyer market thickness at the property-transaction level. Additionally, I include property characteristics, location, and time fixed effects to ensure a comparison of similar properties. The results for the city of Cologne are displayed in Figure 6. There is no significant relationship between price uncertainty and buyer market thickness for both proxies. Indeed, the data does not indicate that properties with higher price uncertainty are transacted in thinner buyer markets.

Figure 6: *Price uncertainty and buyers' bargaining power*

(a) *Hits per ad* **(b)** *Contact clicks per ad*



Panel (a): Binscatter showing the relation between price uncertainty and measures of buyers' bargaining power. The output is based on regressions controlling for time and location fixed effects, property characteristics and time on the market. The data is for the city of Cologne and covers the period between 2008 and 2018. The source is Immoscout.

Additionally, bargaining power is not only affected by number of potential buyers, but also by the outside options of the buyers. As I show in Section 6, properties with higher price uncertainty have a lower number of comparable properties on the market. This means that buyers wanting to buy these properties will typically face a lower number of outside options in the market. All else constant, this increases the bargaining power of sellers, which can then raise prices even more. And, as such, this would go against ma empirical findings of a pricing discount for properties with higher price uncertainty.

Overall, I do not find evidence that higher buyers' bargaining power is driving my empirical results. Theoretically, it is also not clear that buyers will have a higher bargaining power for properties with higher price uncertainty, as the number of outside options is typically smaller for this type of properties.

5 Theoretical Framework

Houses are highly heterogeneous goods, and their characteristics are valued differently by potential buyers. For instance, larger houses tend to be more appealing to larger households. Consequently, houses located in close proximity to one another may be traded in markets characterized by distinct types and quantities of potential buyers, exposing them to markets with varying sizes and thicknesses. Intuitively the size and the thickness of the markets will affect the quality of the matching process between sellers and buyers. Thinner markets typically result in less efficient matching between sellers and buyers, thereby generating greater uncertainty surrounding transaction prices.[22]

Additionally, the attributes that render a house attractive for purchase also influence its demand in the rental market. This, in turn, impacts the uncertainty surrounding the rental value at which the property can be rented out. In this section, I develop a theoretical framework to characterise the optimal bid for a risk-averse investor who faces uncertainty regarding both the rental and resale values of the property.

Setup This is a model with three periods, with one seller and one financially unconstrained investor, who wants to buy a house to rent it out. In the first period, the seller puts the house for sale and enters a Nash bargaining process with the investor. The bargaining power of the investor is given by α. The α parameter can be understood as reflecting the buyers' market thickness for the specific house h.[23] After having bought the house in the first period, the investor rents it out in the second period. The rent is exogenous and random. In the third and final period, the investor sells the house for a random price. I assume that, in the first period, the investor knows the reservation value of the seller, $\underline{PV_S}(h)$, and, as such, will not bid below it. The bargaining problem of the first period can then be written as:

$$\max_{V_B, V_S} \quad V_B^{\alpha} V_S^{1-\alpha} \tag{19}$$

$$s.t. \quad V = V_B + V_S \tag{20}$$

$$V = PV_B(h) - \underline{PV_S}(h) \tag{21}$$

$$\alpha \in (0, 1) \tag{22}$$

where the first constraint is the standard constraint from a Nash bargaining problem that splits the bargaining surplus among the seller and the buyer. The second constraint tells us that the value that will be split between buyer and seller equals the difference between the private valuation of house h by the buyer, PV_B, and the reservation value of the seller, $\underline{PV_S}$. In other words, the final transaction price will be between the private valuation of the buyer and the reservation value of the seller and will be determined by the relative bargaining power of each party. For simplification, I assume that housing is the only asset in the economy and, as such, all income generated by housing will be consumed. The private value of the investor of house h is the discounted value of

[22] This is because, all else being equal, the probability of a match occurring between a seller and a buyer who value the house equally is lower in thinner markets (e.g Han and Strange, 2015).

[23] I take the relation between bargaining power and number of buyers as exogenous, however it can be micro-founded in a setting of sequential bargaining as demonstrated in Rubinstein and Wolinsky (1985).

renting out the house in period 2 for R_2 and selling it in period 3 for the expected price $P_3(h)$:

$$PV_B = \beta E_1[u_2(R_2(h)] + \beta^2 E_1[u_3(P_3(h))], \tag{23}$$

where P_3 and R_2 are log-normally distributed with means μ_P and μ_R and variances σ_P^2 and σ_R^2, respectively. $\beta \in (0,1)$ is the discount factor. The mean μ_P can be thought of as the expected market value of the property, while the variance σ_P^2 measures the house-specific price deviation from its expected market value. The same logic holds for the rent. This model describes transactions for one specific house on the market. When the investor rents out the house and resells it, they face a larger market and receive a random rent and price.[24] Furthermore, I assume that the risk-averse buyer has CRRA utility.[25]

Solution To solve the maximization problem, I substitute the Nash bargaining constraint into the problem:

$$\max_{V_S} \quad (V - V_S)^{\alpha} V_S^{1-\alpha} \tag{24}$$

$$s.t. \quad V = PV_B(h) - \underline{PV_S}(h) \tag{25}$$

$$\alpha \in (0,1). \tag{26}$$

Deriving the first-order condition and solving for V, we get:

$$V = \frac{1}{(1-\alpha)} V_S. \tag{27}$$

Plugging in the constraint and using the definition of the private value of the investor yields:

$$\frac{1}{(1-\alpha)} V_S = \beta E_1[u_2(R_2(h))] + \beta^2 E_1[u_3(P_3(h))] - \underline{PV_S}(h). \tag{28}$$

Since the rent in period 2 and the price in period 3 are log-normally distributed, we have that:

$$E_1[ln(P_3)] = ln(E_1[P_3]) - \frac{1}{2} Var_1[ln(P_3)] \tag{29}$$

$$E_1[ln(R_2)] = ln(E_1[R_2]) - \frac{1}{2} Var_1[ln(R_2)] \tag{30}$$

Given that the buyer knows the reservation value of the seller, the optimal bid of the buyer in the first period will equal the bargaining surplus of the seller, $B^* = V_S$. As

[24] The randomness of both sale and rent prices can be justified by extensive empirical evidence demonstrating the substantial unpredictability of prices in housing markets (Kotova and Zhang, 2021; Giacoletti, 2021).

[25] For simplification, I assume log utility.

such, assuming the buyer has log utility, we have the following expression for optimal bid by the buyer in period 1:

$$B^* = V_S = (1-\alpha)\left[\beta(ln(\mu_R) - \frac{1}{2}\sigma_R^2) + \beta^2(ln(\mu_P) - \frac{1}{2}\sigma_P^2) - \underline{PV_S}(h)\right] \tag{31}$$

Since the optimal bid of the buyer will be at or above the reservation value of the seller, the bid will be accepted. I do not explicitly model the sellers' problem, however, ss demonstrated in DeGroot (2005) and under the assumption that the seller knows the distribution of buyers, it is optimal for sellers to accept the first bid above their reservation value.

5.1 Comparative statics

In this subsection, I explore the comparative statics of the models' equilibrium predictions, primarily based on equation (31). In doing so, I investigate the effects of price and rental dispersion, as measured by the idiosyncratic variances, on transaction prices and returns to housing. These predictions are then tested empirically in the following sections of the paper.

P1. Higher idiosyncratic price variance leads to lower transaction prices For properties with a higher expected idiosyncratic price variance, the optimal bid, and consequently, the transaction price will be lower. All else being equal, a risk-averse buyer will choose to bid a lower amount for a property that has a more uncertain resale value.[26] Therefore, the model predicts that these properties should transact at a lower price.

$$\frac{\partial B^*}{\partial \sigma_p^2} = -(1-\alpha)\frac{1}{2}\beta^2 < 0 \tag{32}$$

P2. Higher idiosyncratic price variance leads to higher rental yields Using equation (31) I write the ratio of the expected rental income in period 2 to the transaction price in period 1 as a function of the price idiosyncratic variance. Then taking the derivative with respect to the idiosyncratic price variance I get the following equation:

$$\frac{\partial \frac{E_1(R_2)}{B^*}}{\partial \sigma_p^2} = \frac{(1-\alpha)\frac{1}{2}\beta^2 * E_1(R_2)}{(1-\alpha)^2\left[\beta(ln(\mu_R) - \frac{1}{2}\sigma_R^2) + \beta^2(ln(\mu_P) - \frac{1}{2}\sigma_P^2) - \underline{PV_S}(h)\right]^2} > 0 \tag{33}$$

from which it becomes evident that the ratio of rents to prices, known as rental yield, increases with the idiosyncratic price variance. In other words, an investor will only be willing to offer a lower value for a given rental cash flow if they anticipate higher price uncertainty, which consequently mechanically increases the rental yield.

[26]Please note that even if the buyer does sell the house in the future, the idiosyncratic component might still impact their optimal bid through the use of the house as a collateral.

Nevertheless, if price uncertainty arises as a result of trading frictions in the housing market as shown in Sagi (2021), then we expect properties with higher price uncertainty to be traded in small and illiquid markets. This would mean that there are less potential renters for the property and, as such, the landlord would have less bargaining power. Other things constant, this would lead to a lower rental income, putting downward pressure on the rental yield. Which effect is stronger is not clear ex-ante and is, therefore, an empirical question.

P3. Higher idiosyncratic price variance does not result in excess capital gains but does lead to higher total returns Due to the randomness of the resale price in the third period, this model does not make a prediction about the relation between capital gains and the idiosyncratic price variance. Nevertheless, the theoretical and empirical evidence on idiosyncratic price risk in housing markets shows that this risk primarily materializes at the point of sale and resale (Sagi, 2021; Giacoletti, 2021). Consequently, it arises primarily due to trading frictions in the housing market. In other words, idiosyncratic price variance is not driven by changes in the fundamental characteristics of the house and, thus, should not, on average, impact capital gains.[27] In conclusion, if properties with higher idiosyncratic price variance are expected to yield higher rental yields but no additional capital gains, then these properties should offer higher total returns.

Note that if a buyer owns multiple houses, they might be able to diversify away idiosyncratic price deviations. If this is the case, then in a standard asset-pricing model with complete markets, the idiosyncratic variance should not influence the market's stochastic discount factor. In Appendix B, I demonstrate how the assumption of incomplete markets can lead to a market stochastic discount factor that also incorporates idiosyncratic price variance. This provides a theoretical framework for idiosyncratic risk being priced in housing markets.

In this section, I have provided a stylised theoretical framework that predicts the empirical results about transactions prices and returns I showed in previous sections. In the next section, I will empirically test the assumptions of this model.

6 Market size, liquidity and value uncertainty

Drawing inspiration from theoretical frameworks that establish a relationship between market size and price dispersion in OTC markets (Gavazza, 2011), in this section, I empirically examine the connection between price uncertainty, market size, and market liquidity at the transaction property level. Given that houses are extremely heterogeneous goods and, consequently, exposed to highly diverse markets, it is crucial to

[27] I provide a more detailed and extensive summary of this result in Appendix A.

investigate the relationship between price uncertainty and market size or liquidity at the transaction property level. Aggregation at higher levels might potentially obscure significant variations in the data.

6.1 Value uncertainty and market size

To measure market size at the property level, I rely on the literature on atypical properties (e.g. Bourassa et al., 2009; Haurin, 1988), and build an atypicality index for each transaction i based on the distance between the properties' characteristics and the average property in the neighborhood:

$$ATYP_i = \sum_n |exp(\hat{\beta}_n X_n) - exp(\hat{\beta}_n \overline{X}_n)|,$$

where $\hat{\beta}_n$ represent the shadow price of characteristic n estimated in a log-linear hedonic regression using all the transactions for the respective city. $\overline{X}_n$ is the average of characteristic n for all the transactions in the respective neighborhood. The $ATYP_i$ then measures the relative distance of the properties' characteristics to the mean, or typical, property in the neighborhood weighted by the shadow price of each characteristic. The higher the value of $ATYP_i$ the more atypical a property is with respect to the other properties in the neighborhood. As such, the atypicality index is a direct measure of the sellers' market size of the property. Through general equilibrium effects, we expect the demand for these type of properties to also be relatively low, altogether making the markets for atypical properties relatively small. In Figure 7, I plot a scattered bin plot of predicted dispersion on the atypicality index. For all cities in the sample there is a clear positive relation: properties with higher levels of idiosyncratic risk are also more atypical. In other words, properties transacted in smaller markets display larger levels of value uncertainty.

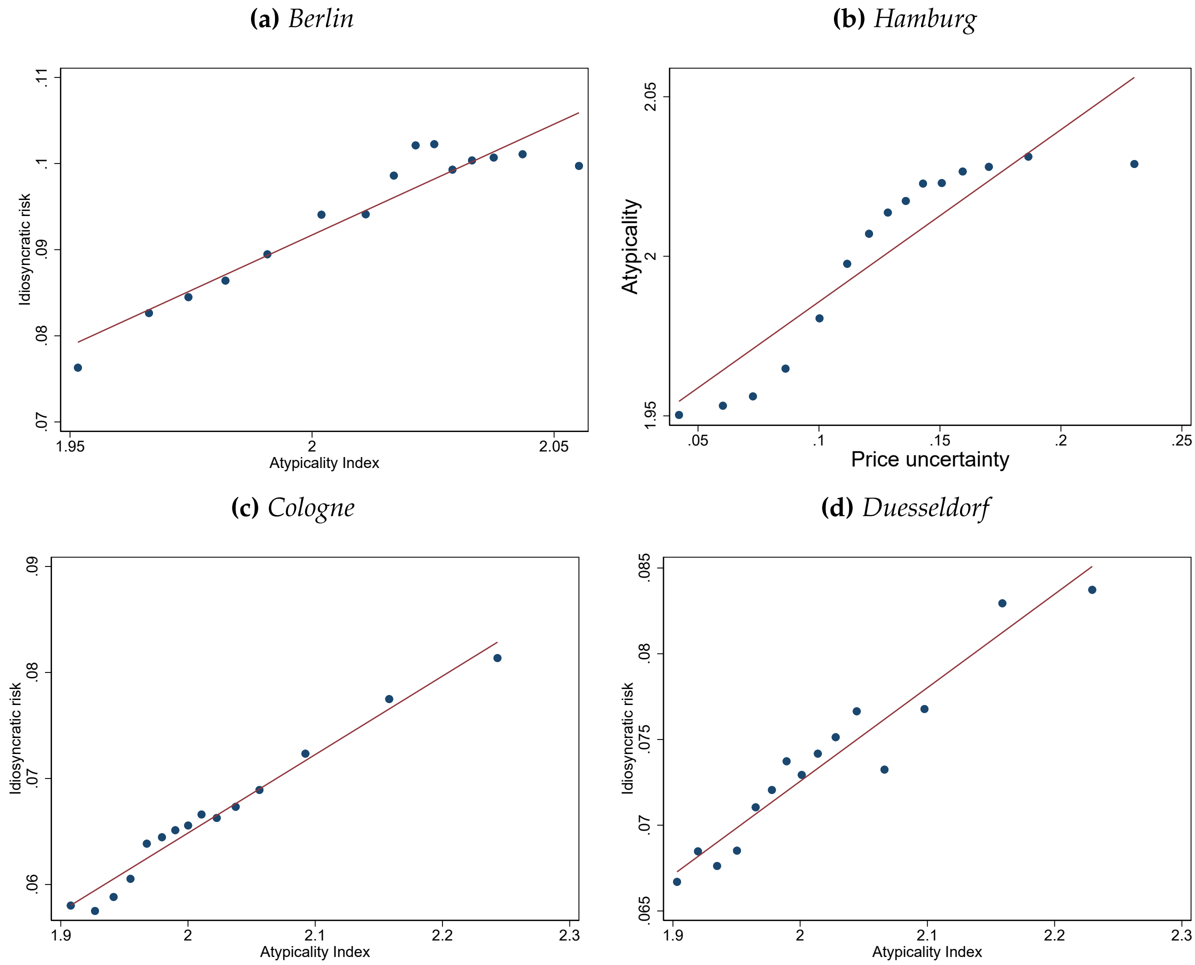
The statistical significance of this relation is confirmed by the regression results in the tables in Appendix **??**, which show that the relation between idiosyncratic risk and the atypicality index is positive and highly significant also when neighborhood and time fixed effects.

6.2 Value uncertainty and asset liquidity

Using the data set with matched transactions and advertisements I build two measures of asset liquidity at the transaction property level. Firstly, I created a measure of time on the market, defined as the number of weeks between the day the ad was posted and the day it was taken offline:

$$TOM_{it} = \frac{Number\ of\ days\ advertised}{7}$$

Figure 7: *Value uncertainty and atypicality of the property*



Note: The figure dispalys binscatters of predicted price dispersion on atypicality index at the transaction property level for all cities in the sample. In all binscatters the underlying regressions include year-quarter and neighborhood fixed-effects as well as controls for property characteristics.

Secondly, I constructed a measure of the spread between the asking price and the transaction price as:

$$Spread_{it} = 100 \cdot \frac{(Sales\ price_{it} - Asking\ price_{it})}{Asking\ price_{it}}$$

To gain insight into the relation between value uncertainty and the expected duration a property stays on the market, it is essential to consider whether the listed property ultimately sells or not. In the literature on housing markets, hazard models have been employed to analyze the expected time a property spends on the market (Haurin, 1988; Han and Strange, 2015). Following the literature, I assume the following hazard function for time on the market:

$$h(tom) = h_0(tom) * exp[\gamma\hat{\sigma} + B_X X + \eta_{tq} + \kappa_n], \tag{34}$$

where $h_0(tom)$ is the baseline hazard rate and its specific shape will depend on the assumption about the distribution of the error term. The hazard rate $h(tom)$ then denotes the probability of a property being sold at time t, conditional on the seller listing the property to that point in time, and subject to the predicted dispersion, $\hat{\sigma}$, the property characteristics, X and the year-quarter, η_{tq}, and neighborhood, κ_n fixed effects. I estimated the hazard rate using various error term distributions and presented the results in Table 4. The first row of the table displays the effect of predicted dispersion on the hazard rate of time on the market, given by its hazard ratio. Across all specifications, it becomes evident that increased value uncertainty, as quantified by predicted price dispersion, is associated with higher expected time on the market. A one unit increase in value uncertainty is associated with more than doubling the probability that the property does not get sold.

Table 4: *Expected time on the market and value uncertainty, Hamburg (2012-2022)*

	Exponential	Weibull	Cox
Price uncertainty	2.58^{**}	2.61^{***}	2.66^{***}
	(0.994)	(0.905)	(0.917)
Year-quarter FEs	Yes	Yes	Yes
Neighborhood FEs	Yes	Yes	Yes
Property characteristics	Yes	Yes	Yes
N	24497	24497	24497

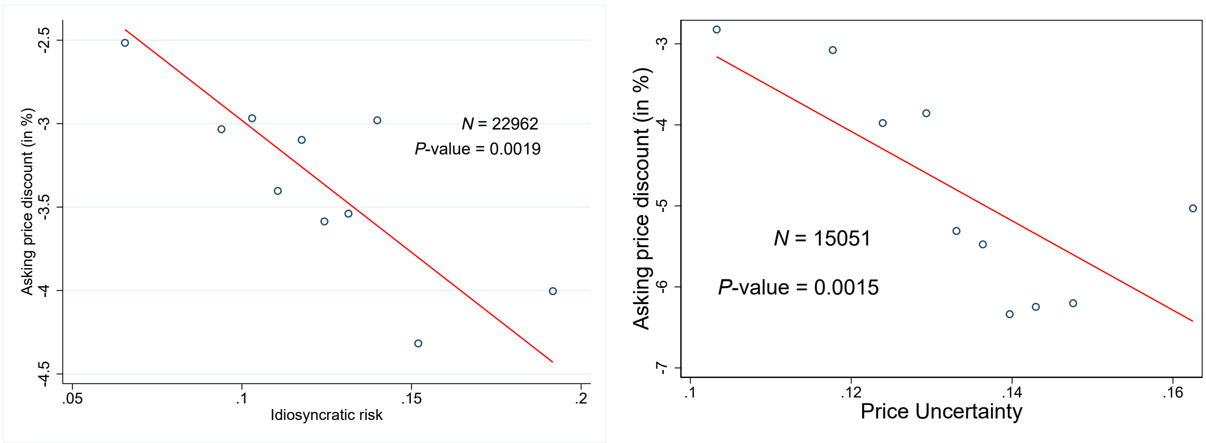
The Table reports the results of three different duration models of time on the market. The first row displays the estimated hazard ratio of predicted dispersion. Standard errors are shown in parenthesis. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

Having established that properties with greater predicted dispersion tend to, on average, spend a longer time on the market, the subsequent analysis narrows down to the subset of properties that have been successfully sold. In this context, I examine whether the transaction prices for these properties significantly differ from their initial asking prices. I regress the spread between the asking price and the transaction price on predicted price dispersion while controlling for property characteristics, neighborhood factors, and year-quarter fixed effects. The outcomes of this analysis are illustrated in Figure 8, and the visual representation makes it evident that properties characterized by higher value uncertainty generally sell at prices considerably lower than their initial asking prices. This finding aligns with Sagi (2021), who demonstrates that buyer-seller heterogeneity in private values plays a central role in explaining the idiosyncratic price dispersion.

Additionally, the trade-off between time on the market and asking price discount becomes greater with value uncertainty. It is a well-known trade-off that, to sell a house quickly, a seller has to accept a cut to the original asking price. I tested the effect of

Figure 8: *Predicted dispersion and asset level liquidity, Cologne (2012-2022)*

(a) *Hamburg - Asking price discount* **(b)** *Cologne - Asking price discount*



Note: These figures display a binned scatter plot, based on a regression of spread on predicted dispersion, property characteristics and year-quarter and neighborhood fixed effects. The data shown is for the city of Cologne and Hamburg for the period between 2012 and 2022.

value uncertainty on this trade-off by running a regression of asking price discount on time on the market at the transaction level, where I also included an interaction term between time on the market and predicted dispersion. The results can be found in Table 5. The sign of the interaction coefficient is negative, meaning that for a given level of time on the market, a seller will need to accept a larger cut to the asking price when selling a property with more price uncertainty.

Table 5: *Trade-off TOM and price discount, Cologne (2012-2022)*

	Price Discount	Price Discount	Price Discount
TOM	-0.03***	0.04	0.04
	(0.007)	(0.026)	(0.025)
TOM × Idiosyncratic risk		-0.64**	-0.63***
		(0.198)	(0.181)
Idiosyncratic risk		-96.90***	-87.49***
		(15.957)	(14.083)
Year FEs	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes
Property characteristics	No	No	Yes
N	12830	12800	12800
R^2	0.02	0.05	0.06

Standard errors are clustered at the neighborhood-level (Stadtbezirk). Singletons were dropped. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 6: *Trade-off TOM and price discount, Hamburg (2012-2022)*

	Price Discount	Price Discount	Price Discount
TOM	0.01	0.08***	0.06***
	(0.009)	(0.019)	(0.019)
TOM × Idiosyncratic risk		-0.56***	-0.54***
		(0.155)	(0.160)
Idiosyncratic risk		-35.95***	-20.39***
		(5.290)	(4.830)
Year FEs	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes
Property characteristics	No	No	Yes
N	22861	22513	22513
R^2	0.02	0.03	0.04

Standard errors are clustered at the neighborhood-level (Stadtbezirk). Singletons were dropped. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

7 Conclusion

Despite the extensive literature on the microstructure of housing markets, which has emphasized the significance of housing liquidity for patterns in housing prices, relatively little attention has been given to the interplay between liquidity in the rental and sales markets and its impact on transaction prices and returns. I begin by constructing a bargaining model involving a risk-averse investor who encounters uncertainty regarding future rental income and the property's value. The model predicts that properties with higher value uncertainty will be traded at lower prices and yield higher returns. To test the model's predictions, I utilize a novel transaction-level dataset encompassing all real estate transactions in major German cities over the past four decades. In each of the four cities in my sample, I find robust evidence that supports all three predictions of the model. In the context of the German housing markets, higher returns on properties with greater value uncertainty are plausible, given the larger size and liquidity of rental markets compared to sales markets.

While Germany has grappled with persistently low homeownership rates for decades, this paper sheds light on how the substantial size and liquidity of the rental market may impede policies aimed at increasing homeownership rates, as they provide higher returns for housing investments, making buy-to-let investments highly attractive.

References

Amaral, Francisco, Martin Dohmen, Sebastian Kohl, et al. (2021). "Superstar Returns". In: *FRB of New York Staff Report*(999).

Amaral, Francisco, Martin Dohmen, Moritz Schularick, et al. (2023). "German Real Estate Index (GREIX)". In.

Bali, Turan G, Robert F Engle, and Scott Murray (2016). *Empirical asset pricing: The cross section of stock returns*. John Wiley & Sons.

Bourassa, Steven C et al. (2009). "House price changes and idiosyncratic risk: The impact of property characteristics". In: *Real Estate Economics* 37(2), pp. 259–278.

Buchak, Greg et al. (2020). *Why is intermediating houses so difficult? evidence from ibuyers*. Tech. rep. National Bureau of Economic Research.

Cannon, Susanne, Norman G Miller, and Gurupdesh S Pandher (2006). "Risk and Return in the US Housing Market: A Cross-Sectional Asset-Pricing Approach". In: *Real Estate Economics* 34(4), pp. 519–552.

Conklin, James N et al. (2023). "An alternative approach to estimating foreclosure and short sale discounts". In: *Journal of Urban Economics* 134, p. 103546.

DeGroot, Morris H (2005). *Optimal statistical decisions*. John Wiley & Sons.

Demers, Andrew and Andrea L Eisfeldt (2022). "Total returns to single-family rentals". In: *Real Estate Economics* 50(1), pp. 7–32.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen (2007). "Valuation in over-the-counter markets". In: *The Review of Financial Studies* 20(6), pp. 1865–1900.

Duranton, Gilles and Diego Puga (2015). "Urban land use". In: *Handbook of regional and urban economics*. Vol. 5. Elsevier, pp. 467–560.

Ehsani, Sina and Juhani T Linnainmaa (2022). "Factor momentum and the momentum factor". In: *The Journal of Finance* 77(3), pp. 1877–1919.

Fama, Eugene F and Kenneth R French (2015). "A five-factor asset pricing model". In: *Journal of financial economics* 116(1), pp. 1–22.

Gavazza, Alessandro (2011). "The role of trading frictions in real asset markets". In: *American Economic Review* 101(4), pp. 1106–1143.

Giacoletti, Marco (2021). "Idiosyncratic Risk in Housing Markets". In: *The Review of Financial Studies* 34(8), pp. 3695–3741.

Glaeser, Edward L and Joseph Gyourko (2007). *Arbitrage in housing markets*. Tech. rep. National Bureau of Economic Research.

Goetzmann, William N, Christophe Spaenjers, and Stijn Van Nieuwerburgh (2021). "Real and private-value assets". In: *The Review of Financial Studies* 34(8), pp. 3497–3526.

Haan, Jan de and WE Diewert (2011). "Handbook on residential property price indexes". In: *Luxembourg: Eurostat*.

Han, Lu (2013). "Understanding the Puzzling Risk-Return Relationship for Housing". In: *The Review of Financial Studies* 26(4), pp. 877–928.

Han, Lu and William C Strange (2015). "The microstructure of housing markets: Search, bargaining, and brokerage". In: *Handbook of regional and urban economics* 5, pp. 813–886.

Haurin, Donald (1988). "The duration of marketing time of residential housing". In: *Real Estate Economics* 16(4), pp. 396–410.

Hill, Robert J (2013). "Hedonic price indexes for residential housing: A survey, evaluation and taxonomy". In: *Journal of economic surveys* 27(5), pp. 879–914.

Jiang, Erica Xuewei and Anthony Lee Zhang (2022). "Collateral Value Uncertainty and Mortgage Credit Provision". In: *Available at SSRN*.

Kohl, Sebastian (2017). *Homeownership, renting and society: Historical and comparative perspectives*. Taylor & Francis.

Kotova, Nadia and Anthony Lee Zhang (2021). *Liquidity in residential real estate markets*. Tech. rep. Working paper.

Mian, Atif and Amir Sufi (2009). "The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis". In: *The Quarterly journal of economics* 124(4), pp. 1449–1496.

Pagano, Marco (1989). "Trading volume and asset liquidity". In: *The Quarterly Journal of Economics* 104(2), pp. 255–274.

Piazzesi, Monika and Martin Schneider (2016). "Housing and macroeconomics". In: *Handbook of macroeconomics* 2, pp. 1547–1640.

Rubinstein, Ariel and Asher Wolinsky (1985). "Equilibrium in a market with sequential bargaining". In: *Econometrica: Journal of the Econometric Society*, pp. 1133–1150.

RWI and Immobilienscout24 (2021). *RWI Real Estate Data - Apartments for rent and sale-suf. RWI-GEO-RED*. RWI -Leibniz Institute for Economic Research.

Sagi, Jacob S (2021). "Asset-level risk and return in real estate investments". In: *The Review of Financial Studies* 34(8), pp. 3647–3694.

Sinai, Todd and Nicholas S Souleles (2005). "Owner-occupied Housing as a Hedge Against Rent Risk". In: *The Quarterly Journal of Economics* 120(2), pp. 763–789.

Van Nieuwerburgh, Stijn and Pierre-Olivier Weill (2010). "Why has house price dispersion gone up?" In: *The Review of Economic Studies* 77(4), pp. 1567–1606.

Appendix

A Idiosyncratic Price Dispersion in Housing Markets

A broad literature, which started with Case and Shiller (1988), has argued that the idiosyncratic component of prices is the largest determinant of capital gains for individual houses. Idiosyncratic house price risk is defined as the property-level capital gains not explained by local market fluctuations and common house or transaction characteristics. It can thus be estimated as the standard deviation of the residuals of a regression of house price appreciation on a set of controls:

$$\Delta p^h_{t+1} \quad = \Delta v_{t+1} + \quad BX^h + \quad \sigma_{l,residual}\ \rho^h_{t+1} \tag{35}$$

where $\Delta v_{l,t}$ represents the average growth of local house prices, X^h is a vector of property and transaction-specific characteristics that might impact the price and ε^h_t can be interpreted as a transaction-specific shock. Using very rich transaction-level data, recent work has demonstrated empirically that most of the variation in house prices is indeed idiosyncratic (Giacoletti, 2021; Sagi, 2021). In addition to finding a large amount of idiosyncratic volatility in house prices, these papers also show that the idiosyncratic component of volatility almost does not scale with the holding period. Instead, idiosyncratic volatility seems to stem mostly from the sale and re-sale of the property. This suggests that transaction frictions might explain most of the idiosyncratic risk in housing.

Sagi (2021) builds and calibrates a heterogeneous agents search model of the housing market, in which idiosyncratic volatility in house prices arises from limited trading opportunities and heterogeneity in valuations. In the model dispersion in the relative valuations of randomly matched counterparties and limited trading opportunities leads to uncertainty about the matching process and, therefore, to transaction risk. Illiquidity can thus amplify the house price risk. Sagi shows that one can write the log total return on a property for a specific period as the sum of a market-wide price shock μ_m, a shock to the value of the housing services η_{inc} and a transaction-specific shock ρ_{trans}:

$$\Delta r^h_{t+1} \quad = \sigma_m \mu_{m,t+1} + \sigma_{inc}\eta_{inc,t+1} + \sigma_{trans}\rho^h_{trans,t+1}, \tag{36}$$

where the housing services shock component can be decomposed into a market and an idiosyncratic component: $\sigma_{inc}\eta_{inc} = \sigma_{inc,m}\eta_{inc,m} + \sigma_{inc,idio}\eta_{inc,idio}$. The idiosyncratic component of the housing services shock together with the transaction shock build the property-level risk. In particular, the transaction shock maps one-to-one to the residual shock in equation 36. Sagi is also able to show empirically that the majority of property-level risk arises from the transaction risk and not from the idiosyncratic housing services shock component. As such, I will focus on this transaction shock as a source of risk premium in the theoretical analysis that follows.

B Idiosyncratic risk in asset pricing

Sagi (2021) constructs a search model of the housing market, demonstrating that heterogeneity in the private valuations of sellers and buyers, coupled with market illiquidity, play a crucial role in explaining the deviations of individual real estate prices from their expected market values. When these deviations persistently occur for specific properties, meaning a property consistently transacts at a price significantly different from its expected market value, it becomes a source of risk for potential buyers. Consequently, buyers may consider this source of risk when determining the value they are willing to pay for a specific house. In other words, this idiosyncratic risk might be priced into housing markets.

However, in most asset-pricing theories, idiosyncratic risk is not priced in equilibrium, as it can be diversified away (Cochrane, 2009). Consequently, existing pricing models of the housing market often rely on this framework, concentrating solely on common risk factors (Piazzesi et al., 2007; Case, Cotter, et al., 2011). However, in an incomplete markets setting, households may not be able to fully diversifying their portfolios, thereby exposing them to idiosyncratic risks.

Inspired by the vast empirical evidence that households' consumption reacts to house price shocks (Attanasio et al., 2011; Mian et al., 2013; Stroebel and Vavra, 2019), I will setup a model in which idiosyncratic consumption depends on house price volatility. Based on household-level data, several papers have been able to isolate confounding factors, such as income expectations, to show that unexpected house price shocks causally lead to changes in consumption (e.g. Campbell and Cocco, 2007). More recently, Berger et al. (2018) show that large consumption responses to house price movements are fully in line with workhorse models of consumption with incomplete markets.

I build on the seminal work by Constantinides and Duffie (1996) to construct a housing asset-pricing model in which idiosyncratic price risk is a priced state variable, as idiosyncratic housing price volatility affects the average household's marginal utility through consumption. Since households cannot completely insulate their consumption from persistent shocks to their income (Blundell et al., 2008), the volatility of households' consumption growth distribution inherits the same factor structure as the volatility in property-level prices. In other words, persistent, idiosyncratic price shocks that hit houses are an important source of undiversifiable risk to households. Housing price risk thus enters the pricing kernel of households and, as a result, is a priced state variable.

This approach is very similar to the work from Herskovic et al. (2016). The authors build a heterogenous agent incomplete markets model, in which the idiosyncratic component of households' consumption follows the same structure as the common idiosyncratic volatility of firms' dividends. Since households cannot diversify away the idiosyncratic component of consumption, the firms' idiosyncratic volatility is priced in equilibrium. In contrast to stocks, idiosyncratic volatility in real estate arises mostly from shocks to house prices at the point of sale and re-sale. As such, it makes more sense to think of house price shocks as being the source of undiversifiable risk to households, rather than shocks to value of housing dividends (rents).

The theoretical framework presented here has two main caveats. Firstly, it considers only one risky asset, namely housing, thus ignoring any common risk sources that might arise from the covariance structure of returns to housing and other assets. Since this paper focuses on idiosyncratic risk, I abstract from several sources of common risk. Secondly, the model views housing as an investment good and does not consider its nature as a consumption good. However, as I demonstrate in more detail, this abstraction does not affect the main results of the model.

B.1 A Consumption Asset-Pricing Model with Idiosyncratic Risk

B.1.1 Setup

Households can invest both in a riskless asset with return r^f and in housing with return $R^h_{t+1} = \frac{P^h_{t+1}+D^h_{t+1}}{P^h_t}$, where P_t is the price and D_t the value of the housing services provided by the house at time t.

Following Berk et al. (1999), I parameterize directly the pricing kernel without explicitly modelling the consumer's problem. The individual log stochastic discount factor is then:

$$m^i_{t+1} = log\beta - \gamma b^i \left[\sigma_{p,t+1} v_{idio,t+1} + \sigma_{idio,t+1}\rho^i_{t+1} - \frac{1}{2}\sigma^2_{idio,t+1}\right] \tag{37}$$

This equation can be motivated by assuming a fictitious consumer side problem with heterogeneous agents with power utility and a relative risk aversion coefficient, γ. The heterogeneity among home buyers comes from the fact that they will have different consumption sensitivities to house price shocks, given by the parameter b^i. This is motivated by the empirical evidence on the heterogeneity of consumption responses to house price shocks, for ex. older households respond more than younger households (Campbell and Cocco, 2007). Under this setup we can write the log individual sdf as:

$$m^i_{t+1} = log\beta - \gamma[b^i \Delta c^i_{t+1}]. \tag{38}$$

Following Constantinides and Duffie (1996) I will write log individual consumption c^i both as function of aggregate log consumption c^A as well as of the individual log share of aggregate consumption s^i:

$$\Delta c^i_{t+1} = b^i(\Delta c^A_{t+1} + \Delta s^i_{t+1}), \tag{39}$$

By linking aggregate consumption and the individual shares to house price shocks in reduced form as:

$$\Delta c^A_{t+1} = \sigma_{p,t+1} v_{t+1} \tag{40}$$

$$\Delta s^i_{t+1} = \sigma_{idio,t+1} \rho^i_{t+1} - \frac{1}{2}\sigma^2_{idio,t+1} \tag{41}$$

Equation 37 follows from the above equations. Note that while aggregate consumption growth is homoskedastic, individual consumption growth is not:

$$E_i[\Delta s^i_{t+1}] = -\frac{1}{2}\sigma^2_{idio,t+1} \tag{42}$$

$$V_i[\Delta s^i_{t+1}] = \sigma^2_{idio,t+1} \tag{43}$$

Assuming there are N housing investors in the economy, and that these investors have the same level of risk aversion we can write the average markets' sdf as the sum of the individual investors' sdf. Define $E_i = \frac{1}{N}\sum_{i=1}^{N}$ and note that since the idiosyncratic shock has mean zero, then, applying the law of large numbers, the term $\sigma_{idio}\rho^i$ converges to zero when we sum over the individual investors. Then we have:

$$E_i m^i_{t+1} = E_i(-\delta - \gamma \Delta c^i_{t+1}) \tag{44}$$

$$m^m_{t+1} = -\delta - \gamma(\sigma_{p,t+1} v_{t+1}) + \frac{1}{2}\sigma^2_{idio,t+1}, \tag{45}$$

where it becomes clear that the markets' sdf not only varies with aggregate housing price volatility, but also with the cross-sectional variance of housing prices, which is determined by the idiosyncratic housing price shocks. Assuming the pricing kernel is derived from the FOC of the consumer problem, I can write the riskless asset log return as:

$$\begin{aligned} 1 &= E_t[m^m_{t+1} r^f_{t+1}] \\ \iff r^f_{t+1} &= \delta + \gamma(\sigma_{p,t+1} v_{t+1}) - \frac{1}{2}\sigma^2_{idio,t+1}, \end{aligned} \tag{46}$$

the log housing premium as:

$$E_t(r^h_{t+1}) - r^f_{t+1} = -r^f_{t+1} * cov(m^m_{t+1}, r^h_{t+1}), \tag{47}$$

where r^h_{t+1} is the total return to house h in period $t+1$. Multiplying and dividing by the variance of the markets' stochastic discount factor we can write the log housing premium in the standard beta representation form as:

$$E_t(r^h_{t+1}) - r^f_{t+1} = \beta_p \gamma \sigma^2_{p,t+1} + \beta_{idio} \gamma \sigma^2_{idio,t+1}. \tag{48}$$

This equation provides a linear relation between housing excess returns, systematic risk, σ_p, and idiosyncratic risk, σ_{idio}. In the next sections, I will first measure σ_{idio} and then test whether I can find a significant impact on housing returns.

C Transaction-level data set

C.1 All sales

Table 7: *Summary statistics for all apartment sales by city*

	Berlin					
	N	Mean	SD	P25	Median	P75
Price (thousand €)	190144	216	176.4	93.3	160.1	285
Size (m^2)	190144	76	29.9	55.1	70	92.7
Construction year	190144	1943	45.5	1905	1928	1991
Residuals, $u_{i,tq}$ (%)	190144	0	29.7	-17.6	0	19.2
Rental yield (%)	190144	3.4	1.8	2.2	3	4.1
	Hamburg					
	N	Mean	SD	P25	Median	P75
Price (thousand €)	81840	296	256.6	128	222.5	376
Size (m^2)	81840	77	31	56	72	92.5
Construction year	81840	1972	38.4	1955	1974	2008
Residuals, $u_{i,tq}$ (%)	81840	0	24.8	-13.3	0	15.4
Rental yield (%)	81840	4.3	2	3	3.9	5.2
	Cologne					
	N	Mean	SD	P25	Median	P75
Price (thousand €)	108103	159	124.4	80	122	195
Size (m^2)	108103	71	25.8	54	69	86
Construction year	108103	1972	25.6	1959	1972	1990
Residuals, $u_{i,tq}$ (%)	108103	0	25.1	-14.3	0	15.8
Rental yield (%)	108103	5.5	2.3	3.9	5.2	6.7
	Duesseldorf					
	N	Mean	SD	P25	Median	P75
Price (thousand €)	48893	184	175.7	81.3	126	214
Size (m^2)	48893	76	30.2	55	72	93
Construction year	48893	1965	27.5	1954	1965	1982
Residuals, $u_{i,tq}$ (%)	48893	0	26.3	-15.3	0	15.8
Rental yield (%)	48893	4.9	2.2	3.5	4.6	5.8

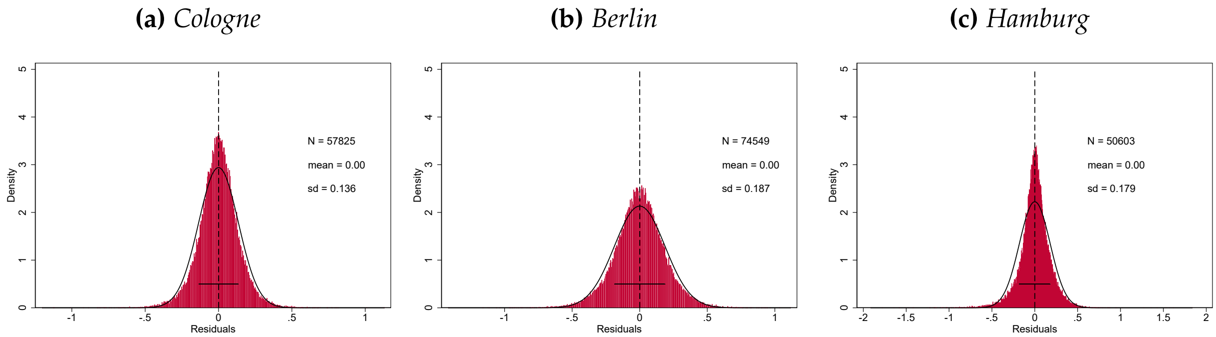
Note: Table reports summary statistics for all apartment sales for Berlin (1986-2022), Hamburg (2002-2022), Cologne (1989-2022) and Duesseldorf (1984-2022). Note that before 1992 the data for Berlin refers only to West-Berlin. Prices are in nominal terms.

C.2 Distribution of idiosyncratic price deviations

C.3 Transaction data for Hamburg

In this section of the appendix, I provide a more detailed description of the method used to measure price dispersion at the transaction apartment level in the city of Hamburg. In the original dataset containing transactions for the city of Hamburg, information on apartment identification is missing in most cases. Consequently, it is impossible to identify repeated transactions of the same apartments over time. This limitation accounts for the low number of repeated sales available for the analysis of the

Figure 9: *Distribution of idiosyncratic price deviations by city*



This figure shows the distribution of residuals from equation (1) *for Cologne (a), Berlin (b) and Hamburg (c).*

effects of predicted dispersion on total returns and capital gains. Therefore, for Hamburg, I measure price dispersion without including apartment fixed effects. I employ the following specification to measure price deviations at the transaction apartment level:

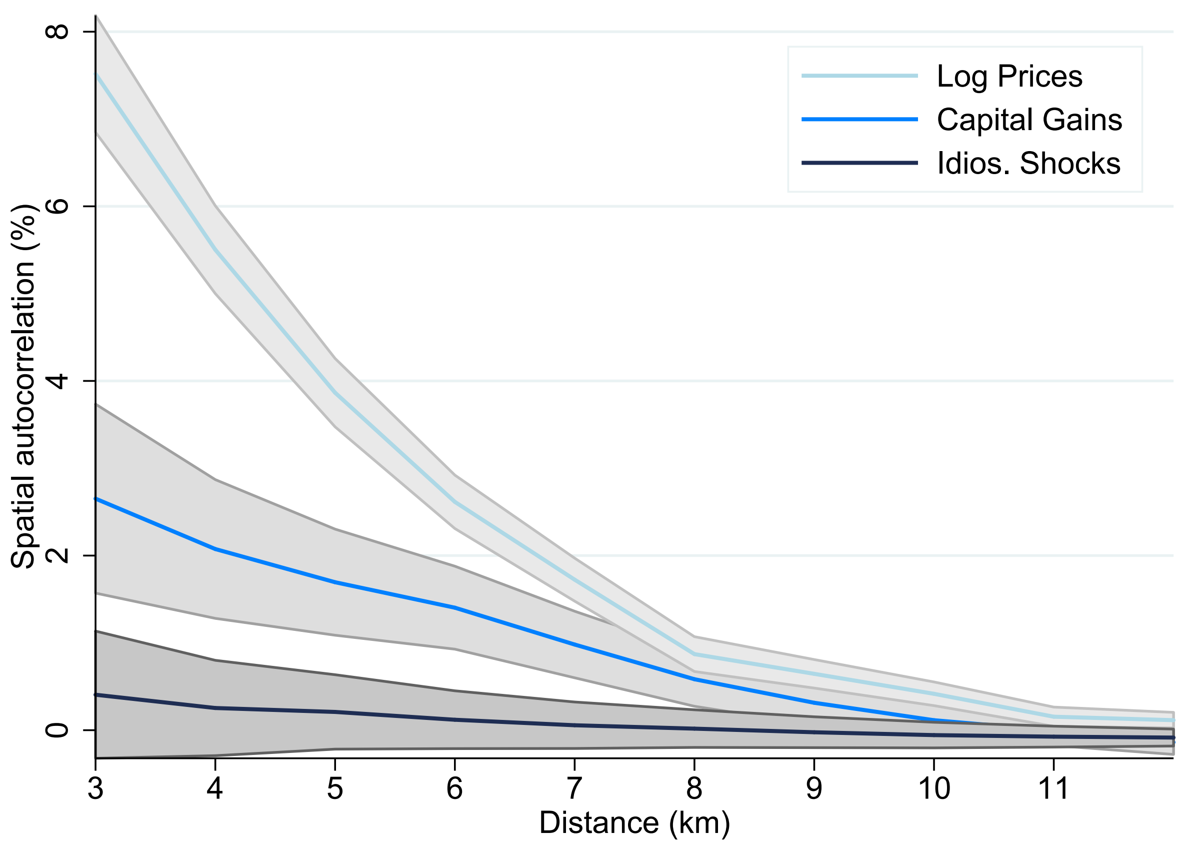
$$ln(p_{i,tq}) = b_i + \eta_{tm} + \kappa_{n,tq} + f_c(x_i, ty) + u_{i,tq}, \qquad (49)$$

where $u_{i,tq}$ is a mean-0 error term with variance σ^2 and b_i is a building fixed-effect. The other terms in the regression are the same as in the baseline specification (1). The most significant deviation from the baseline specification is that I am no longer controlling for apartment-specific features. Instead, I am accounting for features that remain constant within the building, such as the exact location.

D Distribution of dispersion across space and time

By definition, a idiosyncratic shock should be uncorrelated with common shocks. More precisely, for each property i $E[e_{it}\mu_t] = 0$. Measuring common shocks directly is extremely complicated in housing markets, since this would require additional data on the supply and demand of housing markets. However, it is possible to measure common movements in the market. In other words, if the value of an apartment in the changes in response to a common shock, then we would expect the values of similar apartments nearby to also change. On the other hand, if the value changes in response to an idiosyncratic (property-level) shock, then we do not expect the value of similar apartments nearby to change. The idiosyncratic component of housing prices should be independently distributed across apartments. To test for this, I estimate spatial correlation in idiosyncratic shocks using Morans'I . A positive Morans'I indicates that apartments with positive residuals are surrounded by other apartments with positive residuals.[28] Figure 10 plots Morans'I for sales prices, property-level capital gains and idiosyncratic shocks of apartments sold in the same year in Cologne for the period between 1989 and 2022. Log sales prices and capital gains show a positive and significant spatial autocorrelation, which intuitively decreases with distance. If an apartment is sold for a high price, then probably the neighboring apartments will also sell for a high price. For idiosyncratic price shocks, I cannot reject the hypothesis that the correlation is 0, even when looking only at apartments sold within a three kilometer radius.

Figure 10: *Spatial autocorrelation in housing market outcomes*



Note: Morans'I is estimated for apartments sold in the same year and within the given km radius for the city of Cologne between 1989 and 2022. The Figure shows the simple average across years of Morans'I for each radius. 95% confidence interval bands are shown in the shaded areas.

For the idiosyncratic shocks to matter, their variance needs to be persistent over time. If idiosyncratic shocks to housing prices would be transitory, then one could easily make the argument that a buyer should not care about such shocks. In other words, I want to

[28]Morans'I test for spatial autocorrelation is estimated as:

$$I = \frac{N}{\sum_i \sum_j w_{i,j}} \frac{\sum_i \sum_j w_{i,j}(x_j - \bar{x})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

with d_{ij} being the distance between apartment i and j in kilometers, k is the maximum radius in km and w_{ij} is one if the distance between i and j is smaller than k.

test whether a large shock to a specific property todays, predicts a large shock in the future. Following the empirical evidence on idiosyncratic housing price shocks, here I am considering shocks that occur at the points of sale and re-sale. Specifically, I test for all pairs of transactions in the data set whether the variance of the shock at the point of sale predicts the variance of the shock at the point of re-sale:

$$u_{i2}^2 = \beta_1 u_{i1}^2 + \beta_2 hp_i + \kappa_{nt} + \lambda_m + \epsilon_{it}, \tag{50}$$

where e_{i2} and e_{i1} are the idiosyncratic price shocks at the points of re-sale and sale respectively of property i. hp_i measures the holding period in months for property i, while δ_m are monthly fixed effects and κ_t are neighborhood fixed effects. The results can be found in Table 8, which shows that properties sold and re-sold on the same in the same month and neighborhood show considerable persistence in their idiosyncratic shocks. An increase in one standard deviation of the sales' shock predicts an increase in 0.66 standard deviations in the resale shock. One concern is that these results are being driven by the buyers, if a specific buyer is bad at pricing a house at the moment of sale, then probably as well at the moment of re-sale. This could potentially explain the high level of persistence in the variance. To address this concern, I show that the persistence in variance is also strongly positive and statistically significant when testing the relation between first and third sale. The results can be found in Table 9 in Appendix **??**. Additionally, the cross-sectional correlation at the point of sale and re-sale of idiosyncratic shocks is 0.66, which is higher that than most risk factors used in the stock pricing literature (Bali et al., 2016). The results can also be found in Appendix **??**.

Table 8: *Persistence in the variance of idiosyncratic shocks*

	u_{i2}^2	u_{i2}^2
u_{i1}^2	0.6485***	0.6479***
	(0.0166)	(0.0167)
Holding period	Yes	Yes
Month-sale FEs	Yes	Yes
Neighborhood FEs	No	Yes
N	34060	34060
R^2	0.43	0.43

Standard errors are clustered at the neighborhood-level (Stadtbezirk). Coefficients are standardized. Singletons were dropped. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

In Table 9, I test whether the variance of the idiosyncratic shock at the point of the first sales predicts the variance of the shock at the third sale.

In Figure 11, I plot the Pearson cross-sectional correlation in idiosyncratic shock

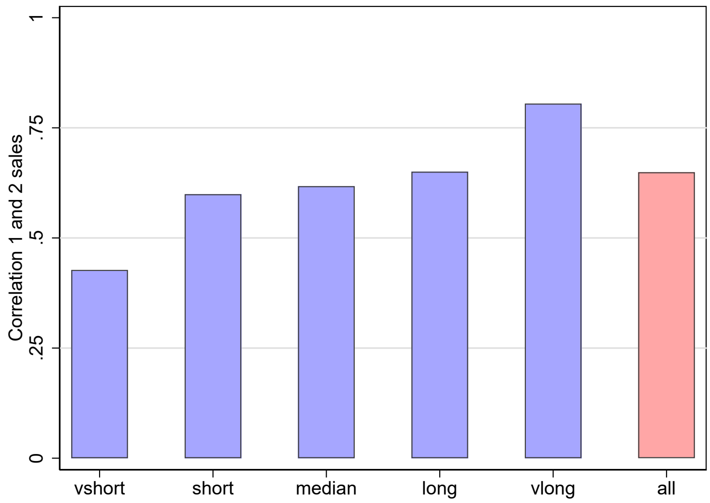
Table 9: *Persistence in the variance of idiosyncratic shocks*

	u_{i3}^2	u_{i3}^2
u_{i1}^2	0.3161***	0.3144***
	(0.0257)	(0.0260)
Holding period	Yes	Yes
Month-sale FEs	Yes	Yes
Neighborhood FEs	No	Yes
N	7244	7244
R^2	0.13	0.14

Standard errors are clustered at the neighborhood-level (Stadtbezirk). Coefficients are standardized. Singletons were dropped. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

variance by holding period. The average cross-section correlation is 0.66 indicating that the variance of the shocks is highly persistent over time. In contrast, the cross-sectional correlation of market betas is at most 0.60 and decreases with the time distance. In the case of idiosyncratic housing price shocks, the cross-section correlation increases with the holding period, indicating that the variance of the shocks is more persistent for pairs of sales more distant in time.

Figure 11: *Pearson cross-section correlation in the predicted price dispersion*



Note: Figure shows Pearson cross-section correlation of standardized residuals from sale 1 and 2 for different holding periods.

E Hedonic price and rental yield indices

In this section of the appendix, I describe the hedonic methods employed to construct the two components of the total housing return portfolio quarterly time-series. Both for the price index as well as the rental yield index, I employ a rolling-window time-dummy hedonic index. The rolling-window component assures that the coefficients can change over time, i.e. the effect of age on the price can change over time. I set the rolling window at 20 quarters. More specifically, I employ the following log-linear specification:

$$ln(y_{i,tq}) = \beta^0 + \sum_{\tau}^{20} \gamma_\tau D_\tau + \sum_{k=1}^{K} (\beta^k x_i^k) + \epsilon_{i,tq}, \tag{51}$$

where the log dependent variable (transaction price, rental yield) for property i in quarter tq is regressed on a time-dummy D_τ and a set of property characteristics x_i, which consist of apartment size, age and neighborhood.

F Idiosyncratic price uncertainty, sales prices and rents

F.1 Regression results for the main sample

In this section of the appendix, I provide the regression output tables for the analyses that form the basis of Figure **??** in the paper. Figure **??** illustrates the relationship between predicted price dispersion and both sales prices and rents for each city in the sample. The regression output is displayed in the following tables for each city separately. Please note that the specification that underlies the binned scatter in the paper is always in columns 2 and 4 for sales price and net rent respectively. From the tables, it is visible that the coefficient of predicted price dispersion on sales prices is much larger than the one on net rents. Additionally, the coefficient on net rents is mostly statistically insignificant, indicating that rents decrease very only slightly with idiosyncratic price uncertainty.

Table 10: *Predicted price dispersion, sales prices and rent (Berlin)*

	Sales Price	Sales Price	Net Rent	Net Rent
Predicted dispersion, $\hat{\sigma}_{it}$	-13.88***	-0.65***	-22.03***	0.15
	(1.341)	(0.129)	(2.708)	(0.183)
Year-quarter FEs	Yes	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes	Yes
Property characteristics	No	Yes	No	Yes
N	67194	67194	67194	67194
R^2	-3.89	0.70	-8.08	0.93

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). The outcome variables have been standardized to have mean 0 and standard deviation of one. Singletons were dropped. The explanatory variable of interest is predicted price dispersion. The coefficients are estimated in the 2SLS regression framework of (7). $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

Table 11: *Predicted price dispersion, sales prices and rent (Hamburg)*

	Sales Price	Sales Price	Net Rent	Net Rent
Predicted dispersion, $\hat{\sigma}_{it}$	-9.72***	-1.51***	-12.13***	0.31
	(0.976)	(0.209)	(1.338)	(0.227)
Year-quarter FEs	Yes	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes	Yes
Property characteristics	No	Yes	No	Yes
N	52647	52647	52647	52647
R^2	-2.06	0.70	-2.61	0.92

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). The outcome variables have been standardized to have mean 0 and standard deviation of one. Singletons were dropped. The explanatory variable of interest is predicted price dispersion. The coefficients are estimated in the 2SLS regression framework of (7). $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

Table 12: *Predicted price dispersion, sales prices and rent (Cologne)*

	Sales Price	Sales Price	Net Rent	Net Rent
Predicted dispersion, $\hat{\sigma}_{it}$	-34.36***	-1.45***	-42.12***	0.80*
	(3.671)	(0.443)	(4.456)	(0.424)
Year-quarter FEs	Yes	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes	Yes
Property characteristics	No	Yes	No	Yes
N	50029	50029	50029	50029
R^2	-4.30	0.82	-6.84	0.96

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). The outcome variables have been standardized to have mean 0 and standard deviation of one. Singletons were dropped. The explanatory variable of interest is predicted price dispersion. The coefficients are estimated in the 2SLS regression framework of (7). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 13: *Predicted price dispersion, sales prices and rent (Duesseldorf)*

	Sales Price	Sales Price	Net Rent	Net Rent
Predicted dispersion, $\hat{\sigma}_{it}$	-17.71***	-2.47***	-18.58***	-0.80***
	(1.326)	(0.314)	(1.379)	(0.189)
Year-quarter FEs	Yes	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes	Yes
Property characteristics	No	Yes	No	Yes
N	25971	25971	25971	25971
R^2	-3.78	0.68	-4.63	0.86

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). The outcome variables have been standardized to have mean 0 and standard deviation of one. Singletons were dropped. The explanatory variable of interest is predicted price dispersion. The coefficients are estimated in the 2SLS regression framework of (7). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

F.2 Regression results for the sub-sample

In this subsection, I examine the relationship between predicted price dispersion and sales prices and rents, utilizing a subsample of observations for which data on both rent and sales price are available. The results mirror the patterns observed in the analysis of the full sample, indicating that prices tend to decrease more than rents in response to idiosyncratic price uncertainty. However, it is noteworthy that in some cities, the coefficients loose statistical significance. This was expected given the considerable reduction in the sample size.

Table 14: *Idiosyncratic price uncertainty, sales prices and rent (Berlin-subsample)*

	Sales Price	Sales Price	Net Rent	Net Rent
Predicted dispersion, $\hat{\sigma}_{it}$	-1.93***	-1.31***	-2.48***	-0.60*
	(0.454)	(0.326)	(0.516)	(0.317)
Year-quarter FEs	Yes	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes	Yes
Property characteristics	No	Yes	No	Yes
N	13466	13466	13466	13466
R^2	0.04	0.72	-0.07	0.62

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). The outcome variables have been standardized to have mean 0 and standard deviation of one. Singletons were dropped. The explanatory variable of interest is predicted price dispersion. The coefficients are estimated in the 2SLS regression framework of (7). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 15: *Idiosyncratic price uncertainty, sales prices and rent (Hamburg-subsample)*

	Sales Price	Sales Price	Net Rent	Net Rent
Predicted dispersion, $\hat{\sigma}_{it}$	-1.89***	-0.86***	-1.88***	-0.42
	(0.378)	(0.247)	(0.494)	(0.322)
Year-quarter FEs	Yes	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes	Yes
Property characteristics	No	Yes	No	Yes
N	8651	8651	8651	8651
R^2	0.03	0.66	-0.03	0.56

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). The outcome variables have been standardized to have mean 0 and standard deviation of one. Singletons were dropped. The explanatory variable of interest is predicted price dispersion. The coefficients are estimated in the 2SLS regression framework of (7). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 16: *Idiosyncratic price uncertainty, sales prices and rent (Duesseldorf-subsample)*

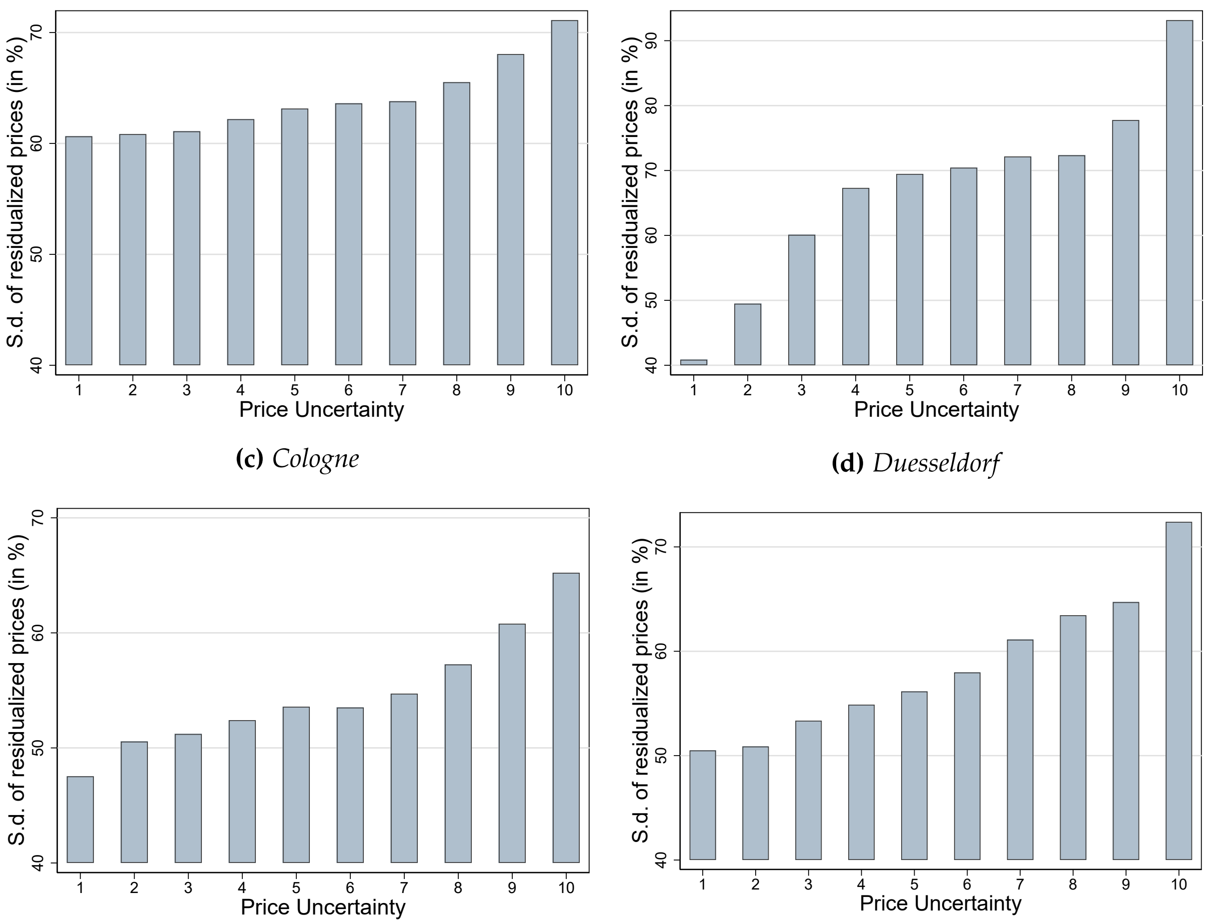
	Sales Price	Sales Price	Net Rent	Net Rent
Idiosyncratic uncertainty, $\hat{\sigma}_{it}$	-5.85***	-0.50	-6.67***	0.08
	(0.766)	(0.408)	(0.868)	(0.169)
Year FEs	Yes	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes	Yes
Property characteristics	No	Yes	No	Yes
N	1321	1321	1321	1321
R^2	0.40	0.86	0.25	0.96

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). The outcome variables have been standardized to have mean 0 and standard deviation of one. Singletons were dropped. The explanatory variable of interest is predicted price dispersion. The coefficients are estimated in the 2SLS regression framework of (7). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

G Price uncertainty and dispersion of prices

In this section of the appendix, I show how price uncertainty also affects the second moment of the distribution of transaction prices. Since I do not observe properties that have been sold every period, my analysis will explore the cross-sectional variation in prices of similar houses. My analysis will follow two steps. In the first step, I residualize the transaction prices using property characteristics and time and location fixed effects. In a second step, I split the transactions into 10 different bins depending on their level of price uncertainty. I then calculate the standard deviation of prices within those bins and plot them in Figure 12. From the Figure it becomes clear that properties with higher ex-ante price uncertainty also have a higher standard deviation of prices. This means that price uncertainty predicts lower prices and higher dispersion.

Figure 12: *Price uncertainty and dispersion of prices*



(a) *Berlin* **(b)** *Hamburg* **(c)** *Cologne* **(d)** *Duesseldorf*

Note: The figure displays the standard deviation of residualized log transaction prices across the price uncertainty distribution for the different cities in the data set.

H Predicted dispersion and returns to housing - regression output

Table 17: *Predicted dispersion and total returns, Berlin (1984-2022)*

	Rental Yields	Capital Gains	Total Returns
Predicted dispersion, $\hat{\sigma}_{i,tq}$	2.56***	4.70	7.71**
	(0.445)	(2.806)	(3.061)
Year-quarter FEs	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes
Property characteristics	Yes	Yes	Yes
Holding period FEs	No	Yes	Yes
N	67194	33309	33309
R^2	0.12	0.35	0.32

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk) and are adjusted for the estimated regressors. Singletons were dropped. The explanatory variable of interest is predicted dispersion. The first column displays the results of 2SLS regressions as in (7). *Columns 2 and 3 display the results of the two-step regression as in* (13). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 18: *Predicted dispersion and total returns, Hamburg (2001-2022)*

	Rental Yields	Capital Gains	Total Returns
Predicted dispersion, $\hat{\sigma}_{i,tq}$	8.72***	-1.04	1.74
	(0.768)	(7.500)	(8.281)
Year-quarter FEs	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes
Property characteristics	Yes	Yes	Yes
Holding period FEs	No	Yes	Yes
N	49506	1741	1741
R^2	-0.06	0.27	0.28

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk) and are adjusted for the estimated regressors. Singletons were dropped. The explanatory variable of interest is predicted dispersion. The first column displays the results of 2SLS regressions as in (7). *Columns 2 and 3 display the results of the two-step regression as in* (13). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 19: *Predicted dispersion and total returns, Cologne (1989-2022)*

	Rental Yields	Capital Gains	Total Returns
Predicted dispersion, $\hat{\sigma}_{i,tq}$	16.28***	7.61	16.77**
	(3.368)	(4.544)	(6.220)
Year-quarter FEs	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes
Property characteristics	Yes	Yes	Yes
Holding period FEs	No	Yes	Yes
N	49963	27069	27069
R^2	-0.17	0.31	0.28

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk) and are adjusted for the estimated regressors. Singletons were dropped. The explanatory variable of interest is predicted dispersion. The first column displays the results of 2SLS regressions as in (7)*. Columns 2 and 3 display the results of the two-step regression as in* (13)*.* $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

Table 20: *Predicted dispersion and total returns, Duesseldorf (1984-2022)*

	Rental Yields	Capital Gains	Total Returns
Predicted dispersion, $\hat{\sigma}_{i,tq}$	6.59***	1.22	5.76**
	(0.730)	(1.605)	(2.049)
Year-quarter FEs	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes
Property characteristics	Yes	Yes	Yes
Holding period FEs	No	Yes	Yes
N	25238	13037	13037
R^2	0.15	0.27	0.25

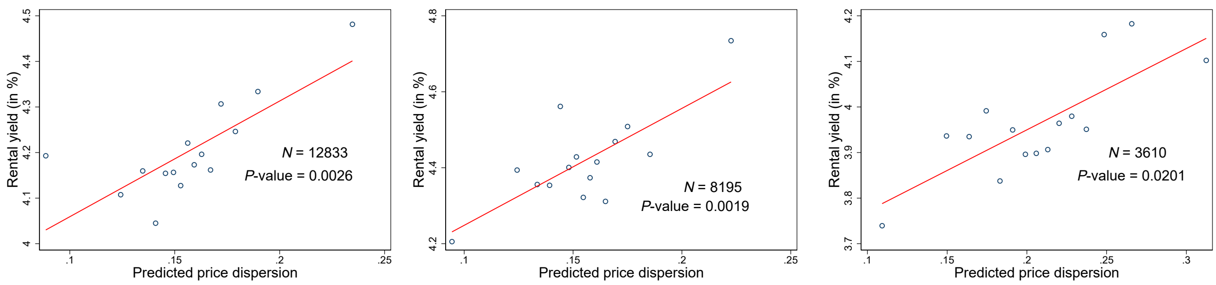
Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk) and are adjusted for the estimated regressors. Singletons were dropped. The explanatory variable of interest is predicted dispersion. The first column displays the results of 2SLS regressions as in (7)*. Columns 2 and 3 display the results of the two-step regression as in* (13)*.* $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

H.1 Predicted dispersion and rental yields - robustness

To address any biases that might arise from the matching process between transaction prices and rental values, I replicate the exercise in the previous sectios using only the subsample of transactions for which I also observe the rent data at the point of transaction. The results can be found in Figure 13, which shows that the main results hold. Comparing transactions in the same neighborhood and year-quarter and controlling for size and property characteristics, the data shows that properties with higher predicted dispersion, on average, have significantly higher rental yields than the rest.

Figure 13: *Rental yields and predicted dispersion using observed rent data*

(a) *Berlin* **(b)** *Hamburg* **(c)** *Duesseldorf*



Note: The first and last columns display a binscatter of log sales price and rental yields on idiosyncratic risk respectively. The second columns displays a binscatter of capital gains on the sum idiosyncratic risk from sale and re-sale. In all binscatters the underlying regressions include year-quarter and neighborhood fixed-effects as well as controls for property characteristics.

H.2 Predicted dispersion and rental yields - multi-family housing

In this section of the appendix, I present the regression outputs that form the basis for the results regarding the relationship between predicted dispersion and rental yields in the multi-family housing market. The tables that follow are generated from regression equation (7) with the ratio of net rental income to transaction price of multi-family houses as the dependent variable. The data samples used are drawn from multi-family housing transactions in Berlin spanning the period from 1970 to 2022 and in Hamburg from 1991 to 2022. For this analysis, I include only those transactions for which both the rental income and the transaction price are observed.

Table 21: *Predicted dispersion and rental yields for multi-family housing, Berlin (1970-2022)*

	Rental Yield	Rental Yield	Rental Yield (wo mixed-use)
Predicted Dispersion, $\hat{\sigma}_{i,tq}$	2.38*	5.03**	4.17**
	(1.306)	(1.031)	(1.139)
Year-quarter FEs	Yes	Yes	Yes
Year $\times$ Neighborhood FEs	Yes	Yes	Yes
Property characteristics	No	Yes	Yes
N	14332	14332	6841
R^2	0.00	-0.07	-0.02

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk) and are adjusted for the estimated regressors. Singletons were dropped. The explanatory variable of interest is predicted dispersion. All the columns display the results of 2SLS regressions as in (7) with ratio of net rental income to transaction price as the outcome variable The third column displays results only for the sample of multi-family housing that do not have any kind of commercial properties. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 22: *Predicted dispersion and rental yields for multi-family housing, Hamburg (1991-2022)*

	Rental Yield	Rental Yield	Rental Yield (wo mixed-use)
Predicted Dispersion, $\hat{\sigma}_{i,tq}$	7.93**	8.60**	6.76***
	(3.259)	(3.366)	(2.006)
Year-quarter FEs	Yes	Yes	Yes
Year $\times$ Neighborhood FEs	Yes	Yes	Yes
Property characteristics	Yes	Yes	Yes
N	7633	7633	5171
R^2	-0.00	-0.02	0.02

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk) and are adjusted for the estimated regressors. Singletons were dropped. The explanatory variable of interest is predicted dispersion. All the columns display the results of 2SLS regressions as in (7) with ratio of net rental income to transaction price as the outcome variable The third column displays results only for the sample of multi-family housing that do not have any kind of commercial properties. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

I Robustness Tests

I.1 All sales

In this section of the Appendix, I present the results for the analysis in which I utilize all property sales data to measure value uncertainty at the transaction property level, not just focusing on repeated property sales. The objective of this analysis is to ensure that the results are not influenced by specific characteristics of properties that are sold more frequently, which could distinguish them from the rest of the housing stock. The results for each city can be found in the tables below. These results corroborate the findings from the baseline analysis. Properties with higher value uncertainty are, on average, transacted at lower prices and yield higher rental returns. Once again, there is no statistically significant relationship between value uncertainty and the rental value of the property, affirming that the rental market is relatively liquid, and therefore, these properties are not rented out at a discount.

Table 23: *Predicted price dispersion, sales prices and rent using all sales (Berlin 1984-2022)*

	Sales Price	Sales Price	Net Rent	Net Rent	Rental Yield	Rental Yield
Predicted dispersion, $\hat{\sigma}_{it}$	-5.91***	-1.11***	-8.68***	-0.09	2.23**	4.72***
	(0.545)	(0.182)	(1.132)	(0.235)	(0.930)	(0.530)
Year-quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes	Yes	Yes	Yes
Property characteristics	No	Yes	No	Yes	No	Yes
N	190144	190144	190144	190144	190144	190144
R^2	-1.70	0.63	-3.27	0.90	0.09	0.07

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). The outcome variables in columns 1 to 4 have been standardized to have mean 0 and standard deviation of one. Singletons were dropped. The explanatory variable of interest is predicted price dispersion. The coefficients are estimated in the 2SLS regression framework of (7). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 24: *Predicted price dispersion, sales prices and rent using all sales (Hamburg 2001-2022)*

	Sales Price	Sales Price	Net Rent	Net Rent	Rental Yield	Rental Yield
Predicted dispersion, $\hat{\sigma}_{it}$	-6.72***	-1.39***	-7.76***	0.85***	7.87***	9.25***
	(1.273)	(0.211)	(1.678)	(0.147)	(0.817)	(1.128)
Year-quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes	Yes	Yes	Yes
Property characteristics	No	Yes	No	Yes	No	Yes
N	81840	81840	81840	81840	81840	81840
R^2	-1.78	0.69	-1.93	0.89	-0.06	-0.21

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). The outcome variables in columns 1 to 4 have been standardized to have mean 0 and standard deviation of one. Singletons were dropped. The explanatory variable of interest is predicted price dispersion. The coefficients are estimated in the 2SLS regression framework of (7). $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

Table 25: *Predicted price dispersion, sales prices and rent using all sales (Cologne 1989-2022)*

	Sales Price	Sales Price	Net Rent	Net Rent	Rental Yield	Rental Yield
Predicted dispersion, $\hat{\sigma}_{it}$	-6.81***	-0.27	-7.33***	0.87**	7.70***	5.52***
	(1.169)	(0.303)	(1.056)	(0.292)	(2.012)	(0.552)
Year-quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes	Yes	Yes	Yes
Property characteristics	No	Yes	No	Yes	No	Yes
N	108103	108103	108103	108103	108103	108103
R^2	-1.44	0.71	-1.79	0.91	-0.03	0.13

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). The outcome variables in columns 1 to 4 have been standardized to have mean 0 and standard deviation of one. Singletons were dropped. The explanatory variable of interest is predicted price dispersion. The coefficients are estimated in the 2SLS regression framework of (7). $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

Table 26: *Predicted price dispersion, sales prices and rent using all sales (Duesseldorf 1984-2022)*

	Sales Price	Sales Price	Net Rent	Net Rent	Rental Yield	Rental Yield
Predicted dispersion, $\hat{\sigma}_{it}$	-7.54***	-0.53***	-8.02***	0.74***	9.72***	6.90***
	(1.087)	(0.164)	(1.163)	(0.111)	(1.592)	(0.914)
Year-quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year × Neighborhood FEs	Yes	Yes	Yes	Yes	Yes	Yes
Property characteristics	No	Yes	No	Yes	No	Yes
N	48893	48893	48893	48893	48893	48893
R^2	-2.34	0.73	-2.84	0.85	-0.28	0.06

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). The outcome variables in columns 1 to 4 have been standardized to have mean 0 and standard deviation of one. Singletons were dropped. The explanatory variable of interest is predicted price dispersion. The coefficients are estimated in the 2SLS regression framework of (7). $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

I.2 Different measures of predicted dispersion

This suggests that a significant proportion of the impact of my measure of idiosyncratic risk on sales prices can be attributed to variations in observable property traits. To tackle this issue, I devise an instrumental variable that approximates the idiosyncratic price risk without being directly dependent on the property characteristics. Following Jiang and Zhang (2022), I build an instrument based on the distance of the properties' i characteristics to the mean characteristics of the properties' sold in the same city and within the same period:

$$Z_{it}^{m} = (X_{it}^{m} - \bar{X}_{ct}^{m})^2, \ \forall m \in \{size, age, location\}. \tag{52}$$

This measure captures the degree of thinness in the local property market for property i, which in turn reflects the uncertainty surrounding its sales price. For instance, pricing an old and large apartment in a neighborhood predominantly composed of new and small apartments can be challenging. Additionally, I also build a measure based directly on the relative frequency of the combination of characteristics of an apartment. Every quarter I assign each transaction a specific bin depending on its size, location and age. The idea is to capture how frequently a specific combination of characteristics appears on the market at a given point in time:

$$Z_{it}^{m} = \frac{obs_{it}}{obs_t}, \ \forall m \in \{size, age, location\} \tag{53}$$

Building upon these concepts, I perform two-stage least squares (2SLS) regressions by utilizing the distances and the relative frequency Z_i as instruments to approximate the variance of the idiosyncratic price deviations:

$$\text{Stage 1:} \quad u_{it}^2 = \alpha + \beta_1 Z_{it}^{age} + \beta_2 Z_{it}^{size} + \beta_3 Z_{it}^{location} \tag{54}$$

$$+ B_X X_i + \kappa_{nt} + \mu_d + e_{it} \tag{55}$$

$$\text{Stage 2:} \quad ln(P_{it}) = \alpha + \gamma \hat{u}_{it} + B_X X_i + \kappa_{nt} + \mu_d + \epsilon_{it}. \tag{56}$$

The outcome of the 2SLS regressions for Berlin are presented in Table 27 and for Hamburg in Table 28. For the purpose of comparison, I have also included the results of my main baseline analysis in the first column. The coefficients remain negative and highly statistically significant, and it are of similar size to the coefficient of my baseline analysis. This indicates that these measures directly capture a significant portion of the variation in the sales price that is not explained by the property characteristics or the time-fixed effects, but rather by the degree of idiosyncratic sales price risk.

Table 27: *Log sales prices and idiosyncratic risk* (Berlin, 1989-2022)

	Benchmark	Distance	Frequency
Log idiosyncratic risk, $\hat{\sigma}_{it}$	-0.0444***	-0.0323***	-0.0484***
	(0.0144)	(0.0063)	(0.0081)
Year-month FEs	Yes	Yes	Yes
Quarter × Neighborhood FEs	Yes	Yes	Yes
Property characteristics	Yes	Yes	Yes
N	69123	69123	69123
R^2	0.63	0.64	0.63

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). Coefficients are estimated via two-stage least squares. Singletons were dropped. The explanatory variable of interest is idiosyncratic risk. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

Table 28: *Log sales prices and idiosyncratic risk* (Hamburg, 2002-2022)

	Benchmark	Distance	Frequency
Log idiosyncratic risk, $\hat{\sigma}_{it}$	-0.0297***	-0.0407***	-0.0261***
	(0.0073)	(0.0043)	(0.0077)
Year-month FEs	Yes	Yes	Yes
Quarter × Neighborhood FEs	Yes	Yes	Yes
Property characteristics	Yes	Yes	Yes
N	53824	53824	53824
R^2	0.70	0.69	0.70

Note: Standard errors are clustered at the neighborhood-level (Stadtbezirk). Coefficients are estimated via two-stage least squares. Singletons were dropped. The explanatory variable of interest is idiosyncratic risk. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

I.3 Building renovations and predicted dispersion

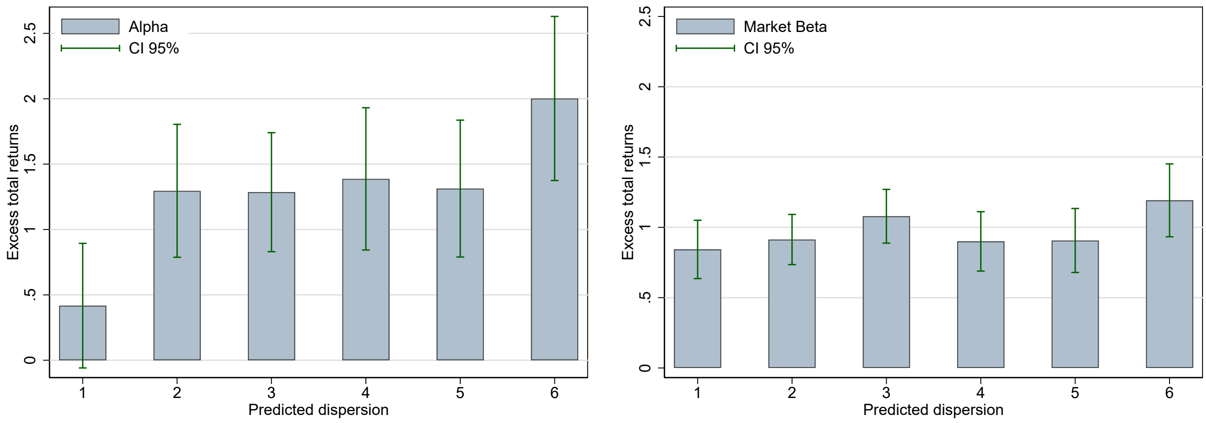
In Figure 14, panel a, I plot the excess returns adjusted for the exposure to the market portfolio for all the six portfolios. Again, the portfolio containing the properties with the highest level of idiosyncratic risk overperforms all the othe portfolios. Panel b shows that the exposure to the market portfolio is almost flat across the idiosyncratic risk ditribution.

I.4 Length of holding periods and predicted dispersion

Following the same steps as before, I build different portfolios based on this new measure of the variance of shocks. I then compare the returns to these portfolios and

Figure 14: *Excess returns with building-time fixed effects, Cologne 1989-2022*

(a) *alpha* **(b)** *market beta*

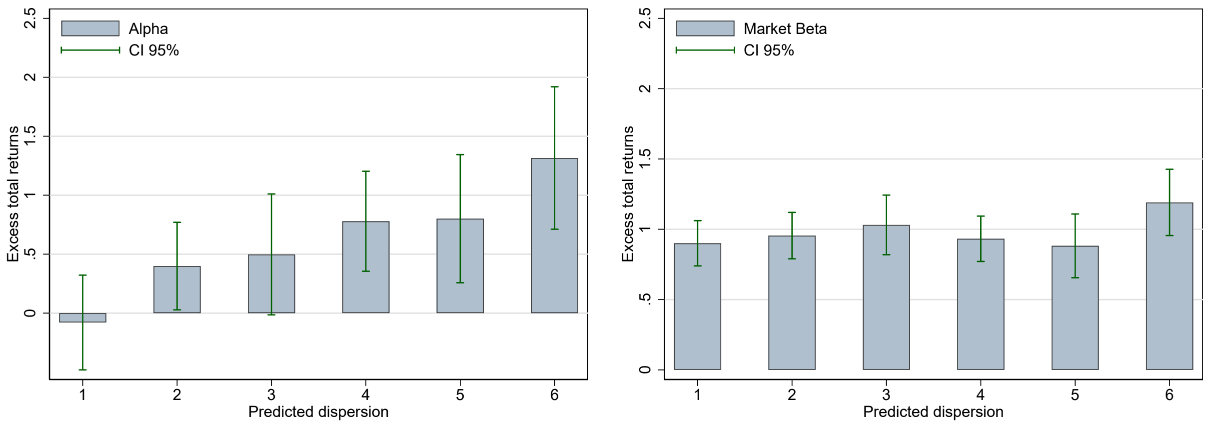


Note: The six equally-sized portfolios are built based on predicted variance of idiosyncratic shocks quantiles from equation 17. Standard errors are adjusted for time-series autocorrelation using Newey-West with 4 quarter lags.

find the same pattern as in the baseline analysis.

Figure 15: *Excess returns controlling for holding periods, Cologne 1989-2022*

(a) *alpha* **(b)** *market beta*



Note: The six equally-sized portfolios are built based on predicted variance of idiosyncratic shocks quantiles from equation **??***. Standard errors are adjusted for time-series autocorrelation using Newey-West with 6 quarter lags.*

J Rent data sources

J.1 Source

The rent data comes from the so-called 'Mietspiegel,' which are documents produced at the city level containing estimates for the average rent per square meter for apartments in the private market, depending on their size, building year, location and condition. The data is collected via a survey, and then the aggregate estimates are published every two years. The Mietspiegel provides benchmark rents that can be used by landlords to set their rents. If the rents deviate significantly from the benchmarks provided in the Mietspiegel, tenants have the option to file an official complaint. In such cases, landlords are obliged to adjust the rents to a level that aligns more closely with the Mietspiegel benchmarks. The publication of the Mietspiegel is typically the responsibility of the city. Until the 2000s, the data collection and estimates in the Mietspiegel were mostly produced by the city itself. Since then, most cities have started to hire specialized companies, which typically survey larger samples and produce rent estimates based on hedonic regressions. As a result, the quality of the estimates has substantially improved in the last 20 years (Steffen and Memis, 2021). The quality of the estimates provided by the Mietspiegel has been examined in several research papers. Specifically, researchers have sought to analyze the reliability of rental estimates derived from the Mietspiegel by comparing them to estimates based on alternative sources of rent data. In a study by Thomschke (2022), micro-level rents from the German census and online asking rents were utilized to generate rent estimates for the segments represented in the Mietspiegel for major German cities. The authors then compared their own estimates with those from the Mietspiegel and found minimal differences, thus affirming the validity of the estimates provided by the Mietspiegel. On the other hand, Rendtel et al. (2021) claims that the Berlin Mietspiegel underestimates the average value of rents by approximately 14% due to oversampling of large landlords. However, it is worth noting that this bias appears to be evenly distributed across all rent classes and, therefore, does not significantly impact the comparisons across rent classes, which are more relevant for the results I present.

Overall, the key point is that rental returns tend to increase with idiosyncratic risk. This relationship is clearly evident in cases where I have access to both rental and price data for the same property. Furthermore, it appears that this relationship holds true when using the Mietspiegel data as well, indicating its reproducibility.

J.2 Matching process

All the rent estimates provided in the Mietspiegel are net of utilities, meaning they do not include heating, water, electricity, and maintenance costs. The rent estimates are provided based on different criteria such as building years, size, location, and condition of the apartments. Additionally, only monthly rent per square meter estimates are provided. Regarding building years, the Mietspiegel typically distinguishes between apartments built before WWI, between WWI and WWII, and provides estimates for each post-WWII decade. In terms of size, the Mietspiegel typically categorizes apartments as less than 40 square meters, between 40 and 60 square meters, between 60 and 90 square meters, and more than 90 square meters. For location, the Mietspiegel usually differentiates between regions of varying quality within the city, commonly referred to as bad, middle, and good-quality regions. In most cases, the information about location quality is already available in the transaction data set. Lastly, the Mietspiegel also distinguishes between apartments with an own bathroom or central heating and those that have both of these amenities. As is expected, the rents are higher for those apartments, which have both of these amenities. This distinction is typically only done for apartments that were built before the 1970s, as the majority of apartments built after have both of these amenities.

Overall, the Mietspiegel provides a wide range of rent per square meter estimates based on the mentioned characteristics. By using the building year, size, and location quality, I am able to match the transaction data with the corresponding rent per square meter estimates. The only category that presents challenges in terms of matching is whether the apartment has its own bathroom or central heating. As previously stated, this issue primarily affects properties built before the 1970s.

In my primary analysis, I focus solely on rent estimates for apartments with both central heating and an own bathroom. However, it's important to acknowledge that this approach may introduce potential bias into the results by potentially overestimating the rental yields for properties with higher idiosyncratic risk.

To address this concern, I conduct a robustness analysis where I match transactions to the rent data based on the relative value of the transactions in that specific year, along with their corresponding characteristics. If a property is sold for a price above the median considering the year, size, and building year, it is matched with the rent estimate for properties with both central heating and an own bathroom. Conversely, if a property is sold for a price below the median, it is matched with the rent estimate for properties with either an own bathroom or central heating. In this case the results are also hold through.

References

Attanasio, Orazio, Andrew Leicester, and Matthew Wakefield (2011). "Do House Prices Drive Consumption Growth? The Coincident Cycles of House Prices and Consumption in the UK". In: *Journal of the European Economic Association* 9(3), pp. 399–435.

Bali, Turan G, Robert F Engle, and Scott Murray (2016). *Empirical asset pricing: The cross section of stock returns*. John Wiley & Sons.

Berger, David et al. (2018). "House prices and consumer spending". In: *The Review of Economic Studies* 85(3), pp. 1502–1542.

Berk, Jonathan B, Richard C Green, and Vasant Naik (1999). "Optimal Investment, Growth Options, and Security Returns". In: *The Journal of finance* 54(5), pp. 1553–1607.

Blundell, Richard, Luigi Pistaferri, and Ian Preston (2008). "Consumption Inequality and Partial Insurance". In: *American Economic Review* 98(5), pp. 1887–1921.

Campbell, John Y and Joao F Cocco (2007). "How do House Prices Affect Consumption? Evidence from Micro Data". In: *Journal of Monetary Economics* 54(3), pp. 591–621.

Case, Karl E, John Cotter, and Stuart A Gabriel (2011). "Housing risk and return: evidence from a housing asset-pricing model". In: *Journal of Portfolio Management* 37(5).

Case, Karl E and Robert J Shiller (1988). *The efficiency of the market for single-family homes*.
Constantinides, George M and Darrell Duffie (1996). "Asset Pricing with Heterogeneous Consumers". In: *Journal of Political Economy* 104(2), pp. 219–240.
Giacoletti, Marco (2021). "Idiosyncratic Risk in Housing Markets". In: *The Review of Financial Studies* 34(8), pp. 3695–3741.
Herskovic, Bernard et al. (2016). "The Common Factor in Idiosyncratic Volatility: Quantitative Asset Pricing Implications". In: *Journal of Financial Economics* 119(2), pp. 249–283.
Jiang, Erica Xuewei and Anthony Lee Zhang (2022). "Collateral Value Uncertainty and Mortgage Credit Provision". In: *Available at SSRN*.
Mian, Atif, Kamalesh Rao, and Amir Sufi (2013). "Household Balance Sheets, Consumption, and the Economic Slump". In: *The Quarterly Journal of Economics* 128(4), pp. 1687–1726.
Piazzesi, Monika, Martin Schneider, and Selale Tuzel (2007). "Housing, consumption and asset pricing". In: *Journal of Financial Economics* 83(3), pp. 531–569.
Rendtel, Ulrich, Steffen P Sebastian, and Michael Frink (2021). "Ist der Berliner Mietspiegel 2019 qualifiziert? Ein alternativer Mietspiegel mit Daten des Mikrozensus". In: *Stadtforschung und Statistik: Zeitschrift des Verbandes Deutscher Städtestatistiker* 34(1), pp. 72–91.
Sagi, Jacob S (2021). "Asset-level risk and return in real estate investments". In: *The Review of Financial Studies* 34(8), pp. 3647–3694.
Steffen, Sebastian and Halil Memis (2021). "gif-Mietspiegelreport 2021". In: *Gesellschaft für Immobilienwirtschaftliche Forschung*.
Stroebel, Johannes and Joseph Vavra (2019). "House Prices, Local Demand, and Retail Prices". In: *Journal of Political Economy* 127(3), pp. 1391–1436.
Thomschke, Lorenz (2022). "Einfache Mietspiegel qualifizieren: Alternative Daten zur Ermittlung ortsüblicher Vergleichsmieten". In: *Stadtforschung und Statistik: Zeitschrift des Verbandes Deutscher Städtestatistiker* 35(1), pp. 97–107.